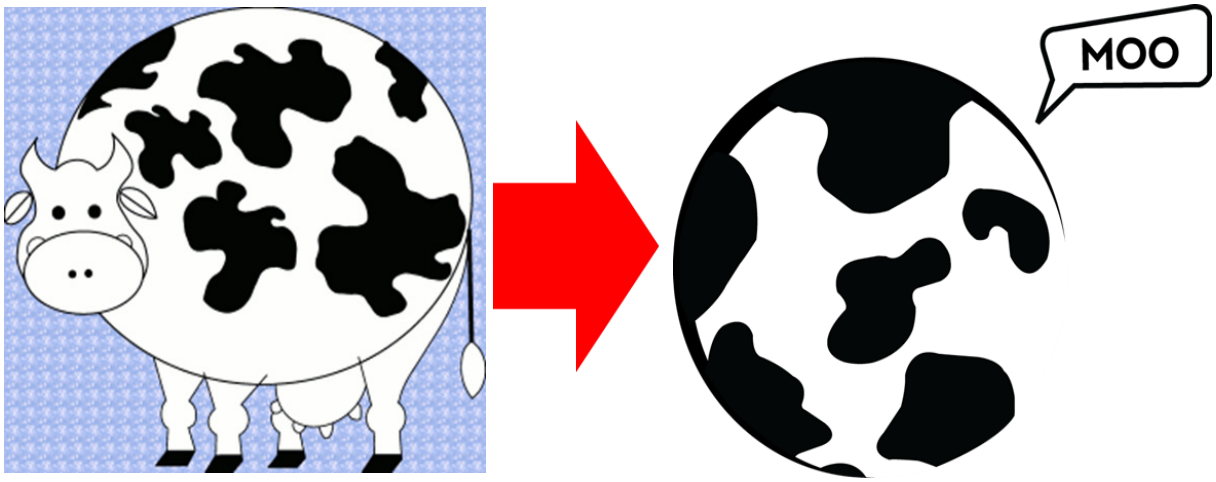
Complex Physics

Kim Sneppen and Jan O. Haerter



Contents

Perspectives

Complexity is a somewhat undefined concept. This may be so because a complex system is more than the simple sum of its parts. Complex systems often show emergent behavior, which is inherently not apparent from the basic interaction or basic rules between the individual parts of the system. Complex systems share this feature with fractals and systems that can be characterized using power laws. In fact the study of such scale-free systems has inspired much of complex system research. Both fractals and complex systems take some time to create, even when the basic mathematics defining them is simple — a notion that might inspire a quest of inventing a simple algorithm that, when executed, only requires time, perhaps a lot of time, in order to generate complexity.

Fractals repeat themselves at disparate scales and thereby connect phenomena across these scales, fueling our basic dream of physics to connect apparently different phenomena — and with such connections to perhaps understand that the patterns we see may not depend on details of the system or the explicit events considered.

Coherence. A typical feature of complex systems is, that they often display some sort of coherence, i. e. *that different parts of the system appear to march to the same drummer*. In contrast to such coherent, or partly coordinated dynamics, other large systems may behave more like a collection of many independent smaller systems. In that case, the overall behavior of the smaller systems does not depend on that of the total system. This would then be a more boring "equilibrium-like" behavior associated to the addition of the many uncorrelated sources.

Think power laws! These lectures aim to educate the student in ways to think about the complex world that surrounds us. These notes often do this in terms of examples, drawn from statistical mechanics and complex systems science. A recurring theme is that of "power laws," and the wide array of natural phenomena that repeat themselves across many scales. These systems are then said to "scale". We will see that such "scaling" is relatively common, and that it can have several origins. However, it nearly always emerges from some far from equilibrium dynamics, with a taint of positive feedback.

These notes aim to give the students the following skills:

- Chapter 1 provides an introduction to **critical phenomena**. Fundamental concepts from equilibrium statistical mechanics are first laid out and it is then discussed, how abrupt transitions of an observable can occur in the thermodynamic limit. These basic concepts set the background for subsequent chapters, in particular by emphasizing scaling properties.

- Chapter 2 revisits the concept of criticality in terms of scaling properties around the critical point in **percolation**. This analysis allows us to

introduce the concept of fractals and the relation between fractal dimensions and power laws.

- Chapter 3 introduces **random walkers and branching processes**. We show the connection between non-linearity in terms of stick-slip dynamics and self-organization. We will see that, when individuals parts are either moving or completely at rest, a dynamics termed "**self-organized criticality**" can emerge, provided that time scales are infinitely separated.

- Chapter 4 introduces the student to basic concepts of **complex networks**, including the ubiquitous scale-free networks and their potential origins. The discussion includes two processes for emergence of power laws, historically used to describe very different systems — from human wealth to sizes of asteroids in the universe. We also illustrate, how to analyze systems with many components in terms of null models and algorithms for network construction.

- Chapter 5 teaches the student about **agent-based models**, stochastic event-based simulations and how to describe self-organization from individual to collective behavior.

- Chapter 6 introduces **concepts from economics** that can be analyzed and modelled using methods from physics. We discuss basic time series analysis, agent-based models and bet-hedging aspects of game theory.

- Chapter 7 aims to broaden the student's view of non-linear physics in dynamics of extended systems (systems with many degrees of freedom). This is accomplished through the discussion of dynamical **fronts and interfaces**. Interfaces provide examples of stochastic dynamics, chaos and self-organized criticality, thus drawing a link to Ch. 3.

... some practical notes:

Homework exercises will be assigned as the course progresses. They will usually be listed at the end of the respective chapters in these lecture notes. Note that, in some cases, exercises titled similarly will be available, however, one version will contain a plus (+) symbol. The exercises labeled with a "+", are more open versions of the alternative ones, but lead you to similar results. For completion of the course, you can go with the more detailed ones, but if you get bored, the "+" version will simply be more challenging, as less explicit guidance will be available in those variants. So, it is entirely up to you, which one you work on — pick one, and try to get through it. Additionally, group work on homework problems is explicitly encouraged. You should always work through all problems and if you get stuck, discuss with your classmates or visit me in my office. Please work through the assignments already at home,

before you come to exercise sessions. This will help you get the most out of the tutorials we offer.
Please do the *computer exercises*, they are a very important part of the course. They are key methodology that is absolutely needed for succeeding in complex systems science!

Mini tutorials (marked green) are interspersed throughout the text. They intend to quickly raise some thought which could be recapitulated after reading a section. These should usually be quick to answer and not difficult after having read the previous paragraphs.

Overall, with the lectures and exercises we intend to provide the student with knowledge of a set of model types and simulation algorithms that are useful in understanding our surrounding world. They aim to give the student a feeling of playfulness when thinking about putting "Life, the Universe and Everything" on a computer:
"The popular view that scientists proceed inexorably from well-established fact to well-established fact, never being influenced by any unproved conjecture, is quite mistaken. Provided it is made clear which are proved facts and which are conjectures, no harm can result. Conjectures are of great importance since they suggest useful lines of research."

- Alan Turing; The Enigma

"The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work-that is, correctly to describe phenomena from a reasonably wide area."

- John von Neumann

"Truth is much too complicated to allow anything but approximations."

- John von Neumann

"A complicated idea is a confused idea."

- Marty Rubin

"It's fun to invent systems and meanings and then poke holes in them."

- Marty Rubin

"I'm all in favour of the democratic principle that one idiot is as good as one genius, but I draw the line when someone takes the next step and concludes that two idiots are better than one genius. "

- Leo Szilard

"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk."

- John von Neumann

"Simplicity is a great virtue but it requires hard work to achieve it and education to appreciate it. And to make matters worse: complexity sells better."
- Edsger W. Dijkstra

"Don't be fooled by the many books on complexity or by the many complex and arcane algorithms you find in this book or elsewhere. Although there are no textbooks on simplicity, simple systems work and complex don't."
- Jim Gray

"Truth is ever to be found in simplicity, and not in the multiplicity and confusion of things."
- Isaac Newton

Chapter 1

Statistical Mechanics and critical phenomena

1.1 Review of statistical mechanics

1.1.1 Entropy: Measuring Ignorance

At the base of statistical mechanics lies the assumption that entropy, which is a measure of the "lack of information" about a given many-particle ensemble, must be maximized [1]. Consider a large number of systems N_S, i.e. $N_S \to \infty$. Each of these systems can be in a certain state i, and n_i counts the number of systems that are in the state i, one of a total of q possible states (Fig. 1.1.1). If there were only one state, then we would have complete knowledge, since the state of all systems would be the same. If we had a number of possible states, we can measure the *ignorance* I we have about the system. I is the number of ways the N_S systems can be re-arranged, i.e. the factorial $N_S!$ divided by the number of ways the systems of equal state can be re-arranged:

$$I \equiv \frac{N_S!}{n_1!\, n_2!\, \ldots\, n_q!} \,. \tag{1.1}$$

I simply measures the multiplicity of the outcome, where the states are populated according to the numbers n_i. The aim is to find the numbers n_i such that ignorance is maximized while satisfying a constraint.

We are not forced to maximize I, instead, we could just as well maximize any monotonically increasing function of I. Instead of the ignorance I, for mathematical reasons it is convenient to use a different quantity S, the *entropy*, defined as

$$S \sim \frac{1}{N_S} \ln(I) \,. \tag{1.2}$$

When I is maximized, so is S. We use the proportionality symbol in order to indicate, that one has the liberty to choose the proportionality constant as one pleases, the important aspect about S is that it is an increasing function of I. The use of S is convenient, as for large numbers N_S and n_i, Stirling's approximation can be used to transition from the discrete factorial to continuous functions:

$$\lim_{N\to\infty} \ln N! = N \ln N - N + (1/2) \ln N + (1/2) \ln(2\pi) + 1/(12N) + \ldots \,, \tag{1.3}$$

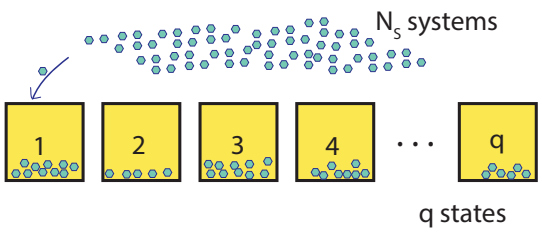


Figure 1.1: **Organizing N_S systems into q states.**

where for large N retaining only the first two terms of the RHS is a very good approximation[1]. Using only these, for $N_S \to \infty$, which we sometimes call the *thermodynamic limit*, the entropy S in Eq. 1.2 can be re-written as

$$S \sim \frac{1}{N_S}\left(N_S \ln N_S - \sum_i n_i \ln n_i - N_S + \sum_i n_i + \dots\right) \tag{1.4}$$

$$\sim -\sum_i p_i \ln p_i \ . \tag{1.5}$$

The probabilities $p_i \equiv n_i/N_S$ thereby denote the likelihood that a given system is in state i. The proportionality allows free choice of a constant, and for physical systems

$$S = -k_B \sum_i p_i \ln p_i \ , \tag{1.6}$$

with $k_B = 1.38 \times 10^{-23}\mathrm{JK}^{-1}$ is conventional, hence entropy has units of energy divided by temperature. We note that it would be equally reasonable to absorb k_B into the definition of temperature T, to be defined below. In that case, temperature would simply be measured in units of energy and entropy would remain dimensionless.

There are several aspects to point out about Eq. 1.5: Given that there is only one possible state n_1, then $n_1 = N_S$ and the entropy S vanishes. For cases with more than one occupied state, $0 \leq p_i < 1 \ \forall i$, and entropy is always positive. As a measure of "disorder", entropy S is the fundamental quantity of statistical physics.

1.1.2 Lagrange multipliers — maximization under constraints.

As mentioned, statistical mechanics builds on the principle that entropy must be maximized. What we did not mention above is, that such a maximization takes place under a given number of so-called *constraints*. The entropy (Eq. 1.5) is a multidimensional function of the values p_i. In 2D, i.e., $i \in \{1, 2\}$, its value can be visualized as a surface (Fig. 1.1.2) and the gradient ∇S is a vector pointing in the direction of steepest slope. Any constraint, $C(p_1, p_2, \dots) = 0$, is also a function of the p_i, and we define such constraints to equal zero. In the 2D example (Fig. 1.1.2), a possible (but not at all physically inspired) constraint could be that $p_1^2 + p_2^2 - 1 = 0$, i.e. that the values of p_i lie on a unit circle in the p_1-p_2-plane. The gradient of the constraint is $\nabla C(p_1, p_2, \dots)$ and points in the direction in which the constraint is most effectively modified. For the example of the circle, the gradient $\nabla C = 2(p_1, p_2)$ is the normal to the circle line.

[1] To loosely motivate Stirling's approximation, consider that $\ln N! = \ln(1 \cdot 2 \cdot \dots \cdot N) = \sum_n \ln n \approx \int_1^N \ln n = N \ln N - N$.

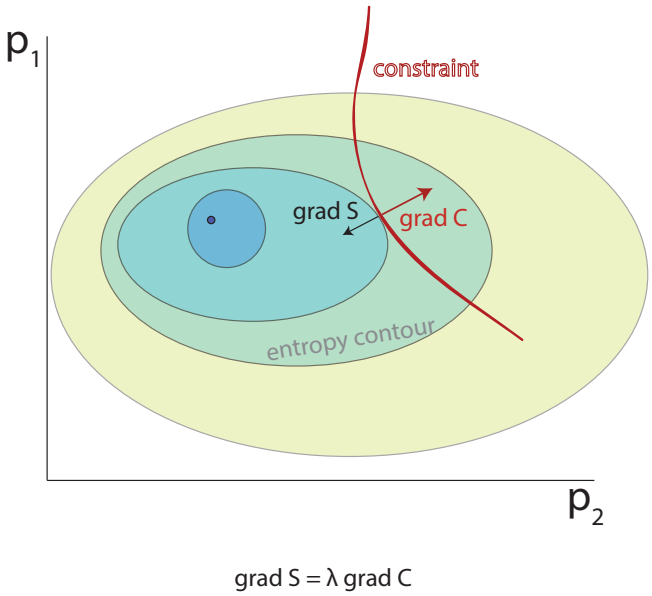


Figure 1.2: **Lagrange multiplier.** Cartoon illustrating the method of constraints for 2D probability space. Entropy is maximized under the constraint C where the gradients of C (the red line) and the entropy (colored patches) match.

Maximizing S subject to C means to find the values of p_i where the gradients ∇S and ∇C align, i.e.

$$\nabla S = \lambda \nabla C \ ,$$

the gradients may thereby differ by a constant λ, which is termed *Lagrange multiplier*. To ensure the constraint and express it analogously to the gradient, one often employs the notation

$$\frac{\partial}{\partial \lambda}(S - \lambda C) = 0 \ .$$

Hence, the Lagrange multiplier λ can be seen as an extra dimension to the problem. By adding this dimension, one has the advantage of obtaining a maximization problem without constraint.

Mini Tutorial: Entropy is defined in a q-dimensional space. Which dimension is a constraint condition defined in? How many constraints can one maximally have?

Constraint of normalization of probability. The previous example of a circle was entirely fabricated and physically not meaningful. In practice, one basic constraint is always present, namely that each system must be in one of the

available states, i.e. total probability must equal unity, or

$$\sum_j p_j - 1 = 0 \; . \tag{1.7}$$

In this case, the problem to solve is:

$$\begin{aligned} \frac{\partial}{\partial p_i} k_B \left(- \sum_j p_j \ln p_j - \lambda \left[\sum_j p_j - 1 \right] \right) &= 0 \; , \\ \frac{\partial}{\partial \lambda} k_B \left(- \sum_j p_j \ln p_j - \lambda \left[\sum_j p_j - 1 \right] \right) &= 0 \; . \end{aligned}$$

These equations give

$$\ln p_i = -\lambda - 1 \; , \qquad \text{hence } p_i = \exp(-\lambda - 1) \; , \tag{1.8}$$

as well as the imposed constraint $\sum_j p_j = 1$. Eq. 1.8 hence expresses the uniform distribution of probabilities, that is, the fact that all states are equally likely — given that there are no additional constraints. The only constraint, that of normalized probabilities, is fulfilled by proper choice of $\lambda = \ln q - 1$.
Mini Tutorial: Using $S = -k_B \sum_{i=1}^{q} p_i \ln p_i$, assume $q = 2$ and forget about all constraints: maximize S and compare to the entropy with probability constraint.

1.1.3 Conserved quantities

In the previous example we have already encountered a *conserved quantity*, namely total probability, stating that each system must be in one of the available states. However, in many physical contexts, such as a laboratory experiment, other quantities are also conserved — they might be controlled by the experimental setup, e.g. a closed container which constrains particle number or the total energy of the system.
Constraint of total energy. Consider total energy E, and allow each possible state i to come with a specific energy ϵ_i. Then the constraint on total energy is

$$\sum_j p_j \epsilon_j - E = 0 \; . \tag{1.9}$$

This additional constraint now requires one additional Lagrange multiplier and the complete equation for the gradients w.r.t. the p_i becomes

$$\frac{\partial}{\partial p_i} k_B \left(- \sum_j p_j \ln p_j - \lambda \left[\sum_j p_j - 1 \right] - \beta \left[\sum_j p_j \epsilon_j - E \right] \right) = 0 \; , \tag{1.10}$$

where β is the Lagrange multiplier corresponding to the total energy constraint. Evaluating the derivatives yields the probabilities

$$p_i = \exp(-1 - \lambda - \beta \epsilon_i) \sim \exp(-\beta \epsilon_i) \; . \tag{1.11}$$

Hence, the probability of occupying a given state is now dependent on the respective state energy and decays exponentially with that energy. Note that the exponentials in Eq. 1.11 are the usual Boltzmann probabilities. However, the new multiplier β must be fixed by ensuring that total energy equal E. β is often referred to as "inverse temperature", i.e. $\beta = 1/k_B T$. We emphasize that the Boltzmann constant k_B generally appears together with T, hence it is often easiest to absorb k_B into the definition of temperature — making the entropy dimensionless.

Constraint of total particle number. When total particle number is conserved, an additional Lagrange multiplier is required. In that case, the probabilities become

$$p_i = \exp(-1 - \lambda - \beta\epsilon_i - \alpha N_i) \sim \exp(-\beta\epsilon_i - \alpha N_i) \ , \tag{1.12}$$

where N_i are the particle numbers for the different states i and α is the Lagrange multiplier for particle number. For practical purposes, α is often re-expressed as $\alpha = -\mu/k_B T$, where μ is the "chemical potential" and T temperature.

Mini Tutorial: If l (out of q) distinct states have the same energy, what does this do to the joint probability weight corresponding to this energy?

1.1.4 Statistical ensemble.

Depending on which quantities are allowed to vary, statistical mechanics distinguishes several types of *statistical ensembles*. The *micro-canonical ensemble* considers both energy and particle number to be "fixed". By "fixed" it is hereby meant that the system has a specific value of total energy or particle number, no fluctuations regarding their values are allowed. This ensemble assumes that each state i has the same energy and probability.

In practical terms, the microcanonical ensemble is less realistic, since an experimental system would generally allow for some uncertainty regarding the fluctuations of energy. The *canonical ensemble* relaxes the need for identical energy levels. It allows for the states i to have distinct energy values but does require the total energy to be some average, or expectation value, E. The microstates are then occupied statistically, by the maximization of entropy under the total energy constraint, as discussed above. The *grand canonical ensemble* further relaxes the need for a fixed particle number in each state, but considers a total average N, which is again enforced as a constraint.

1.1.5 Partition function.

With probabilities proportional to $\exp(-\beta\epsilon_i - \alpha N_i)$ and the normalization constraint on unit total probability, the normalization constant Z can be defined as:

$$Z \equiv \sum_i \exp(-\beta\epsilon_i - \alpha N_i) \ , \tag{1.13}$$

through which the probabilities p_i are

$$p_i = \frac{1}{Z}\exp(-\beta\epsilon_i - \alpha N_i)\ . \tag{1.14}$$

Z is commonly referred to as the *partition function*, and can be useful in expressing observables, e.g. the average energy

$$\langle E\rangle = \frac{\sum_i \epsilon_i \exp(-\beta\epsilon_i - \alpha N_i)}{Z} = -\frac{\partial}{\partial\beta}\ln Z = T^2\frac{\partial}{\partial T}\ln Z \tag{1.15}$$

or total particle number

$$\langle N\rangle = \frac{\sum_i N_i \exp(-\beta\epsilon_i - \alpha N_i)}{Z} = -\frac{\partial}{\partial\alpha}\ln Z = T\frac{\partial}{\partial\mu}\ln Z\ . \tag{1.16}$$

Note that, in the previous two equations, it is assumed that α, respectively β, is held fixed when evaluating the derivatives.

Notably, the partition function is also related to the entropy:

$$\begin{aligned} S &= -k_B\sum_i p_i \ln p_i = k_B\sum_i p_i(\ln Z + \beta\epsilon_i + \alpha N_i) &(1.17)\\ &= k_B\ln Z + \langle E\rangle/T + \mu\langle N\rangle/T\ . &(1.18)\end{aligned}$$

Mini Tutorial: Using $S = -k_B\sum_i p_i\ln p_i$, show how the entropy in the canonical ensemble relates to that in the microcanonical one. (*Hint*: Use a change of variables to sum over energies rather than individual states).

1.1.6 Thermodynamic potentials

Logarithms of partition functions are often referred to as *thermodynamic potentials* or *free energies*. When both particle number and energy are allowed to vary, the corresponding potential is referred to as *grand canonical potential*,

$$\Omega_{GC} \equiv k_B T\ln Z_{GC} = TS - \langle E\rangle + \mu\langle N\rangle\ , \tag{1.19}$$

which is a reformulation of Eq. 1.18.

In the canonical case, where only states of the same particle number are considered, the term μN is missing and the potential is referred to as the *(Helmholtz) free energy*,

$$F \equiv -k_B T\ln Z_C = \langle E\rangle - TS\ , \tag{1.20}$$

where we note that the opposite sign is conventional as compared to Eq. 1.19.

In the microcanonical case, the potential is just the entropy.

Once the partition function is known, observables can be evaluated by taking appropriate derivatives, e.g. the internal energy

$$\langle E\rangle = -\frac{\partial\ln Z}{\partial\beta}\ , \tag{1.21}$$

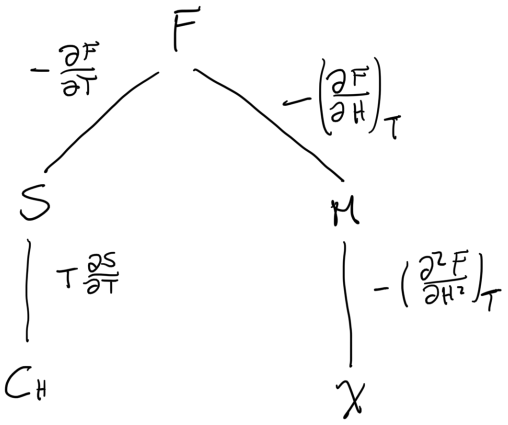


Figure 1.3: **Free energy and possible derivatives.** Here, a magnetic system is assumed, where M represents total magnetization, H is the external magnetic field and χ is the magnetic susceptibility.

specific heat

$$C_H \equiv \left(\frac{\partial\langle E\rangle}{\partial T}\right)_H = \frac{\partial\langle E\rangle}{\partial\beta}\frac{d\beta}{dT} = -\frac{1}{k_B T^2}\frac{\partial\langle E\rangle}{\partial\beta} = \frac{1}{k_B T^2}\frac{\partial^2 \ln Z}{\partial\beta^2} \,, \tag{1.22}$$

where H is some quantity that is held constant. In the following sections, the state energy will often be modified by a term $-M\,H$, where M is magnetization and H plays the role of the external magnetic field. it is hence assumed that the energy has a term $-M\,H$ in addition to the internal degrees of freedom. H can e.g. be controlled within the experimental setting and acts as the "generalized force", while M acts as the "generalized displacement". This should be compared to the usual term $-p\,V$, where the pressure p is the generalized force and the volume V the generalized displacement.

Note that the derivative w.r.t. temperature could also be carried out explicitly in the partition function, as

$$\frac{\partial\langle E\rangle}{\partial\beta} = -\frac{1}{k_B T^2}\frac{\partial}{\partial\beta}\frac{\sum \epsilon_n \exp(-\beta\epsilon_n)}{\sum \exp(-\beta\epsilon_n)} \tag{1.23}$$

which leads to a relation between (microscopic) root-mean-square fluctuations σ_E and the (macroscopic) specific heat (*see*: exercises). One can hence measure the specific heat without perturbing the external temperature, it is sufficient to observe the fluctuations in equilibrium.

From the free energy further quantities can be obtained by taking derivatives w.r.t. the controllable variables (Fig. 1.1.6):

$$S = -\left(\frac{\partial F}{\partial T}\right)_H = \frac{\langle E\rangle - F}{T} \,, \tag{1.24}$$

magnetization

$$M = -\left(\frac{\partial F}{\partial H}\right)_T \ , \tag{1.25}$$

or susceptibilities

$$\chi_T = -\left(\frac{\partial^2 F}{\partial H^2}\right)_T \ , \tag{1.26}$$

where the subscript T denotes that temperature is held constant while evaluating the derivative.

Mini Tutorial: Show how the canonical free energy relates to the microcanonical one. (*Hint*: Again use a change of variables and start by defining a microcanonical free energy.)

1.1.7 Exercises

1. Two electrons.
Consider two single-particle levels with energies $-\epsilon$ and ϵ. In these levels place two electrons (no more than one electron of the same spin per level). As a function of T find: (a) the partition function; (b) the average energy; (c) the entropy; (d) for microcanonical ensembles corresponding to each system energy level, compute the entropy; (e) for a—c, discuss the limits $T = 0$ and $T \to \infty$.

2. Fluctuations. [2]
(i) Verify that

$$\langle (M - \langle M \rangle)^2 \rangle = \langle M^2 \rangle - \langle M \rangle^2 = k_B^2 T^2 \frac{\partial^2}{\partial H^2} \ln \mathcal{Z} = k_B T \chi_T \ .$$

(ii) Show in a similar way that the fluctuations in the energy are related to the specific heat at constant volume by

$$(\Delta E)^2 = \langle (E - \langle E \rangle)^2 \rangle = k_B T^2 C_V \ .$$

Use this equation to argue that $\Delta E \sim N^{1/2}$ where N is the number of particles in the system.

[2] Yeomans: Problem 2.1

1.2 Monte Carlo method

It is important to thoroughly understand so-called *Monte Carlo* methods due to their wide range of applicability. Monte Carlo methods are generally computer simulations which help to compute the ensemble average when analytical approaches fail or are too cumbersome — a situation that is often encountered in statistical physics and condensed matter physics. Monte Carlo methods are widely used in science and technology, also in areas far away from lattice models as we study them here, e.g. traffic flow.

Plainly speaking, the aim generally is to obtain an approximation to the expectation value of an observable A, i.e.

$$\langle A \rangle = \frac{\sum_{\{s\}} A \exp(-\beta \mathcal{H})}{\sum_{\{s\}} \exp(-\beta \mathcal{H})} \ .$$

The Ising model in a nutshell. In this text we will repeatedly make use of the spin-$\frac{1}{2}$ Ising model as our "canonical" example. Spin-$\frac{1}{2}$ means that there are two states for each spin. We will discuss the Ising model in more detail in Sec. 1.4, but here briefly introduce the model for the sake of being able to work directly with the Monte Carlo method.
The Ising model is defined as

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - h \sum_i s_i \ ,$$

where s_i can take the values $+1$ or -1 and represents the spin at site i, J is the coupling between neighboring spins and h is an external magnetic field. The bracket specifies that sites i and j only interact if they are nearest neighbors, i.e., the sum is carried out over all possible pairs of neighboring sites. In a two-dimensional square lattice with N sites, there will be $2N$ such pairs to sum over.
For $J > 0$, spins will minimize the energy when aligned (same sign), while for $J < 0$ energy will be lowered if spins are anti-aligned (opposite sign). Similar considerations go for the external magnetic field, which will tend to align spins when sufficiently strong.

Common expectation values are

$$\langle s_i \rangle = \langle s \rangle = \frac{\sum_{\{s\}} s_i \exp(-\beta \mathcal{H})}{\sum_{\{s\}} \exp(-\beta \mathcal{H})} \ ,$$

$$\langle s_i s_j \rangle = \frac{\sum_{\{s\}} s_i s_j \exp(-\beta \mathcal{H})}{\sum_{\{s\}} \exp(-\beta \mathcal{H})} \ ,$$

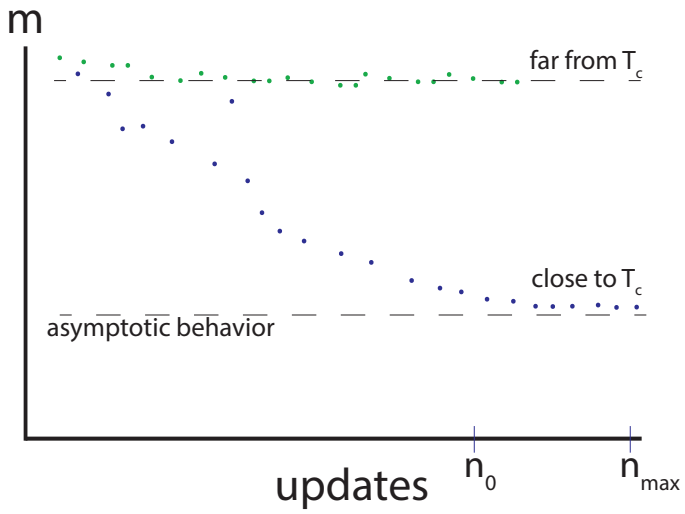


Figure 1.4: **Cartoon of possible Monte Carlo timeseries.** Points show system averages of magnetization m after various system updates for a Monte Carlo simulation with fixed T and parameters J and H, n_0 and n_{max} indicate system updates between which the asymptotic behavior is estimated. A "system update" corresponds to N attempted spin flips, where N is the system size. Far from T_c one expects relatively fast convergence, close to T_c an aspect known as "critical slowing down" causes the convergence time to diverge.

$$\langle \mathcal{H} \rangle = \frac{\sum_{\{s\}} \mathcal{H} \exp(-\beta \mathcal{H})}{\sum_{\{s\}} \exp(-\beta \mathcal{H})} \, .$$

One could hence imagine to simply sum over all configurations and obtain an exact number for the expectation value of interest. However, even on modern-day computers, summing over 2^N configuration for an N-site lattice of more than a few dozen sites is prohibitively costly.

But do we really need to sample the entire space of configurations to get a reasonable estimate of the expectation values? The idea of Monte Carlo techniques is, to sample mainly those configurations that are likely to occur, while ensuring that each state is represented as much as it probabilistically should be. Take intermediate temperatures, where $J/k_BT \sim 1$. Further, take the external field to be absent, i.e. $h = 0$. When all N spins are aligned, the contribution to the partition function is $\exp(Nz/2)$, where z is the coordination number. Conversely, a state where all spins are anti-aligned gives a contribution of $\exp(-Nz/2)$. Notably, the former configuration is $\exp(Nz)$ times more likely that the latter, an enormous number even for modest N.

1.2.1 Importance sampling

Now the idea is to sample phase space by importance. This can be thought of a *Markov chain*, where, in principle, each state of the system can be visited in finite time. However, the probability of entering states of low likelihood will be reduced — to finally yield the proper Boltzmann distribution of weights as it appears in the partition function. The goal is, to achieve the equilibrium distribution

$$P_l^{eq} = \frac{e^{-\beta E_l}}{\sum_m e^{-\beta E_m}} ,$$

when continuing the Markov chain infinitely long. l labels a given spin configuration and P_l^{eq} is the equilibrium probability of this configuration. This requirement puts constraints on the transition probability between states.

The probability to reside in state l at time $t+1$ is

$$P_l(t+1) = P_l(t)\left(1 - \sum_{m\neq l} w_{l\to m}\right) + \sum_{m\neq l} P_m(t) w_{m\to l} , \qquad (1.27)$$

where $w_{i\to j}$ labels the transition probability from configuration i to j and the summations are carried out to include all possible configurations m. It is further useful to define $w_{l\to l} = 1 - \sum_{m\neq l} w_{l\to m}$, i.e. the probability to remain in configuration l. Note also that $\sum_m w_{l\to m} = 1$, hence compactly

$$P_l(t+1) = \sum_m P_m(t) w_{m\to l} .$$

We check that probabilities are normalized correctly:

$$\sum_l P_l(t+1) = \sum_{m,l} P_m(t) w_{m\to l} = \sum_{ml} P_l(t) w_{l\to m} = \sum_l P_l(t) = 1 .$$

For the stationary solution we therefore have

$$P_l(t+1) = P_l(t) ,$$

hence

$$\sum_m [P_l^{eq} w_{l\to m} - P_m^{eq} w_{m\to l}] = 0 . \qquad (1.28)$$

A simple way to achieve the condition in Eq. 1.28 is to ensure that every term vanishes, i.e.

$$P_l^{eq} w_{l\to w} = P_m^{eq} w_{m\to l} ,$$

i.e.

$$\frac{w_{l\to m}}{w_{m\to l}} = \frac{P_m^{eq}}{P_l^{eq}} = e^{-\beta(E_m - E_l)} \equiv e^{-\beta\Delta E} . \qquad (1.29)$$

1.2.2 Requirements for importance sampling

Notably, Eq. 1.29 ensures that probabilities are assigned by the standard Boltzmann weights in the canonical partition function. This condition is referred to as *detailed balance*, which stresses that a balance of probabilities exists between any two states individually. In principle, detailed balance is not required in the Monte Carlo procedure, what *is* strictly required is that equilibrium probabilities are consistent with the Boltzmann distribution. There can be other ways than detailed balance to achieve this, but they are generally much more difficult to prove. A further requirement is *ergodicity*, which in this context means that each state can be visited within infinite time during a Monte Carlo simulation. In practice it is often hard to prove that ergodicity is fulfilled and there are situations where a Monte Carlo simulation can become "trapped" in a local minimum of the free energy.

Mini tutorial: The detailed balance condition (Eq. 1.28) is a sufficient starting point to ensure importance sampling. But is it strictly necessary?

1.2.3 Various options

There are several ways to accomplish this. We therefore label a particular choice of transition probabilities as $a_{l\to m}$, to distinguish from the general $w_{l\to m}$. We make the ansatz that $a_{l\to m}$ should be a function of $e^{-\beta\Delta E}$, i.e.

$$a_{l\to m} = F(e^{-\beta\Delta E}) \,,$$

by symmetry hence

$$a_{m\to l} = F\left(\frac{1}{e^{-\beta\Delta E}}\right) \equiv F\left(\frac{1}{x}\right) \,,$$

with $x \equiv e^{-\beta\Delta E}$.

It results that

$$\frac{a_{l\to m}}{a_{m\to l}} = \frac{F(x)}{F(1/x)} = x \,, \tag{1.30}$$

whereby it must be ensured that $0 \leq F(x) \leq 1$ for meaningful transition probabilities. The choice of $F(x)$ is not unique. Popular choices are

1. $F(x) = \min(x, 1)$, Metropolis algorithm,
2. $F(x) = \frac{x}{1+x}$, heat bath algorithm,
3. $F(x) = \frac{1}{2}\left(1 - \tanh(\beta\Delta E/2)\right)$, Glauber dynamics.

Mini tutorial: As the function $F(x)$, required for obtaining consistency with Boltzmann equilibrium statistics, is not unique, which feature then does change, when a different choice is made for $F(x)$?

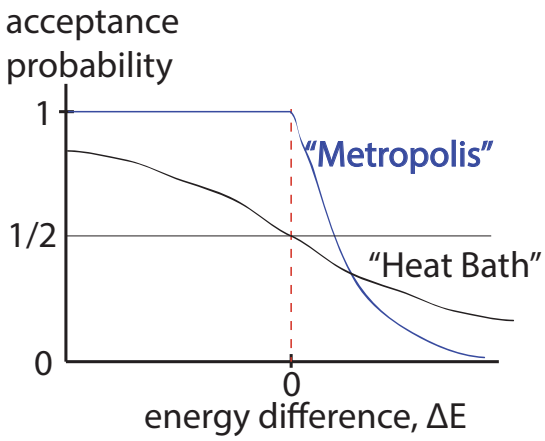


Figure 1.5: **Choices for Monte Carlo update function F(x).**

1.2.4 Practical implementation on a computer.

The basic set-up of a Monte Carlo simulation is rather straightforward:

- set up lattice sites i and spins s_i, define Hamiltonian $\mathcal{H}$, define total number of steps n_{max} and $n_0 < n_{max}$,
- set the system to an initial configuration of spins (either a random configuration or one that is meaningful for the problem of interest),
- flip a spin (note the example in Fig. 1.2.3), compute $r \equiv e^{-\Delta E/k_B T}$, generate random number q between $[0, 1]$, if $r > q$ perform the Monte Carlo move (e.g. flip the spin), otherwise do not,
- compute A_n, i.e. the expectation value of A at time step n, if $n > n_0$, if $n \leq n_0$, we are still in the period considered transient,
- repeat until n_{max} is reached,
- calculate $\langle A \rangle = \frac{1}{n_{max}-n_0} \sum_{n>n_0} A_n$.

Difficulties usually lie in the proper choice of system size N, the choice of the transient period n_0, and the duration of the sampling period n_{max}. Generally, fluctuations increase in the vicinity of T_c and achieving robust results may require increasing both system size and sampling time.

Fig. 1.2.4 shows an example of a Monte Carlo simulation of the spin-$\frac{1}{2}$ Ising model on a 400×400 lattice. The timeseries show the average magnetization $m(t) = N^{-1} \sum_i s_i(t)$ for two different values of temperature. Note that, for the lower temperature ($T = 2.2\ J/k_B$) a relatively steady value of $m(t)$ is approached rapidly. At the higher temperature ($T = 2.3\ J/k_B$), which is very close to $T_c \approx 2.27\ J/k_B$, compare Sec. 1.7.3, the approach is much slower and substantial temporal fluctuations remain. The spatial plots show the state of

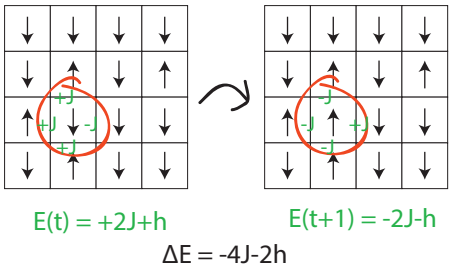


Figure 1.6: **Performing the Monte Carlo step.** Example of a spin flip by which the change of energy is negative for $J > 0$ and $h > 0$. In the Metropolis algorithm this step would always be accepted. Imagine going the opposite direction: The energy difference would be $\Delta E = +4J + 2h$. The probability of accepting this move would then become $\exp -\beta(4J + 2h)$. Note also that ΔE can be computed entirely locally, since the remainder of the lattice maintains its energy. Considering this speeds up the computation enormously.

the system near the end of the simulation (i.e., after $\approx 10^4$ system updates). Note the substantial spatial fluctuations, especially for the higher value of T, and the absence of any typical scale for the clusters shown.

1.2.5 Critical slowing down *

One well-known issue with Monte Carlo simulations is that near the critical temperature T_c the correlation length of the lattice diverges (i.e. $\xi \sim t^{-\nu}$, *see* also the discussion following Eq. 1.35), hence clusters of arbitrary size form. Within these clusters, any Monte Carlo update is unlikely to lead to any long-lasting changes and the overall *correlation time* τ, which for an observable A can be defined as

$$\tau \equiv \frac{\int_0^\infty dt\; t[A(t) - \langle A\rangle]}{\int_0^\infty dt\; [A(t) - \langle A\rangle]}\,, \tag{1.31}$$

can be shown to diverge, i.e. $\tau \sim \xi^z \sim t^{-z\nu}$. The number z is thereby a dynamical critical exponent associated with the observable A, and $\langle A\rangle$ is the equilibrium average of A, i.e. the average of A after the system has reached an equilibrium configuration. Since the autocorrelation time diverges, it becomes more and more time consuming to achieve *statistically independent* samples of the system, as T_c is approached. There are sophisticated methods to (partially) alleviate the problem of statistical slowing down, e.g. cluster update algorithms, which however come with their own set of complications [2, 3]. A useful precaution to at least quantify the required sampling time is to measure the *autocorrelation* $C_A(k)$ as the simulation is running:

$$C_A(t) = \frac{\langle A_n A_{n+k}\rangle - \langle A_n\rangle\langle A_{n+k}\rangle}{\langle A_n^2\rangle - \langle A_n\rangle^2}\,.$$

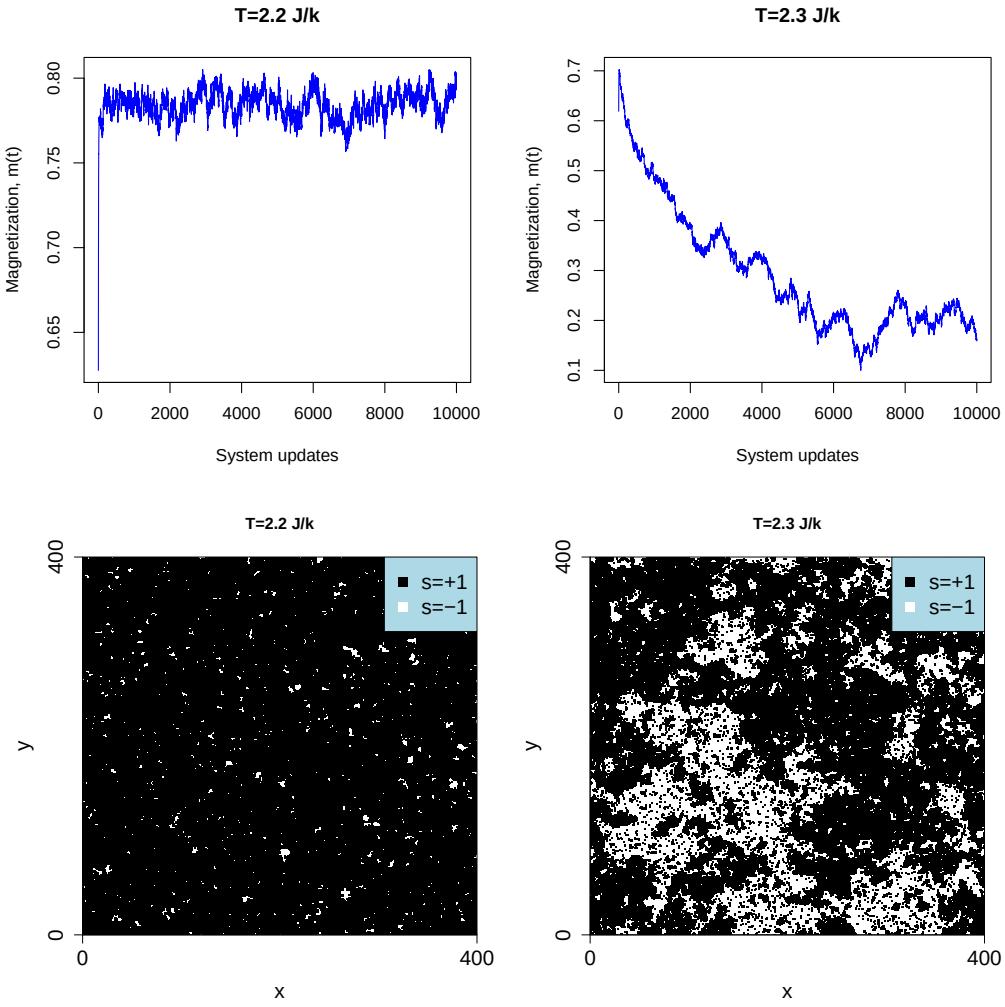


Figure 1.7: **Example of a simulation result.** Computation on a system of 400×400 lattice sites. Timeseries for average magnetization and corresponding spatial spin patterns for two distinct values of T as labeled.

n and k are "times" measured in units of Monte Carlo updates, where one Monte Carlo update represents the attempted update of N sites (N is system size). In general, $C_A(t) \sim \exp(-t/\tau_{auto})$, hence τ_{auto} can be estimated and one can decide how many steps are required for a large enough, independent, sample. More more details on Monte Carlo methods the reader is referred to the literature [3].

1.2.6 Exercises

1A. Monte Carlo Simulation. +
Consider a 2D spin-$\frac{1}{2}$ Ising model, define appropriate neighborhoods for all sites and implement a Monte Carlo method to compute the internal energy and magnetization as a function of temperature. Are you able to find the critical temperature and estimate any critical exponent? Discuss differences for finite/zero magnetic field h. Investigate the robustness of your results, by modifying appropriate parameters of your simulation, and increasing the replicas, i.e. using a set of different initial conditions or random number seeds. Discuss various lattice geometries, in particular, a change of dimension. Make a literature search for available exact results and compare your findings with these. Consider the Wolff cluster update procedure to speed up your simulation near criticality [2].

1B. Monte Carlo Simulation.
Consider a 2D spin-$\frac{1}{2}$ Ising model on a square lattice. Write a computer program (choose your favorite programming language) where you define the sites i and their spins s_i. Make sure to store your program, so that we can build on it later on. Start with a small number of sites, perhaps 20×20. Consider periodic boundary conditions, i.e. each site has four nearest neighbors that are cyclically defined at the boundaries.

1. Define variables for the constants J, h and k_BT as well as the maximum number of spin configurations n_{max}, i.e. the number of configurations the program will sample before terminating. Define also a number $n_0 < n_{max}$ beyond which expectation values should be computed. Define also a function that computes the energy corresponding to a given spin configuration, i.e. $E = -J\sum_{\langle ij\rangle} s_i s_j + h\sum_i s_i$. Further, define a function that flips a (random or deterministic) spin, as well as the exponential $x \equiv \exp(-\Delta E/k_BT)$ with ΔE the energy difference between two states l and m and define the acceptance procedure for a transition $a_{l\to m}$.

2. Make a loop that iterates over the procedure. Starting from a random configuration, carry out the Monte Carlo simulations.

3. For $h = 0$, compute the internal energy $\langle E\rangle$ and the magnetization per site $\langle s\rangle$ as function of k_BT, by evaluating the expectation value for all

$n > n_0$. Plot a timeseries of $\langle E \rangle$ and $|\langle s \rangle|$ as function of n_0. Determine a minimal n_0 so that the expectation values are not affected by the transient behavior. (Note that you may need to make adequate adjustments to n_{max} in this process.)

4. Obtain the expectation values for various temperatures and plot them as function of $k_B T/J$.

5. Repeat the simulation several times for each observable and temperature to obtain a distribution of results for that data point. Use the distribution to quantify the sampling error and plot the error bars.

6. Try to determine $k_B T_c/J$ and β numerically and compare your results to the exact results $\langle s \rangle^8 = 1 - (\sinh 2J/kT)^{-4}$ and $\beta = 1/8$ (Onsager's solution).

7. Plot also the corresponding temperature dependence using a finite magnetic field h.

8. Repeat for a larger system size and make notes of your findings for n_0 and n_{max} and discuss (qualitatively) how these values and the error bars depend on the reduced temperature t.

9. Change your lattice geometry to the 4d spin-$\frac{1}{2}$ nearest-neighbor Ising model, find T_c and plot $\langle s \rangle$ as well as $\langle E \rangle$ vs. $k_B T/J$. Compare to the mean field solution with the corresponding value of coordination number z.

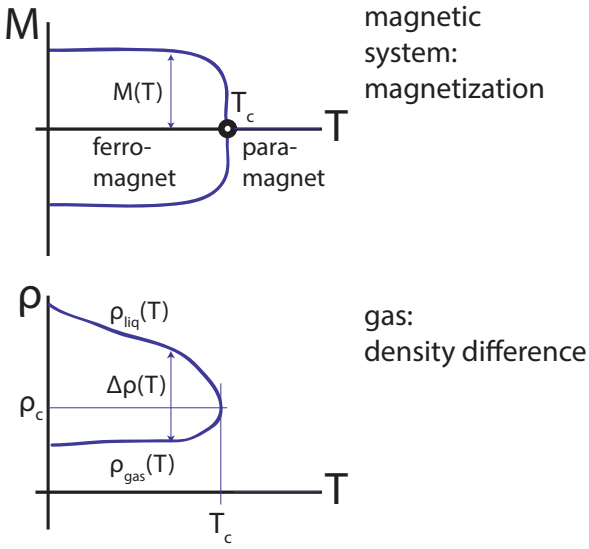


Figure 1.8: **Cartoon of order parameters.** For a zero-external-field ($H = 0$) magnetic system with symmetry regarding spin orientation (top) and a gas, where the order parameter can be defined as the difference in density for the liquid and gas phases.

1.3 Definition of phase transitions

Phase transitions are singularities in the free energy or one of its derivatives. Examples are the liquid-gas transition, the transition from a normal to a superconductor, or the transition from a paramagnet to a ferromagnet.

A phase transition can be measured in terms of an order parameter, which changes with the phase transition. In the case of the liquid-gas transition, the order parameter is the difference in density of the liquid $\rho_{liq}(T)$ and that of the gas $\rho_{gas}(T)$. In magnetic systems, the order parameter is the magnetization $M(T)$ (Fig. 1.8).

In Sec. 1.4 we will analyze the Ising model [3] as a description of magnetism in some magnetic materials. The Ising Hamiltonian is

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - h \sum_i s_i \ , \tag{1.32}$$

where s_i is the "spin" at a lattice site i and can take one of the values ± 1. h is the external magnetic field and J is the *coupling parameter*. For a ferromagnet, which we qualitatively discuss here, $J > 0$, i.e. energy is minimized when spins have the same sign.

[3] First studied by Lenz and Ising in 1925, see Brush [4] and Wolf [5] for reviews on the model and its vast applications.

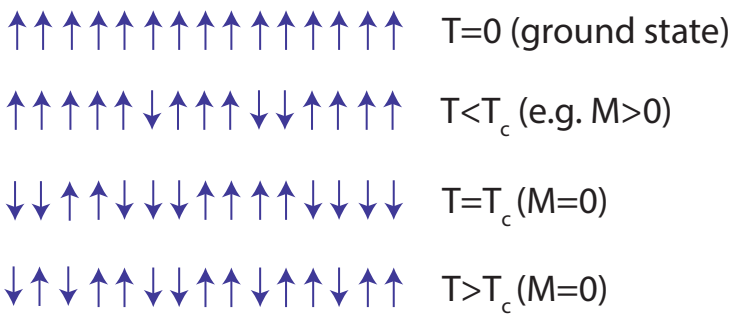


Figure 1.9: **Illustration of spin ordering.** Cartoon of spin ordering for the Ising model for increasing temperature (top to bottom).

We qualitatively sketch some limits (Fig. 1.9): For very high temperature, spins are randomly oriented, all order disappears and there is no net magnetization. As temperature is lowered, the *correlation length* increases, i.e. the length at which spins are correlated and point in the same direction — "patchiness" increases. At the so-called *critical temperature* T_c, the correlation length "diverges", there are patches of correlated spins of all patch sizes. When an external field is absent ($h = 0$) there is however still no net magnetization. As temperature is lowered below T_c, nonzero magnetization emerges spontaneously, i.e. the system breaks the symmetry w.r.t. positive and negative spin, with one orientation dominating randomly. At $T = 0$, all spins are entirely aligned.

Mini tutorial: What is the entropy (per site) for the Ising model in the limit of infinite temperature?

1.3.1 First order vs. continuous phase transition

A first order phase transition in a given quantity is present, when a derivative of the thermodynamic potential has a finite discontinuity. In the case of our magnetic system, the free energy is the appropriate thermodynamic potential, because the particle number is fixed (no particles enter or leave the system) but energy fluctuations are possible, that is, spins are allowed to flip and create fluctuation around the average internal energy $\langle E \rangle$ — we are hence working in the canonical ensemble (see Sec. 1.1.4). For $T < T_c$, there is a line of first order transitions at zero field H, where the free energy shows a kink and magnetization consequently is discontinuous. For $T > T_c$, the free energy is a smooth function of H and magnetization varies continuously. When $T = T_c$, the magnetization varies continuously, hence the transition is *not* of first order. However, the slope at $H = 0$ is infinite, signaling a divergence in the derivative $\partial M/\partial H|_T$, i.e. the isothermal susceptibility. Hence, the transition here is of

second order (continuous transition).

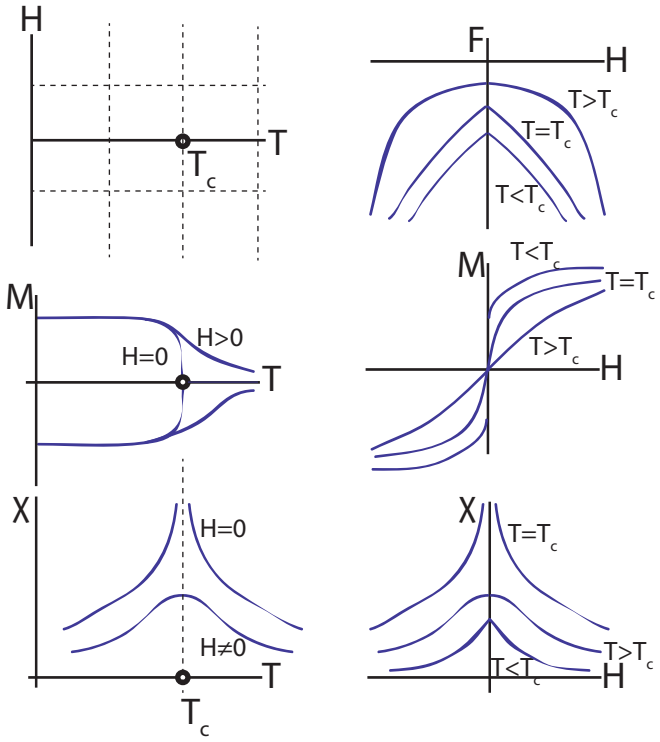


Figure 1.10: Dependencies near the critical point.

Mini tutorial: What would be a zeroth-order phase transition?

1.3.2 Definition of correlation length

While phase transitions are related to the macroscopic properties of a system, we will realize that many of such macroscopic properties are related to the microscopic configuration of the system. One crucial quantity that describes the microscopic state is the spin-spin correlation function. It is defined as

$$\Gamma(\mathbf{r}_i, \mathbf{r}_j) = \langle (s_i - \langle s_i \rangle)(s_j - \langle s_j \rangle) \rangle \ , \tag{1.33}$$

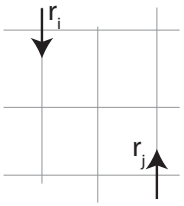
where s_i is the value of spin at lattice position $\mathbf{r}_i$, and $\langle ... \rangle$ denotes the ensemble average.

For translationally invariant systems, $\langle s_i \rangle = \langle s_j \rangle \equiv \langle s \rangle$ and therefore the correlation function only depends on the distance vector between the two spins. It simplifies to

$$\Gamma(\mathbf{r}_i - \mathbf{r}_j) \equiv \Gamma_{ij} = \langle s_i s_j \rangle - \langle s \rangle^2 \ . \tag{1.34}$$

Away from the critical temperature T_c, spins tend to be uncorrelated, i.e.

$$\Gamma(\mathbf{r}) \sim r^{-\tau} \exp(-r/\xi) \ , \tag{1.35}$$



where τ is a number related to the critical exponent η (defined below) and ξ is the correlation length. Away from the critical temperature, the correlation length ξ is finite and the correlation between spins decays exponentially. However, as T_c is approached, the correlation length ξ diverges, i.e. $\xi \to \infty$ and $\exp(-r/\xi) \to 1 = const.$ Indeed, experiments as well as some exactly soluble models show that near criticality $T \to T_c$, the value of the correlation function decays as a power law with distance, i.e.

$$\Gamma(\mathbf{r}) \sim \frac{1}{r^{d-2+\eta}} \, . \tag{1.36}$$

In this equation, η depends on some of the system properties and is a example of a so-called *critical exponent*.

1.3.3 Magnetic susceptibility

One can relate the spin-spin correlation function to the susceptibility, i.e. the fluctuations in magnetization: The magnetic susceptibility at constant temperature is

$$\begin{aligned}
\chi_T &= k_B T \frac{\partial^2}{\partial H^2} \ln Z && (1.37)\\
&= \frac{1}{k_B T} \left(\langle M^2 \rangle - \langle M \rangle^2 \right) && (1.38)\\
&= \frac{1}{k_B T} \langle (M - \langle M \rangle)^2 \rangle && (1.39)\\
&= \frac{1}{k_B T} \langle \sum_i (s_i - \langle s_i \rangle) \sum_j (s_j - \langle s_j \rangle) \rangle && (1.40)\\
&= \frac{1}{k_B T} \sum_{ij} \Gamma_{ij} \, , && (1.41)
\end{aligned}$$

where the total magnetization M was written as the sum over all spins. For the translationally invariant lattice, $\sum_{ij} \Gamma_{ij} = N \sum_j \Gamma_{0j}$, which can be approximated by an integral near criticality, where the lattice structure is unimportant:

$$N \sum_j \Gamma_{0j} \sim N \int dr \; \Gamma(r) r^{d-1} \sim \chi_T \, . \tag{1.42}$$

Universality class	symmetry	α	β	γ	δ	ν	η
2D Ising	2-component scalar	0	1/8	7/4	15	1	1/4
3D Ising	2-component scalar	0.10	0.33	1.24	4.8	0.63	0.04
3D XY	2-dimensional vector	0.01	0.34	1.3	4.8	0.66	0.04
3D Heisenberg	3-dimensional vector	-0.12	0.36	1.39	4.8	0.71	0.04
mean field		0	1/2	1	3	1/2	0
2D Potts, $q = 3$	q-component scalar	1/3	1/9	13/9	14	5/6	4/15

Table 1.1: **Critical exponents for several models.**

Hence, for correlations to remain, one needs to require $\eta < 2$. Overall, divergent susceptibility (a macroscopic quantity) implies divergence also in the fluctuations of magnetization (a microscopic property).

Mini tutorial: Why could it be useful to relate a microscopic material property to a macroscopic one?

1.3.4 Definitions of critical exponents

To measure the deviation from the critical temperature, it is convenient to define the dimensionless quantity

$$t \equiv \frac{T - T_c}{T_c} \,, \tag{1.43}$$

termed "reduced temperature". In terms of t, a critical exponent is generally defined as the limit

$$\lambda = \lim_{t \to 0} \frac{\ln |F(t)|}{\ln |t|} \,, \tag{1.44}$$

or equivalently $F(t) \sim |t|^{\lambda}$.

Commonly used critical exponents are

$$C_H \sim |t|^{-\alpha} \text{ zero-field specific heat} \tag{1.45}$$
$$M \sim (-t)^{\beta} \text{ zero-field magnetization} \tag{1.46}$$
$$\chi_T \sim |t|^{-\gamma} \text{ zero-field isothermal susceptibility} \tag{1.47}$$
$$H \sim |M|^{\delta} sgn(M) \text{ critical isotherm } (t = 0) \tag{1.48}$$
$$\xi \sim |t|^{-\nu} \text{ correlation length} \tag{1.49}$$
$$G(\mathbf{r}) \sim \frac{1}{r^{d-2+\eta}} \text{ pair-correlation function at } T_c \,. \tag{1.50}$$

Why are critical exponents so interesting?
The critical exponents are largely universal, meaning that they depend only on a few fundamental parameters, e.g. the dimensionality of space and the symmetry of the order parameter. That means, that different materials that may differ by several microscopic properties and interactions, can be considered on equal footing. Consider, for example, the fluids studied by Guggenheim in

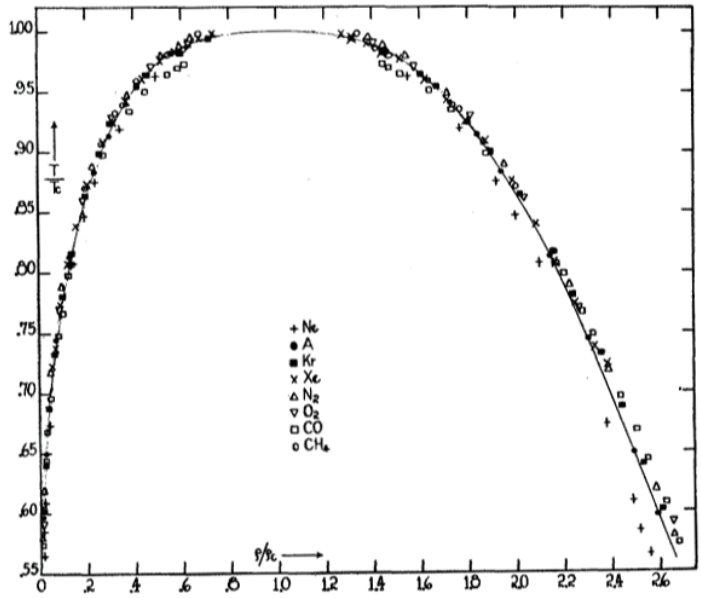


Figure 1.11: **Measurement on eight fluids of the coexistence curve.** The solid line is a fit to a cubic equation, i.e. to the choice $\beta = 1/3$, where $\rho - \rho_c \sim (-\epsilon)^\beta$ [6].

1945 (Fig. 1.11). The diagram shown is the actually measured version of the one in Fig. 1.8. When rescaling temperature by the respective critical temperature of each of the fluids (T/T_c) and similarly for density (ρ/ρ_c), the measurements all collapse on a single line. This line can be well approximated by a cubic equation, which requires only a single critical exponent β. This result shows, that, by knowing about the critical behavior of one of the fluids, one can deduce the behavior of all the others — and this holds also remarkably far away from the critical temperature (*compare*: Fig. 1.11)!

In terms of theoretical modeling, the notion that only the symmetry of the order parameter and the dimensionality of the system matters, this means that very simple models can be chosen to describe the critical behavior of systems, which, a priori, entail much more complicated microscopic interactions. This is why some of the models are so heavily discussed in seemingly unrelated contexts, even though it often seems that they constitute oversimplifications. Concerning critical behavior, this is not so.

Several basic models. For a brief overview, we list several common models along with their critical exponents (Tab. 1.1). Note that the critical exponents for the mean field system can be thought of as corresponding to four dimensional space, i.e. when the exponents for lower dimensions are known,

the mean field exponents in some sense are an "extrapolation" to the next dimension. In fact, in 4D the mean field critical exponents are exact, $d = 4$ is therefore sometimes called the *upper critical dimension*. The mean field approach does not consider the dimensionality of the problem, only the *coordination number* of the lattice, i.e. the number of nearest neighbors, enters into the calculation. The mean field assumption implies that neighboring spins are uncorrelated. This assumption becomes more and more reasonable, when the number of nearest neighbors increases with the dimensionality of the problem. Therefore it seems intuitive that most critical exponents in the 3D Ising model (numerical solution) agree better with the mean field exponents than those of the 2D Ising model.

1.3.5 Exercises

1A. Paramagnet +
A paramagnetic solid contains a large number of non-interacting spin-1/2 particles. This substance is placed in a uniform magnetic field.

Obtain and sketch the magnetization and magnetic response function as well as the entropy of the paramagnet in the field.
Check appropriately chosen limiting behavior for the external variables. Do the limits make physical sense?

1B. Paramagnet[4]
A paramagnetic solid contains a large number N of non-interacting, spin-1/2 particles, each of magnetic moment μ on fixed lattice sites. This substance is placed in a uniform magnetic field H.
(i) Write down an expression for the partition function of the solid, neglecting lattice vibrations, in terms of $x = \mu H/k_B T$.
(ii) Find the magnetization M, the susceptibility χ, and the entropy S, of the paramagnet in the field H.
(iii) Check that your expressions have sensible limiting forms for $x \gg 1$ and $x \ll 1$. Describe the microscopic spin configuration in each of these limits.
(iv) Sketch M, χ, and S as a function of x.
(Answers: (i) $Z = (2\cosh x)^N$, (ii) $M = N\mu \tanh x$, $\chi = N\mu^2/(k_B T \cosh^2 x)$, $S = Nk(\ln 2 + \ln(\cosh x) - x \tanh x)$).

2. Critical Exponents[5] Determine the critical exponents λ for the following functions as $t \to 0$:

- $f(t) = At^{1/2} + Bt^{1/4} + Ct$
- $f(t) = At^{-2/3}(t + B)^{2/3}$
- $f(t) = At^2 e^{-t}$

[4] Yeomans: Problem 2.2
[5] Yeomans: Problems 2.3 and 2.5

- $f(t) = At^2 e^{1/t}$
- $f(t) = A \ln(\exp(1/t^4) - 1)$

Consider a model equation of state that can be written

$$H \sim aM(t + bM^2)^\theta \tag{1.51}$$

where $1 < \theta < 2$, $a, b > 0$ near the critical point. Find the exponents β, γ, and δ and check if they obey the inequality $\gamma \geq \beta(\delta - 1)$ as an equality.

3. Rushbrooke inequality. As the different observables are not independent, also the corresponding critical exponents are related to one another. Consider the specific heats at constant field H and constant magnetization M, respectively:

$$C_H \equiv T \left(\frac{dS}{dT} \right)_H , \tag{1.52}$$

$$C_M \equiv T \left(\frac{dS}{dT} \right)_M , \tag{1.53}$$

as well as the magnetic susceptibility

$$\chi_T \equiv \left(\frac{\partial M}{\partial H} \right)_T . \tag{1.54}$$

Consider now the entropy $S = S(T, H)$ and the total derivative dS. Use the Maxwell relation $\left(\frac{\partial S}{\partial H}\right)_T = \left(\frac{\partial M}{\partial T}\right)_H$ and the chain rule $\left(\frac{\partial z}{\partial x}\right)_y \left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial x}{\partial y}\right)_z = -1$ to obtain a relation between the above observables:

$$\chi_T (C_H - C_M) = T \left(\frac{\partial M}{\partial T} \right)_H^2 . \tag{1.55}$$

Using the definitions of the critical exponents for these observables, verify the Rushbrooke inequality $\alpha + 2\beta + \gamma \geq 2$.

1.4 Definition of the Ising Model

The Ising model represents one example of a lattice model, where one variable is located at each site of a regular grid. The state of the variables is determined by a Hamiltonian. Such models have been successful in the description of critical phenomena, (quantum) magnetism and models for high-temperature superconductivity and phase diagrams, disordered and non-equilibrium systems. With its simplicity, the Ising model is the most heavily studied lattice model in physics. The Ising model can, e.g., be used in illustrating the following topics: phase transitions and critical exponents, mean field theory, series expansion techniques, as well as phenomenological models, such as Landau theory. We will also address the Monte Carlo technique, a numerical method to approximate the dynamics of a many particle system. Finally, the Ising model can be used to discuss the renormalization group procedure.

The Ising model encompasses a lattice of N sites i, each of which contains an object s_i (originally representing the magnetic dipole moment of an atomic spin or simply *spin*), which can be in one of two states that take values ± 1. The Hamiltonian of the Ising model is

$$\mathcal{H} \equiv -\sum_{\langle ij \rangle} J s_i s_j - h \sum_i s_i \,, \tag{1.56}$$

where $\langle ij \rangle$ denotes that a sum is to be carried out over all nearest-neighbor pairs of sites i and j, and J is the *coupling* between these neighboring sites. The quantity h represents an external magnetic field which interacts with the magnetic moment s_i. The magnetization is then defined as the system's macroscopic magnetic moment $M = \sum_i s_i$.

1.4.1 Ferromagnetic and anti-ferromagnetic coupling

It is important to note that the sign of J plays an important role in determining the ordering of the system. For $J > 0$, energy is minimized if all spins align, i.e. neighboring s_i and s_j have the same sign. Such interaction is commonly referred to as *ferromagnetic*, and at low temperatures magnetic order is expected. If $J < 0$, neighboring spins tend to anti-align, in order to minimize energy. Depending on the lattice geometry, at low temperatures, a checkerboard pattern may result, which is referred to as *anti-ferromagnetism*. Notably, when the lattice is not *bipartite*, i.e. in cases where two sites can have a common nearest neighbor, an anti-ferromagnetic coupling ($J < 0$) can cause disordering effects, known as *frustration* (Fig. 1.4.1).

This is most easily exemplified by a triangle with only three spins. For $J > 0$, all spins align and $E = -3J$ for the ground state. There will be two configurations for the ground state, all spins either pointing "up" or "down". For $J < 0$, the situation is more complicated. Out of the three bonds in the triangle, always one will be forced to have spins aligned, i.e. boost the energy along that bond to $+J$, yielding a ground state energy of $-J$. It

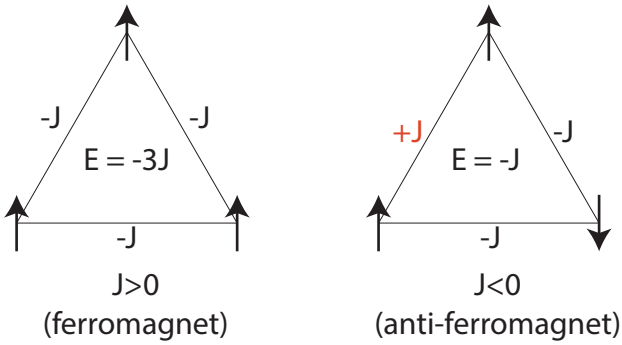


Figure 1.12: **Significance of the sign of coupling.**

is easy to verify that the ground state of such systems is far from unique and the number of states in the ground state increases with system size. In other words, the ground state entropy per site (F/N) is finite for an anti-ferromagnetic triangular lattice, while it is zero for the ferromagnetic case. This is an example of *ground state entropy* (*see* Exercises).

At high temperatures, spins fluctuate thermally and order is generally destroyed. The macroscopic magnetic moment will vanish. This phase is referred to as *paramagnetic phase*. Note that the situation becomes already more complicated, when the lattice is not square, i.e. a simple anti-ferromagnetic order of the checkerboard-type is not possible. Consider a triangular lattice, where two neighboring sites may have a common neighbor. In this case, anti-alignment is not consistently possible, a case referred to as a *frustrated spin system*. We will however focus primarily on the square lattice geometry or one-dimensional systems.

To give an overview, whether an analytical solution exists for the Ising model depends on the dimension of the lattice. In 1D, an analytical solution exists, which we will discuss in Sec. 1.6. In 2D, Lars Onsager in 1944 obtained an analytical solution, which is however very technical. In 3D, no analytical solution exists to date. In 4D, it has been shown that the exact solution is identical to the mean-field solution, which we will discuss in Sec. 1.5.

Mini Tutorial: What do frustration effects do to the ground state of an anti-ferromagnet?

1.4.2 Applications of the Ising model: Exact mapping

Originally, Ising received the model from his Ph.D. supervisor as an exercise, and it was intended as a simple model for magnetism. While this is indeed one (qualitative) application of the model, many others exist.

Lattice gas. One intriguing variant is that of the lattice gas, where particles

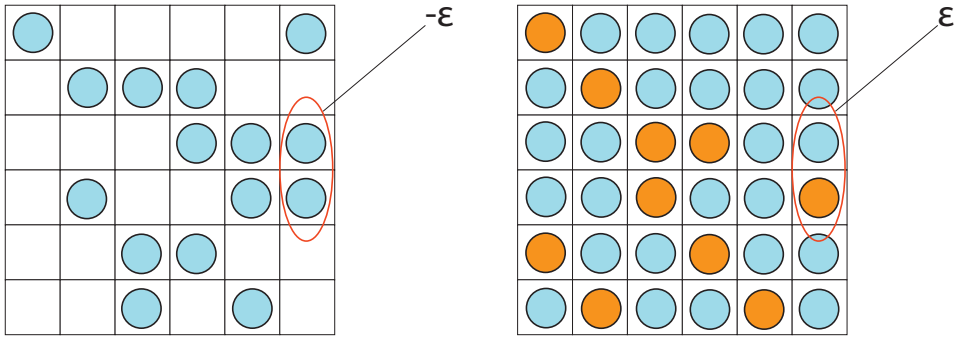


Figure 1.13: **Exact mappings of Ising model.** Lattice gas (left) and incompressible binary mixture of two chemical species (right).

(they might be atoms or molecules) are considered to be located on the sites of a lattice, but the sites can also be empty (Fig. 1.4.2). Further, only nearest neighbors are taken to interact, an interaction which can be thought of as a lowering (or raising) of potential energy, depending on the sign. In short, we have

- states: occupied or empty,
- energy of interaction: $-\epsilon$ if neighboring sites are occupied.

The goal of the lattice gas model is that the occupation density (average number of particles per site) can vary at a fixed number of lattice sites. This means that, as temperature is lowered, one might find that the system spontaneously chooses a particular configuration of density, such that the respective thermodynamic potential is minimized. Since particle number is now *not* fixed, one must consider the Gibbs free energy and the grand canonical partition function (*see* Sec. 1.1.5). This brings in another Lagrange multiplier, namely the chemical potential μ, which is conjugate to the total particle number. We will later see that μ can be associated with the external magnetic field of the Ising model, and that the grand canonical partition function is associated with the canonical one in the case of the Ising model. This means that, below T_c, small changes in chemical potential can bring about a first order transition in density, moving between the "liquid" (i.e. density ρ_{liq}) and the "gas" (i.e. density ρ_{gas}) state abruptly. At temperatures close to T_c but below this value the system will spontaneously collapse to either ρ_{liq} or ρ_{gas}.

The fortunate feature of the lattice gas model is that it maps *exactly* onto the Ising model, hence, finding a solution for one means having *the* solution for the other — there is *no* further approximation. How does this mapping work? For each site ("cell"), we define the spin as $s_i = +1$ (occupied) respectively $s_i =$

-1 (empty). To be able to count particles, we make the unique transformation

$$n_i = (s_i + 1)/2 \ ,$$

where n_i now measures the number of particles at the site, i.e. either 0 or 1. For a lattice of N sites, the total number of particles is

$$N_p = \frac{1}{2}\sum_i (s_i + 1) = \frac{1}{2}\sum_i s_i + \frac{N}{2} \ .$$

The interaction between neighboring sites i and j is

$$\begin{aligned} \epsilon_{ij} &= -\epsilon \text{ if } s_i = s_j = 1 \ , \\ \epsilon_{ij} &= 0 \text{ otherwise.} \end{aligned}$$

This can equivalently be written as

$$\epsilon_{ij} = -\frac{1}{4}\epsilon(s_i + 1)(s_j + 1)$$

and the total energy is

$$E_p = -\frac{\epsilon}{8}\sum_{\langle ij\rangle}(s_i + 1)(s_j + 1) = -\frac{\epsilon}{8}\sum_{\langle ij\rangle} s_i s_j - \frac{z\epsilon}{4}\sum_i s_i - \frac{zN\epsilon}{8} \ .$$

In its original form the lattice gas model requires a grand canonical ensemble, since the total number of particles N_p may be varied (Sec. 1.1.5). The grand partition function then reads

$$Z_G = \sum_{\{s\}} \exp(\beta\mu N_p - \beta E_p) \ . \tag{1.57}$$

Notably, the probability weights in Z_G increase for large values of μ (promoting larger numbers of particles, hence larger density), or by smaller values of the total energy. Re-writing Z_G in the "language" of the Ising model, we obtain an equivalent *canonical* partition function

$$Z_C = \sum_{\{s\}} \exp(-\beta E_{\text{eff}}) \ , \tag{1.58}$$

where

$$E_{\text{eff}} = -\frac{\epsilon}{8}\sum_{\langle ij\rangle} s_i s_j - \left(\frac{\mu}{2} + \frac{z\epsilon}{4}\right)\sum_i s_i - \frac{zN\epsilon}{8} + \frac{\mu N}{2} \ . \tag{1.59}$$

This effective energy now highlights the correspondence between the two models:

- J corresponds to $\frac{\epsilon}{8}$
- h corresponds to $\frac{\mu}{2} + \frac{z\epsilon}{4}$

- M corresponds to density $\rho \equiv N_p/N$
- susceptibility χ corresponds to compressibility α.

Note that the symmetry of the Ising model regarding $\pm h$ is not preserved regarding occupied/unoccupied sites in the lattice gas model, i.e. there is *no* symmetry $\rho \leftrightarrow (1-\rho)$.

Mini Tutorial: What can a lattice gas model teach us about a liquid-gas phase transition (at least, as a metaphor)?

1.4.3 Models related to the Ising model. *

A number of models are related to the Ising model, while they cannot be transformed to *be* the Ising model (*see* Sec. 1.4.2). It is useful to know about these models, to be able to compare them with the Ising model solution, which is often known or more easily available (e.g. by a simple computation).

Potts model.

One simple extension of the Ising model are so-called Potts models, where s_i can take more than two values, but only when neighboring sites have the same value, is the energy value changed, i.e.

$$\epsilon(i,j) = \delta_{(s_i - s_j)} \ ,$$

where δ_x specifies the delta function which is unity when $x = 0$ and zero otherwise. Such Potts models can describe e.g. opinion dynamics in a population, where "agreement" of opinion could cause a negative value of energy.

The spins s_i could also take vectorial values, such that $\mathbf{s}_i \cdot \mathbf{s}_j$ would become the inner product of two vectors. This model is called the Heisenberg model and describes isotropic magnetic moments in a lattice, i.e. moments that are not confined to one of the crystal axes. The Heisenberg model can also be applied to quantum spins, such that the vectors are interpreted as quantum spin operators $\hat{\mathbf{s}}_i$.

Often, more complex lattices are introduced with non-trivial unit cells, e.g. fcc, bcc, or hexagonal lattices. These introduce further complications, which are often necessary when describing metals more quantitatively. Realistic descriptions generally also demand inclusion of further-neighbor interactions, beyond the range of nearest neighbors.

When the coupling parameter J is made entirely random, then so-called "glassy" materials can be described, e.g. spin-glasses. Also time-dependent models are possible, leading altogether away from equilibrium statistical physics.

Ising-like model for volatility.

The volatility clustering can be captured by a herding model inspired by the cartoon in Fig. 1.4.3, and rephrased in terms of an Ising like model at under-critical temperature supplemented by a one-parameter coupling to the absolute value of the total magnetization (by S. Bornholdt, Int. Mod. Phys. C 12 (2001) page 667). The overall philosophy is a system of traders that tend to

1. Copy their friends: Buy if your friends buy, sell if your friends sell.

2. If the global trend is strong, then sell or buy randomly.

This is formulated in terms of an Ising-like model with local interactions and coupling to global magnetization:

$$h_i = \left(\sum_j s_j \right) - \alpha \cdot s_i |M(t)| \ , \tag{1.60}$$

where

$$m(t) = \frac{1}{N} \sum_k s_k \tag{1.61}$$

is the average magnetization at time t.

The dynamics proceeds as *heat bath dynamics* where one selects a random site i and sets

$$s_i(t+1) = +1 \tag{1.62}$$

with probability given by the equilibrium expectation

$$p = \frac{\exp\left(\beta \cdot h_i\right)}{\exp\left(\beta \cdot h_i\right) + \exp\left(-\beta \cdot h_i\right)} = \frac{1}{1 + \exp\left(-2\beta \cdot h_i\right)} \tag{1.63}$$

Figure 1.14: **Cartoon of herding behavior.** A market driven by herding can lead to large volatility (Kalton, in the Economist).

·and otherwise sets $s_i(t+1) = -1$. The model is always considered for subcritical β where the spins tend to align for $\alpha = 0$. In absence of global coupling it will accordingly give large positive or large negative magnetization. With global coupling $\alpha > 0$, switch to negative spins will be favored if the spin is positive ($s_i > 0$). The switching probability for this opposing move to occur is larger if the absolute average spin $|m(t)|$ is large. The $\alpha = 0$ version of the model is the two-dimensional Ising model. For larger α the coupling to the total magnetization makes individual spin tend to take values opposite to itself (not to the overall magnetization). Thus if magnetization deviates substantially from zero, the individual agents tend to shift all the time, introducing an increased volatility in the market that slowly tends to drive the market back

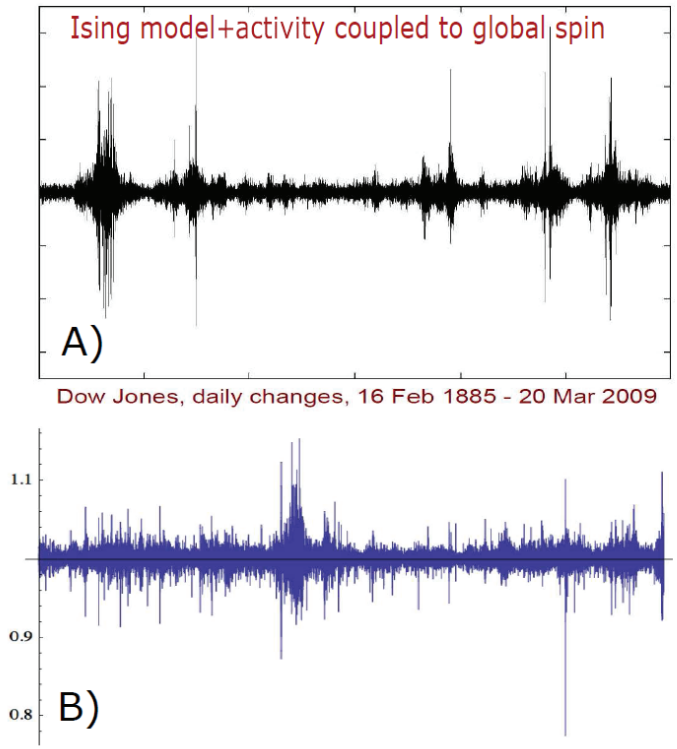


Figure 1.15: **Market model for volatility.** A) which shows $M(t+1) - M(t)$ and the volatility it is supposed to reproduce shown in B) illustrating daily returns, $(S(t) - S(t-1))/S(t-1)$, for the Dow Jones stock market index. Fluctuations are correlated: When variations on one day is large, then it most likely is large again next day [7]. The directions of these fluctuations are uncorrelated! Volatility clustering is sometimes also discussed in terms of the GARCH model (Tim Bollerslev 1986).

to $m \approx 0$. The dynamics of overall volatility measured as $m(t+1) - m(t)$ is shown in Fig. 6.12. The magnetization itself looks like in Fig. 6.13, which does not have an obvious analogy in the market. The connection to a price is explored in the extended model of Kaizoji [8].

1.4.4 Exercises

1. Ground state for simple models. The ground state of a system (stable state at $T = 0$) often serves as the starting point for finite temperature investigations, e.g. the low-temperature expansion technique (later in the course). This is because it can dominate the partition function, even at $T > 0$. It is therefore important to develop some intuition for the ground state of simple models. Find the ground state for the following systems:
(i) The 1d Ising model with first and second neighbor interactions

$$\mathcal{H} = -J_1 \sum_i s_i s_{i+1} - J_2 \sum_i s_i s_{i+2} \ , \tag{1.64}$$

where both positive and negative values of the exchange parameters should be considered.
(ii) The 1d p-state chiral clock model

$$\mathcal{H} = -J \sum_i \cos(2\pi(n_i - n_j + \Delta)/p) \tag{1.65}$$

for $J > 0$ and all values of Δ.
(iii) For the antiferromagnetic, zero-field spin-1/2 Ising model on a *triangular* lattice

$$\mathcal{H} = J \sum_{\langle ij \rangle} s_i s_j \tag{1.66}$$

with $J > 0$, find the ground state energy and a possible representation of it.

2. Lattice binary mixture. This is a model for an incompressible mixture of chemical species A and B (Fig. 1.4.2). In this case, the total number of particles is fixed, but the difference of particles of types A and B may vary, i.e. $N_A - N_B$. One option for defining the energy is that nearest-neighbor contacts of similar species, i.e. AA or BB contribute zero energy while those of different species, i.e. AB, give a contribution ϵ. Species would then attempt to avoid mixing, something that might be observed when oil and water are brought into the same volume. Similar to the lattice gas, we could define a cell i occupied by A to have $s_i = +1$ and those with B as $s_i = -1$.
(a) Write down the energy of interaction ϵ_{ij} between nearest neighbors as well as the total energy E_p.
(b) In terms of the mixing ratio of particle types A and B, write down the "grand" partition function Z_G for the system. The quotation marks are used, since not the total particle number but the difference of particle numbers may be varied. By defining an effective energy E_{eff}, show that the mapping of

parameters $J \to \frac{\epsilon}{4}$ and $h \to \Delta\mu$ maps the lattice binary mixture model onto the standard Ising model (Eq. 1.56). Interpret the mapping in terms of the mixing energy and chemical potential.

3. Volatility model.
a) In the model in Sec. 1.4.3 we use the heat bath method, therefore repeat a simulation of the Ising model for a 10 times 10 system as function of inverse temperature β and plot the energy and average magnetization as function of of temperature. Confirm that it works. (b) Simulate the above Ising inspired model for volatility in a market model using an $N = 10 \times 10$ system with $\beta = 0.7$ and $\alpha = 1$, respectively $\alpha = 2$ and 5. Confirm that: The volatile periods are associated to periods where $M(t)$ is high (when $\alpha \gg 1$).

4. Self-organized criticality.
In self-organized criticality, the system itself controls the tuning parameter, that is, for the case of the Ising model, the internal state will feed back onto the temperature. This feedback is performed in a way such that the system always remains close to the critical value of temperature. Can you find a way to modify your Monte Carlo simulation of the Ising model, such that the system becomes self-organized critical?

1.5 Mean field solution

In two-dimensional lattices, e.g. the square or triangular lattice, an exact solution to the Ising model, that is, an expression for the free energy, is difficult to obtain. In three dimensions, no exact solution is even known to date. In one dimension, where spins are oriented along a line, a relatively simple exact solution does exist and will be discussed in the subsequent chapter (Sec. 1.6). However, various approximate approaches to the solution of the Ising model exist. Some of these are termed "mean field" solutions. In essence, mean field theory builds on the assumption, that the surroundings of each spin, or more generally each particle, act as a joint "field" on this particle, thereby, the individual correlations to each surrounding site are ignored (Fig. 1.5).

In the following, we give two approaches.

1.5.1 Intuitive approach

While more elegant and general, the use of a trial Hamiltonian (Sec. 1.5.5) comes with a more technical nature of the derivation. A simple "hands on" approach is the following: Start with the Ising Hamiltonian

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - h \sum_i s_i \, . \tag{1.67}$$

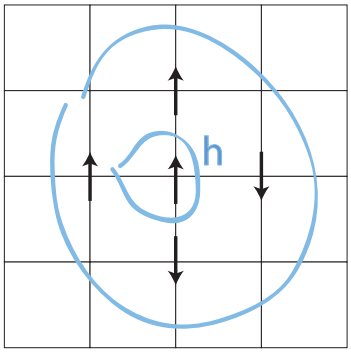


Figure 1.16: **Cartoon of the mean field assumption.** The surroundings of a given site act as an effective "mean field" h, thereby any correlations between individual sites are ignored. The equation is closed by assuming that the site itself contributes to the mean field seen by its neighbors. By this assumption, a self-consistent equation is obtained for the field h, which is proportional to the magnetization of each spin m.

The average magnetization per site is

$$m = \frac{1}{N}\sum_{j=1}^{N}\langle s_j\rangle \ .$$

We can re-write the spin of each particle relative to this average as

$$s_i = m + (s_i - m)$$

and obtain for the product of spins in Eq. 1.67

$$\begin{aligned} s_i s_j &= (m + (s_i - m))(m + (s_j - m)) \\ &= m^2 + m(s_i - m) + m(s_j - m) + (s_i - m)(s_j - m) \ . \end{aligned} \quad (1.68)$$

Assuming that fluctuations are small (generally a poor assumptions), we neglect the second-order fluctuation term (4th term in Eq. 1.68). This is the basis of the mean field assumption, i.e. that products of spins with spins can be ignored and only products of spins and a "mean field" need to be taken into account. This assumption then yields the mean field energy

$$E_{MF} = -J\sum_{\langle ij\rangle}(-m^2 + m(s_i + s_j)) - h\sum_i s_i \ . \quad (1.69)$$

We have hence replaced the (microscopic) interaction between each spin and each neighbor by an *average* magnetic field, produced jointly by all the neighbors. Eq. 1.69 can be simplified by noting that $-J\sum_{\langle ij\rangle}(-m^2) = \frac{JNz}{2}m^2$, where z is the coordination number, i.e. number of nearest neighbors, of the lattice; further, $\sum_{\langle ij\rangle}(s_i + s_j) = z\sum_j s_j$, since all spins are equivalent. The mean field energy then is

$$E_{MF} = \frac{JNz}{2}m^2 - (Jzm + h)\sum_{j=1}^{N} s_j \ .$$

Mini Tutorial: Consider the mean field energy above and discuss its dependence on dimensionality and geometry of the lattice.

1.5.2 Mean field partition function and critical temperature

We can now write the mean field partition function as

$$\begin{aligned} Z_{MF} &= \sum_{\{s_j\}} \exp\left(-\beta E_{MF}\right) \qquad (1.70) \\ &= \exp\left(-\beta \frac{NJzm^2}{2}\right) \sum_{s_1} \cdots \sum_{s_N} \prod_{j=1}^{N} \exp(\beta(Jzm+h)s_j) \\ &= \exp\left(-\beta \frac{NJzm^2}{2}\right) \left[\sum_{s_1} \exp\beta(Jzm+h)s_1\right] \cdots \left[\sum_{s_N} \exp\beta(Jzm+h)s_N\right] \\ &= \exp\left(-\beta \frac{NJzm^2}{2}\right) \left[2\cosh(Jzm\beta+h\beta)\right]^N \ . \qquad (1.71) \end{aligned}$$

It is now straightforward to compute the magnetization per site:

$$m = \frac{1}{N}\sum_{j=1}^{N} \langle s_j \rangle = \langle s \rangle \ ,$$

where translational invariance was assumed. Hence

$$\begin{aligned} m &= \frac{1}{Z_{MF}} \exp\left(-\beta \frac{NJzm^2}{2}\right) \left[2\cosh(Jzm\beta+h\beta)\right]^{N-1} \left[2\sinh(Jzm\beta+h\beta)\right] \\ &= \tanh(Jzm\beta+h\beta) \ , \qquad (1.72) \end{aligned}$$

which yields, for $h = 0$, a self-consistent equation for m (Fig. 1.5.2):

$$m = \tanh(Jzm\beta) \ . \qquad (1.73)$$

Inspecting the plot (Fig. 1.5.2), we note that this equation can either have one or three solutions. In the case of a single solution, only $m = 0$ is possible. We further note that the transition to the case of three solutions is dependent on the value of $Jz\beta$. Considering that J and z are constants, but β can be varied, we ask, at which value of β the transition occurs. This is easy to find when noting that the slopes of the two curves shown must align for this specific β. Hence, we demand that

$$\frac{dm}{dm}|_{m=0} = 1 = \frac{d\tanh(Jzm\beta)}{dm}|_{m=0} = Jz\beta + \mathcal{O}(m^2) \ .$$

A transition between the single and three solution case will hence occur at the *critical* $\beta_c \equiv (Jz)^{-1}$. Alternatively, the mean field critical temperature is

$$T_c \equiv Jz/k_B \ .$$

Notably, the critical temperature T_c increases with coupling J and the number of neighbors z, but does not depend on the dimension or the geometry of the

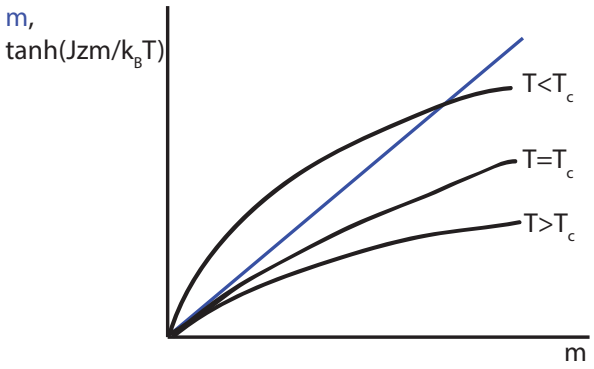


Figure 1.17: **Cartoon of the self-consistency condition.** Schematic shows the curves of $\tanh(Jzm/k_BT)$ for various values of T as well as the curve m.

lattice. This is intuitively reasonable, since our original assumptions leading to the mean-field description did not make any reference to the lattice geometry or the dimensionality.

Mini Tutorial: There are two solutions to Eq. 1.73, where $h = 0$. What is the difference between the two? What would happen to these two, if $h \neq 0$?

1.5.3 Mean field free energy

Now that T_c is obtained, we can proceed to evaluate several critical exponents. Knowing the free energy, all observables can be obtained as derivatives (Fig. **??**). Since these require knowledge of physical observables near T_c, e.g. C_H, M, or χ_T, we first need to work out the free energy in the vicinity of T_c, hence for small values of the reduced temperature

$$t \equiv \frac{T - T_c}{T_c} \ . \tag{1.74}$$

The mean field free energy can be computed from Eq. 1.71 by taking the logarithm of the mean field partition function Z_{MF}, yielding

$$f_{MF} = -\frac{k_B T}{N} \ln Z_{MF} = \frac{Jzm^2}{2} - k_B T \ln\left[2\cosh(Jzm\beta + h\beta)\right] \ . \tag{1.75}$$

It is now useful to work in dimensionless units by dividing Eq. 1.75 through by zJ and introducing the rescaled quantities $h' \equiv h/zJ$ as well as $\theta = T/T_c = k_BT/Jz$. For later use, note that in these units $t = \theta - 1$. We then have

$$\frac{f_{MF}}{zJ} = \frac{m^2}{2} - \theta \ln 2 - \theta \ln\cosh\left(\frac{m + h'}{\theta}\right) \ . \tag{1.76}$$

Mini Tutorial: Discuss the symmetry of the free energy f_{MF}.

1.5.4 Mean field critical exponents

In the following, we will show how to compute the critical exponents for magnetization ($m \sim (-t)^\beta$), specific heat and ($c_H \sim |t|^{-\alpha}$) and susceptibility ($\chi \sim |t|^{-\gamma}$) from the mean field free energy.

β describes the temperature dependence of magnetization $m \sim (-t)^\beta$ in the vicinity of T_C, i.e. in the limit of $t \to 0$. Consider what we have: the mean field free energy is presently a function of temperature θ and magnetization m. The latter, in turn, depends on temperature. Keeping this in mind, we nonetheless want to find a condition for when the free energy is minimized through finite values of m, i.e. when the system "decides" to break the symmetry w.r.t. the two spin configurations.

As will be discussed further (Sec. 1.5.6), a fourth order polynomial in m is a reasonable starting point for observing such a transition. Even without knowing this, one might be tempted to expand $\cosh(m/\theta)$ in Eq. 1.76 in a Taylor series, yielding

$$\frac{f_{MF}}{zJ} = \frac{m^2}{2} - \theta \ln 2 - \theta \ln \left[1 + \frac{1}{2}\frac{m^2}{\theta^2} + \frac{1}{24}\frac{m^4}{\theta^4} + \mathcal{O}\left(\frac{m^6}{\theta^6}\right) \right] . \tag{1.77}$$

Using the additional expansion $\ln(1+x) = x - \frac{x^2}{2} + \ldots$ the final logarithm in Eq. 1.77 simplifies to yield

$$\begin{aligned} \frac{f_{MF}}{zJ} &= \frac{m^2}{2} - \theta \ln 2 - \theta \left[\frac{1}{2}\frac{m^2}{\theta^2} - \frac{1}{12}\frac{m^4}{\theta^4} + \mathcal{O}\left(\frac{m^6}{\theta^6}\right) \right] \\ &= \frac{m^2}{2}\left(1 - \frac{1}{\theta}\right) - \theta \ln 2 + \frac{m^4}{12\theta^3} + \mathcal{O}\left(\frac{m^6}{\theta^6}\right) . \end{aligned} \tag{1.78}$$

We are now in a position to ask for extrema of f_{MF} regarding m, which require that we set the derivative $\partial f_{MF}/\partial m = 0$, hence

$$\frac{1}{zJ}\frac{\partial f_{MF}}{\partial m} = m\left(1 - \frac{1}{\theta}\right) + \frac{1}{3}\frac{m^3}{\theta^3} = 0 . \tag{1.79}$$

For later use we also compute the second derivative w.r.t. m, namely:

$$\frac{1}{zJ}\frac{\partial^2 f_{MF}}{\partial m^2} = \left(1 - \frac{1}{\theta}\right) + \frac{m^2}{\theta^3} . \tag{1.80}$$

Apart from the solution $m = 0$ Eq. 1.79 leads to

$$m^2 = 3(1-\theta)\theta^2 = 3(-t)\theta^2 , \tag{1.81}$$

where we have introduced the reduced temperature $t \equiv (T - T_c)/T_c$. Hence, for small $|t|$ but $t < 0$ the Eq. 1.81 has the two solutions $m(t) = \pm|t|^{1/2}\sqrt{3}\theta$, while for $t > 0$ there are no real solutions. We now check the second derivative

(Eq. 1.80) for any of the real solutions, and find that for $t < 0$ the second derivative is positive for nonzero m and negative for $m = 0$, while for $t > 0$ the only solution $m = 0$ gives a positive value of the second derivative. This finding indicates that minima of f_{MF} are expected for nonzero average magnetization for $t < 0$ and vanishing magnetization for $t > 0$.

Mean-field critical exponent for zero-field magnitization: $m \sim (-t)^{\beta}$. Note that θ^2 in the vicinity of T_c acts as a constant, since the relevant variations occur in the variable t ($1 - \theta$, not θ, is the small quantity). Eq. 1.81 delivers the mean field critical exponent $\beta = 1/2$, and further allows us to substitute m^2 back into the free energy f_{MF}. This yields

$$\frac{f_{MF}}{zJ} = -\frac{3}{2}(1-\theta)^2 - \theta \ln 2 + \frac{3}{4}\theta(1-\theta)^2 + \text{higher order terms}$$

Substituting "solitary" appearances of θ by unity (we are very close to T_c), we have

$$\frac{f_{MF}}{zJ} = -\frac{3}{4}(1-\theta)^2 - \theta \ln 2 + \text{higher order terms} ,$$

and can check that, with $t < 0$ the free energy is actually reduced as compared to the symmetric choice of $m = 0$ — hence, free energy is minimized, not maximized for our choice of m^2.

Mean-field critical exponent for zero-field specific heat: $C \sim |t|^{-\alpha}$. It is also possible to obtain the exponent α corresponding to the mean field specific heat c_H (in the absence of a magnetic field). We will find that c_H has a jump discontinuity (first order transition) as T_c is crossed. Since we have already computed the mean field free energy f_{MF}, we are now in a position to proceed with entropy S and specific heat c_H as derivatives of the free energy, i.e.

$$S = -\frac{\partial f_{MF}}{\partial T} = -\frac{\partial f_{MF}}{\partial \theta}\frac{\partial \theta}{\partial T} = -k_B\left(\frac{3}{2}(1-\theta) + \ln 2\right) .$$

The specific heat becomes

$$c_H = T\frac{\partial S}{\partial T} = \theta\frac{\partial S}{\partial \theta} = \frac{3}{2}k_B\theta .$$

Taking the limit $T \to T_c$ is now simply the statement $\theta = 1$, hence $c_H = \frac{3}{2}k_B$. Since $c_H = \text{const}$, the specific heat critical exponent $\alpha_{MF} = 0$.

For $T > T_c$, $m = 0$ and the paramagnetic free energy only depends linearly on temperature ($f_{MF} = -k_B T \ln 2$), yielding constant entropy $S_{para} = k_B \ln 2$. Hence, also for $T > T_c$ the specific heat critical exponent $\alpha_{MF} = 0$.

Mean-field criticial exponent for the isothermal susceptibility: $\chi_T \sim |t|^{-\gamma}$. To compute the susceptibility

$$\chi = \frac{\partial m}{\partial h}|_{h=0} ,$$

one needs to work with the linear response of magnetization to changes in the external magnetic field. Since only infinitesimal perturbations by h are

required, linear order in h is sufficient. However, for $T < T_c$, magnetization is finite and one needs to consider sufficient order in m. We return to the self-consistency condition given by Eq. 1.72, namely

$$m = \tanh \beta(Jzm + h) \tag{1.82}$$

First, let us Taylor expand Eq. 1.82 to linear order in h, to yield

$$m = \tanh(\beta Jzm) + h\beta(1 - \tanh^2(\beta Jzm)) , \tag{1.83}$$

and expand the hyperbolic tangent out as:

$$\tanh(x) = x - \frac{x^3}{3} + \mathscr{O}(x^5) . \tag{1.84}$$

where the third order term ensures that finite magnetization is possible. Remember now that m vanishes for $T > T_c$, whereas it is finite for $T < T_c$. It is thus important to distinguish the former (super-critical) case from the latter (sub-critical) one.

Sub-critical case: $T < T_c$. To third order in m, by inserting into Eq.1.83 , we obtain

$$m = \beta Jzm - \frac{1}{3}(\beta Jzm)^3 + h\beta - h\beta(\beta Jzm)^2 . \tag{1.85}$$

When applying the derivative w.r.t. h on both sides of the equation and taking the limit $h \to 0$, we have

$$\chi = \frac{\partial m}{\partial h}|_{h=0} \tag{1.86}$$

$$= \beta Jz\chi - (\beta Jz)^3 m^2 \chi + \beta - \beta(\beta Jz)^2 m^2 \tag{1.87}$$

Re-arranging and inserting $m^2 = -3t$ we have

$$\chi \left(1 - \frac{T_c}{T} - \left(\frac{T_c}{T}\right)^3 3t\right) = \frac{1}{kT}\left(1 + 3\left(\frac{T_c}{T}\right)^2 t\right) . \tag{1.88}$$

If we now express occurrences of T and T_c in terms of t, and consider the limit $t \to 0$, $t < 0$, we finally end up with

$$\lim_{t\to 0, t<0} \chi = \frac{1}{k_B T}\frac{1}{(-2t)} . \tag{1.89}$$

Super-critical case: $T > T_c$. In Eq. 1.87 it is now appropriate to disregard the terms involving m^2, since, after taking the derivative regarding h, the limit $h \to 0$ is taken and m takes its zero-field value, namely $m = 0$. Hence, we then have

$$\chi \left(1 - \frac{T_c}{T}\right) = \frac{1}{k_B T} ,$$

which yields

$$\lim_{t\to 0, t>0} \chi = \frac{1}{k_B T_c}\frac{1}{t} = \frac{1}{Jzt} \ . \tag{1.90}$$

From Eqs 1.89 and 1.90 it is now easy to read off the critical exponent γ corresponding to the temperature dependence of χ near the critical temperature. In both cases,

$$\gamma = -1 \ ,$$

hence, χ indeed diverges at $T = T_c$ and the scaling for χ near the critical temperature does not depend on the sign of t.

Mini Tutorial: In how far would the results obtained above differ, if the lattice was a 1D ring?

1.5.5 Using a trial Hamiltonian (less intuitive, more general) *

We here want to approximate the solution to the zero-external-field Ising Hamiltonian. In mean field theory, it is assumed that each spin of the system interacts with its surroundings, but the surroundings represent a "mean field", i.e. there is no explicit interaction with each nearest-neighbor site, but it is assumed that all nearest-neighbor sites produce a joint field, which is the same for each site of the system. The Hamiltonian we want to solve is

$$\mathcal{H} = -J\sum_{ij} s_i s_j \ , \tag{1.91}$$

where a ferromagnetic coupling $J > 0$ is taken.

We use the Bogoliubov inequality, which states that

$$F \leq \Phi = F_0 + \langle \mathcal{H} - \mathcal{H}_0 \rangle_0 \ , \tag{1.92}$$

where F is the true free energy, F_0 is the free energy obtained from a trial Hamiltonian $\mathcal{H}_0$ and $\langle ... \rangle_0$ denotes the expectation value computed in the ensemble defined by $\mathcal{H}_0$. The trial Hamiltonian $\mathcal{H}_0$ thereby depends on a parameter h_0, which is then used to minimize Φ. A common choice is the take $\mathcal{H}_0$ as the free Hamiltonian, i.e.

$$\mathcal{H}_0 \equiv -h_0 \sum_i s_i \ .$$

The trial free energy is (*see* exercises)

$$F_0 = -Nk_BT \ln(2\cosh\ \beta h_0) \ ,$$

and the expectation value of spin is

$$\langle s \rangle_0 = \tanh\ \beta h_0 \ .$$

We now evaluate the expectation value in Eq. 1.92, yielding

$$\begin{aligned}\langle \mathcal{H} - \mathcal{H}_0 \rangle_0 &= \frac{\sum_{\{s\}} \left(-J \sum_{\langle ij \rangle} s_i s_j + h_0 \sum_i s_i \right) exp(\beta h_0 \sum_i s_i)}{\sum_{\{s\}} \exp(\beta h_0 \sum_i s_i)} \\ &= -J \sum_{\langle ij \rangle} \langle s_i \rangle_0 \langle s_j \rangle_0 + h_0 \sum_i \langle s_i \rangle_0 \ ,\end{aligned}$$

note the difference between the symbols $\mathcal{H}_0$ and h_0. Using translational invariance of the lattice (all sites are equivalent),

$$\langle s \rangle_0 \equiv \langle s_i \rangle_0 = \langle s_j \rangle_0 \ ,$$

we have

$$\langle \mathcal{H} - \mathcal{H}_0 \rangle_0 = -JzN \langle s \rangle_0^2 / 2 + N h_0 \langle s \rangle_0 \ .$$

The approximate free energy then is

$$\begin{aligned}\Phi &= -N k_B T \ln(2 \cosh \beta h_0) - JzN \langle s \rangle_0^2 / 2 + N h_0 \langle s \rangle_0 \\ &= -N k_B T \ln(2 \cosh \beta h_0) - \frac{JzN}{2} \tanh^2 \beta h_0 + N h_0 \tanh \beta h_0 \ .\end{aligned}$$

Minimizing w.r.t. h_0 we obtain

$$\begin{aligned}\frac{d\Phi}{dh_0} &= N k_B T \tanh \beta h_0 - \frac{JzN}{k_B T} \frac{\tanh \beta h_0}{cosh^2 \beta h_0} + \frac{N}{k_B T} h_0 \frac{1}{\cosh^2 \beta h_0} + N \tanh \beta h_0 \\ &= \frac{N}{k_B T \cosh^2 \beta h_0} \left(h_0 - Jz \tanh \beta h_0 \right) \ .\end{aligned}$$

Minimization requires the last factor to vanish, i.e. $h_0 = Jz \langle s \rangle_0$. This gives a condition for the mean field magnetization (*compare* Fig. 1.5.2):

$$\langle s \rangle_0 = \tanh(\beta Jz \langle s \rangle_0) \ . \tag{1.93}$$

Inserting this into the approximate free energy Φ yields the mean field free energy:

$$\Phi_{mf} = -N k_B T \ln(2 \cosh \beta Jz \langle s \rangle_0) + \frac{JzN}{2} \langle s \rangle_0^2 \ . \tag{1.94}$$

Finding the critical temperature T_c. Recall that the critical temperature is defined as the temperature, where a transition from a ferromagnetic to a paramagnetic phase is observed. The expression for the mean field magnetization (Eq. 1.93) implicitly defines $\langle s \rangle_0$, even though an analytical solution does not exist. However, one does not need an explicit expression for $\langle s \rangle_0$, if one only is interested in the transition temperature T_c, i.e. where magnetization just barely becomes finite.

A practical way to obtain this transition is to just plot both sides of Eq. 1.93 as a function of $\langle s \rangle_0$. The LHS just gives a straight line of slope unity, while the RHS gives a monotonically increasing concave function, however, the slope depends on temperature. The concavity guarantees that, if the slope is less

than unity at $\langle s\rangle_0 = 0$, there will be no further intersections for positive (or negative, by symmetry) magnetization.

All we need to do is hence to look for an argument of the tanh that gives

$$\lim_{\langle s\rangle_0 \to 0} \frac{\partial \tanh(\beta J z \langle s\rangle_0)}{\partial \langle s\rangle_0} \approx \beta J z = 1 \ ,$$

i.e. the critical temperature T_c becomes

$$T_c = \frac{Jz}{k_B} \ ,$$

a quantity that notably only depends on the coordination number z, that is, the number of nearest neighbors of each site, but not on the dimensionality of the lattice.

Now that T_c is known, we can define the dimensionless temperature

$$t \equiv \frac{T - T_c}{T_c} \ ,$$

i.e. $T = T_c(t+1) = \frac{Jz}{k_B}(t+1)$.

How does magnetization scale as we approach the critical point, i.e. what is the exponent β in $M \sim (-t)^\beta$? Since this still only requires small deviations from T_c and small values of $\langle s\rangle_0$, it is sufficient to expand the tanh in a Taylor series:

$$\langle s\rangle_0 = \tanh(\beta J z \langle s\rangle_0) = \tanh(\frac{1}{t+1}\langle s\rangle_0) \ ,$$

yielding

$$\begin{aligned} \langle s\rangle_0 &= \frac{\langle s\rangle_0}{1+t} - \frac{\langle s\rangle_0^3}{3(1+t)^3} + \mathcal{O}\left(\frac{\langle s\rangle_0^5}{(1+t)^5}\right) \\ &= \langle s\rangle_0(1-t) - \frac{\langle s\rangle_0^3}{3} + \mathcal{O}\left(\langle s\rangle_0 t^2, \langle s\rangle_0^3 t, \langle s\rangle_0^4\right) \ . \end{aligned}$$

Hence, $-t = \langle s\rangle_0^2/3$, or

$$\langle s\rangle_0 = 3(-t)^{1/2} \ , \tag{1.95}$$

i.e. $\beta_{mf} = 1/2$, that is, when temperature is lowered from T_c, the magnitude of total magnetization increases as a square root dependency with the temperature difference $T_c - T$.

Similarly, we can now evaluate the specific heat critical exponent α: $C_H \sim |t|^{-\alpha}$:

$$C_H = T\left(\frac{\partial S}{\partial T}\right)_h \ ,$$

where the external field h is held fixed and the entropy $S = -\left(\frac{\partial F}{\partial T}\right)$ (exercise).

As result, it is found that

$$\begin{aligned} T < T_c\text{: } C_H &= \frac{3}{2}Nk + \mathcal{O}(t) \\ T > T_c\text{: } C_H &= 0 \ , \end{aligned}$$

hence, the specific heat has a jump discontinuity at T_c, but is otherwise constant. Therefore, $\alpha_{mf} = 0$.

It is also possible to compute the critical isotherm exponent δ, which is defined at $t = 0$, as $h \sim |M|^{\delta} sgn(M)$. This requires the addition of a small magnetic field h to the Ising Hamiltonian:

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - h \sum_i s_i \ ,$$

hence,

$$\langle s \rangle_0 = \tanh(\beta(Jz\langle s \rangle_0 + h)) \ .$$

With $T = T_c$, $\beta Jz = 1$, hence $\langle s \rangle_0 = \tanh(\langle s \rangle_0 + h/Jz)$, which can now be expanded for small $\langle s \rangle_0$ and h, yielding

$$\langle s \rangle_0 = \langle s \rangle_0 + \frac{h}{Jz} - \frac{\langle s \rangle_0^3}{3} + \mathcal{O}\left(\langle s \rangle_0^2 h, \langle s \rangle_0 h^2, h^3, \langle s \rangle_0^5\right) \ .$$

Therefore, $h \sim \langle s \rangle_0^3$ and $\delta_{mf} = 3$.

In analogous ways, the susceptibility exponent γ in $\chi_T \sim |t|^{-\gamma}$ can be computed, yielding $\gamma_{mf} = 1$ (left as exercise).

1.5.6 Landau theory

Landau proposed a phenomenological approach that does not consider the details of the interaction, but simply writes an expression for the free energy as a power series of the *order parameter*, the magnetization m, i.e. $\mathcal{F}(m) = \sum_q a_q m^q$. The only constraint imposed was that the functional form should respect the symmetry of the problem, i.e. in the absence of an external magnetic field both orientations of spin should be equivalent. Hence, only even powers of m should be allowed, yielding

$$\mathcal{F}(m) = \mathcal{F}_0 + a_2 m^2 + a_4 m^4 + \text{higher order terms} \tag{1.96}$$

for the first few terms.

Dropping all terms beyond the quartic contribution, one additionally has to consider that free energy should be a minimum at equilibrium. However, a minimum can only be obtained, if the coefficient $a_4 > 0$, otherwise there would be no lower bound to $\mathcal{F}$. The coefficient a_2 may however vary and it was Landau's contribution to consider that it might depend on temperature. Let us distinguish several cases:

- $a_2 > 0$: $\mathcal{F}(m)$ has only one minimum, namely at $m = 0$,
- $a_2 < 0$: $\mathcal{F}(m)$ has two additional minima,
- $a_2 = 0$: This is the transition between the case of a single minimum and three minima.

If we now choose a_2 to have an explicit temperature dependence, $a_2 = \tilde{a_2}t$, then a continuous transition occurs at $t = 0$.

The reader is encouraged to check that our previous expansion of the mean field free energy (Eq. 1.77), obtained from the microscopic partition function, yields a similar temperature dependence as Eq. 1.96 (consider exercise 2, below).

Mini tutorial: Based on physical plausibility, could you suggest further phenomenological expressions for the free energy?

Landau critical exponents

It is even possible to compute critical exponents from Landau's theory. Consider again β, in $m \sim (-t)^\beta$ near $t = 0$. First, we obtain the value of magnetization when $t < 0$ by searching for extrema in $\mathcal{F}$:

$$\frac{d\mathcal{F}(m)}{dm} = 0 = 2\tilde{a}_2 tm + 4a_4 m^3 = m(2\tilde{a}_2 t + 4a_4 m^2) \ ,$$

yielding $m = 0$ and

$$|m| = \sqrt{t\tilde{a}_2/2a_4} \ , \tag{1.97}$$

hence $\beta = 1/2$. Notably, this is the same critical exponent which we previously obtained within the explicit mean field derivation.

The specific heat critical exponent is obtained by differentiating $\mathcal{F}$ twice w.r.t. t. Using Eq. 1.97 in the free energy (Eq. 1.96), we have for $t < 0$

$$\mathcal{F} = \mathcal{F}_0 - \frac{\tilde{a}_2^2 t^2}{4a_4} + \mathcal{O}(t^3) \ , \tag{1.98}$$

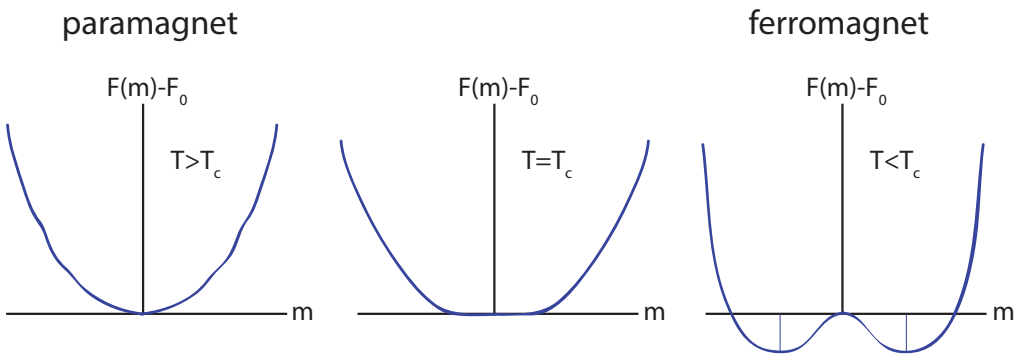


Figure 1.18: **Landau free energy for different values of temperature.** Zero external field ($h = 0$). For $T > T_c$ and $T = T_c$ the minimum of the free energy is located at $m = 0$. For $T < T_c$, there are two minima which are symmetrically located at finite magnetization.

hence, the specific heat tends to a constant as $t \to 0^-$. For $t > 0$, $m = 0$ and the specific heat vanishes, we hence recover the jump discontinuity found in the mean field solution, and $\alpha = 0$.

To obtain γ and δ one needs to add a magnetic field term to the free energy, hence breaking the previous symmetry regarding overall spin flip. The free energy then reads

$$\mathcal{F} = \mathcal{F}_0 - hm + \tilde{a}_2 t m^2 + a_4 m^4 \ , \tag{1.99}$$

Minimizing w.r.t. m gives

$$\frac{d\mathcal{F}}{dm} = -h + 2\tilde{a}_2 t m + 4a_4 m^3 = 0 \ ,$$

which yields the critical isotherm (setting $t = 0$) as $h \sim m^3$, i.e. $\delta = 3$.

By similar means one can also compute the isothermal susceptibility $\chi_T \sim |t|^{-\gamma}$ (exercises).

1.5.7 Exercises

1. Specific heat and susceptibility in the mean field approximation. (reproducing results above)
Follow the steps in Sec. 1.5.4 to obtain the mean field specific heat and susceptibility near $t = 0$. Hence, starting from the mean field free energy with $h = 0$, expand to fourth order in m and find the minimum in free energy to obtain the mean field magnetization as function of temperature (it is useful to introduce the dimensionless temperature $\theta \equiv \frac{k_B T}{Jz}$, where J is the coupling and z the coordination number of the lattice (number of neighbors of a site.) The free energy is now only a function of temperature. By taking the derivative w.r.t. temperature, obtain entropy S. Differentiating again w.r.t. temperature, obtain c_H. Discuss the difference of c_H for $t > 0$ and $t < 0$ near $t = 0$. What is the critical exponent α (in $c_H \sim |t|^{-\alpha}$)?

By using the self-consistency expression

$$m = \tanh(\beta(Jzm + h))$$

and inserting the reduced temperature $t = (T - T_c)/T_c$, obtain the magnetic susceptibility $\chi_T \equiv \partial m/\partial h|_t$ (i.e. the limit $h \to 0$ is taken before the limit $t \to 0$). Can you obtain the critical exponent γ (in $\chi_T \sim |t|^{-\gamma}$)?

2. Landau free energy from mean field free energy.
(a) By expanding out the mean field free energy (Eq. 1.75) to fourth order in m, show that the resulting expression in symmetric regarding the transformation $m \to (-m)$ and that the coefficient a_2 of the quadratic term is temperature dependent (i.e. $a_2(T)m^2$).
(b) By taking derivatives of the expression you found in (a), obtain the Landau theory critical temperature and compare this to the mean field critical temperature.

1.6 1D Ising model *

Ising was able to solve the model in 1D exactly during his thesis. Ising found that the 1D version of the model did not exhibit any phase transitions (except, strictly speaking, at $T = 0$).

In one dimension, the Ising model can be solved exactly by the so-called *transfer matrix method*. For simplicity, consider a periodic 1D lattice consisting of N sites (a ring, see Fig. 1.6). The corresponding 1D Hamiltonian is

$$\mathcal{H}_N = -J \sum_{i=0}^{N-1} s_i s_{i+1} - h \sum_{i=0}^{N-1} s_i \ , \tag{1.100}$$

where periodic boundary conditions mean that $s_N = s_0$.

1.6.1 Partition function

The partition function for the N sites is

$$Z_N = \sum_{\{s\}} \exp(\beta J(s_0 s_1 + s_1 s_2 + \cdots + s_{N-1} s_0) + \beta h(s_0 + s_1 + \cdots + s_{N-1})) \ , \tag{1.101}$$

where the notation $\{s\}$ means that all configurations of the different s_i are summed over in Eq. 1.101, J is again the nearest neighbors coupling and H the external magnetic field. The idea is now to break down the partition function into pairs of each two neighboring spins, yielding

$$Z_N = \sum_{\{s\}} \exp(\beta J s_0 s_1 + \beta h \frac{s_0 + s_1}{2}) \exp(\beta J s_1 s_2 + \beta h \frac{s_1 + s_2}{2}) \cdot . \tag{1.102}$$

$$\cdot \quad \exp(\beta J s_{N-1} s_0 + \beta h \frac{s_{N-1} + s_0}{2}) \ . \tag{1.103}$$

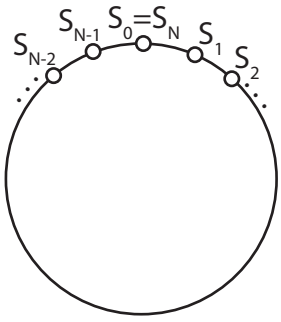


Figure 1.19: **1D Ising model.** Each of the N sites i has two nearest neighbors. Note the periodic boundary conditions, which are enforced by demanding $S_0 = S_N$, i.e. the 1D system becomes a closed loop of N sites.

Noticeable, each of the factors in the argument of exp in Eq. 1.103 can take one of four values, depending on the configuration of the two spins involved. It is more convenient to collect these four terms as the coefficients of a 2×2-matrix,

$$T_{i,i+1} = \exp(\beta s_i s_{i+1} + \beta h(s_i + s_{i+1})/2) \ ,$$

or more explicitly

$$\begin{bmatrix} e^{\beta(J+h)} & e^{-\beta(J)} \\ e^{-\beta(J)} & e^{\beta(J-h)} \end{bmatrix} \tag{1.104}$$

where the rows correspond to the two values of $s_i = \pm 1$ and columns correspond to the two values of $s_{i+1} = \pm 1$. It is more intuitive here to think of summing over configurations of *bonds*, rather than configurations of spins. In more mathematical language, what one is doing here is to switch to the *dual lattice*, the lattice of bonds. Note that in the one dimensional nearest-neighbor Ising model, the dual lattice is still a one-dimensional chain. For a 2D square lattice, the dual lattice is also a square lattice, while for a 2D triangular lattice, the dual becomes the *hexagonal lattice*, since each site there has six nearest neighbors.

1.6.2 Transfer matrix method

We return to the one-dimensional problem: The partition function thus turns into a product of N identical 2×2-matrices, where matrix multiplication beautifully ensures the constraint that the choice of spin orientation at site i has to be consistent from one matrix to the next (the configuration of columns for one matrix matches the configuration of rows for the neighboring). The partition function hence simplifies to

$$Z_N = \sum_{s_0=\pm 1} (T^N)_{0,0} = \mathrm{Tr}(T^N) = \sum_i \lambda_i^N \ ,$$

where λ_i are the eigenvalues of T.

The problem hence boils down to finding the eigenvalues of T, i.e. solving

$$\det(T - \lambda I) = 0 \ .$$

This gives

$$\left(e^{\beta(J+h)} - \lambda)\right)\left(e^{\beta(J-h)} - \lambda)\right) - e^{-2\beta J} = 0 \ ,$$

yielding

$$\lambda_{1/2} = e^{\beta J}\cosh\beta h \pm \sqrt{e^{2\beta J}\sinh^2\beta h + e^{-2\beta J}} \ .$$

1.6.3 Free energy

In the thermodynamic limit, that is, for $N \to \infty$, the free energy per site is easy to compute. Assuming that the eigenvalues are listed in decreasing

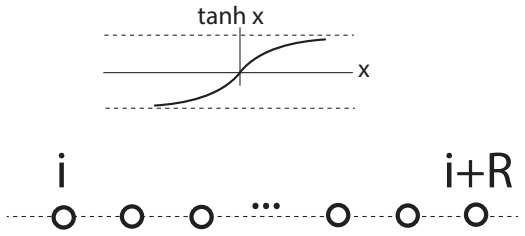


Figure 1.20: **Spin-spin correlation function.** Spin chain with two spins separated by R sites, $\tanh x$, with $x = \beta J$, is the correlation function between two neighboring spins.

magnitude, i.e. $\lambda_1 > \lambda_2$,

$$\begin{aligned} f &= -k_B T \lim_{N\to\infty} \frac{1}{N} Z_N \\ &= -k_B T \lim_{N\to\infty} \frac{1}{N} \ln\left[\lambda_1^N + \lambda_2^N\right] \\ &= -k_B T \lim_{N\to\infty} \frac{1}{N} \ln\left[\lambda_1^N \left(1 + \left(\frac{\lambda_2}{\lambda_1}\right)^N\right)\right] \ . \end{aligned}$$

In the thermodynamic limit, the ratio $\left(\frac{\lambda_2}{\lambda_1}\right)^N \to 0$ and the free energy per site is just

$$\begin{aligned} f &= -k_B T \ln \lambda_1 \\ &= -k_B T \ln\left[e^{\beta J}\cosh\beta h + \sqrt{e^{2\beta J}\sinh^2\beta h + e^{-2\beta J}}\right] \ . \end{aligned}$$

In the zero temperature limit, $\beta \to \infty$,

$$f \to -k_B T \ln\left[e^{\beta J}\left(\cosh\beta h + \sinh\beta h\right)\right] = -J - H \ ,$$

which is the ground state energy of a single spin.

1.6.4 Correlation function

To compute the correlation function (Fig. 1.6.3), in principle, the transfer matrix method could be exploited, both for $h = 0$ and $h \neq 0$. However, for $h = 0$, i.e. in absence of an external magnetic field, there is a simpler way to compute the correlation between two spins that are separated by R lattice sites. Consider first the correlation of any two neighboring spins, i.e. sites with a separation of unity:

$$\Gamma(1) = \langle s_i s_{i+1}\rangle = \frac{\sinh\beta J}{\cosh\beta J} = \tanh\beta J \ ,$$

a result that is simply obtained by summing over the two different values of the *bond energy* between sites i and $i+1$, namely $\pm J$. Now consider two spins that are separated by a distance of R lattice sites instead:

$$\Gamma(R) = \langle s_i s_{i+R} \rangle \ , \tag{1.105}$$

and realize that the expectation value in Eq. 1.105 will not change by inserting products $s_j s_j = 1$, hence

$$\Gamma(R) = \langle s_i s_{i+1} s_{i+1} s_{i+2} s_{i+2} \dots s_{i+R} \rangle \ .$$

Note that the expectation value now factorizes into a product of bond expectation values, when one simply considers the energy of each bond, not the sites themselves, i.e.

$$\Gamma(R) = \tanh^R \beta J \ .$$

Note that as R increases, the correlation falls off exponentially with distance. There is hence no long-ranged order in the 1D Ising model, which would require a power-law dependence on distance.

1.6.5 Exercises

1. Pair correlation function in 1D. (repetition of notes above)
The 1D Ising model has the advantage of showing an exact solution, but has the disadvantage, that it has no finite critical temperature, i.e. $T_c = 0$.

To see a manifestation of this, consider now the spin-spin correlation function

$$\Gamma(1) = \langle s_i s_{i+1} \rangle$$

for two neighboring spins in the absence of an external magnetic field ($h = 0$). Can you compute the correlation function of two spins separated by a distance R, i.e.

$$\Gamma(R) = \langle s_i s_{i+R} \rangle$$

by making use of $\Gamma(1)$?

Discuss that the dependence $\Gamma(R) \sim p(T)^R$ with $p(T) < 1$, i.e. that correlations decay exponentially at finite $T > 0$. What about $T = 0$?
(*Hint:* Make use of multiple insertions of unity and think more of summations over bonds than sites.)

2A. No finite temperature phase transition in 1D Ising model. +
Using the free energy and the concept of domain walls (Fig. 1.6.5), show that domain walls are favored for small but finite temperatures in 1D chains, but disfavored for 2D systems. Argue for a crude lower bound to the critical temperature in 2D.

2B. No finite temperature phase transition in 1D Ising model.
The free energy of the $h = 0$ ferromagnetic Ising model for a state of fixed

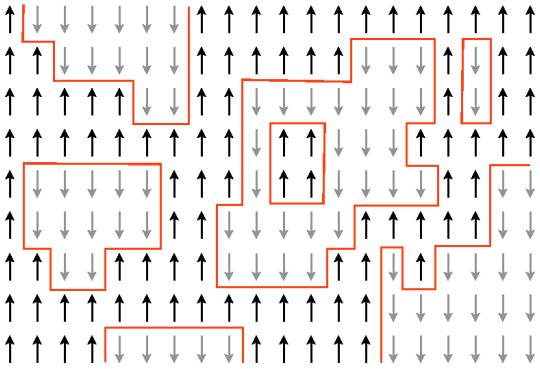


Figure 1.21: **Examples of domain walls in a 2D Ising model.** Domain walls shown as red lines.

energy E_Ω can be calculated from

$$F_\Omega = E_\Omega - TS_\Omega = -J\sum_{\langle ij\rangle} s_i s_j - Tk_B \log(\#states) \ .$$

Here, Ω indicates possible multiplicity of states of the same internal energy.
(i) Consider first a simple spin-$\frac{1}{2}$ Ising chain and compute the ground state internal energy and entropy to obtain F_Ω. Consider then an excited state with a single domain wall along the chain. A domain wall is defined as the boundary between two regions (domains) of different spin orientation (Fig. 1.6.5 shows examples in 2D). Compute the corresponding internal energy and entropy. Use the free energy difference to argue for a breakdown of the ordered state for any $T > 0$.
Hint: There is only one domain wall, but there are still many options where to place it.
(ii) Consider now a square lattice (2D) and again the formation of a domain wall, described by a region of $\uparrow$-spin enclosed by a region of $\downarrow$-spin. Take the region of $\uparrow$-spins to be bounded by a path of length n lattice spacings. Estimate the number of such paths by approximating the upper bound for the number of paths. Then again compute the free energy difference and argue for a finite transition temperature.
Hint: Since you only want an upper bound for the number of paths of length n, you may not need to require the path to be closed (just length n is fine). Also, be generous and even allow the path to cross itself, this will make your calculation much more straightforward).

1.7 Series expansion techniques

While brute force computation allows numerical approximations to the state of a certain, finite, system at a given temperature, external field and system parameters, it is generally advisable to seek results for the infinite system — at the very least to check for consistency of the solutions using numerical computations. Further, numerical simulations, such as the Monte Carlo method (Sec. 1.2), suffer from the strong fluctuations near T_c and *critical slowing down*, by which the convergence to the equilibrium value can require substantial computing time.

1.7.1 High temperature expansion

We again use the zero field Ising model on a 2D square lattice as a simple example. Consider the term

$$e^{\beta J s_i s_j} = \cosh(\beta J) + s_i s_j \sinh(\beta J) \equiv \cosh(\beta J)\,(1 + s_i s_j v)\ ,$$

where $v \equiv \tanh(\beta J)$. This choice is made in order to have a small parameter, which approaches zero at high temperatures, i.e. $v \to 0$ as $\beta \to 0$. In other words, $s_i s_j = 1$ becomes equally likely as $s_i s_j = -1$ in the limit of high temperature.

The partition function is

$$\begin{aligned} Z &= \sum_{\{s\}} \prod_{\langle ij \rangle} e^{\beta J s_i s_j} \\ &= (\cosh \beta J)^{\mathcal{B}} \sum_{\{s\}} \prod_{\langle ij \rangle} (1 + s_i s_j v) \\ &= (\cosh \beta J)^{\mathcal{B}} \sum_{\{s\}} \left(1 + \sum_{\langle ij \rangle} s_i s_j v + v^2 \sum_{\langle ij \rangle ; \langle kl \rangle} s_i s_j s_k s_l + \dots \right)\ , \end{aligned}$$

where $\mathcal{B}$ denotes the total number of bonds on the lattice, i.e. for a 2D square lattice of N sites $\mathcal{B} = 2N$. Since v is the small parameter in inverse temperature, including more orders of v means approaching lower and lower temperatures.

Mini tutorial: Consider a triangle consisting of three sites and work out Z.

Consider now terms of the form

$$\begin{aligned} \sum_{\{s\}} \left(s_i^{n_i} s_j^{n_j} s_k^{n_k} \dots \right) &= 2^N \text{ all } n_i \text{ even}\ , \\ &= 0 \text{ otherwise.} \end{aligned} \tag{1.106}$$

Here, N denotes the total number of spins on the lattice, as before. The result in Eq. 1.106 means that only closed loops contribute to the sum.

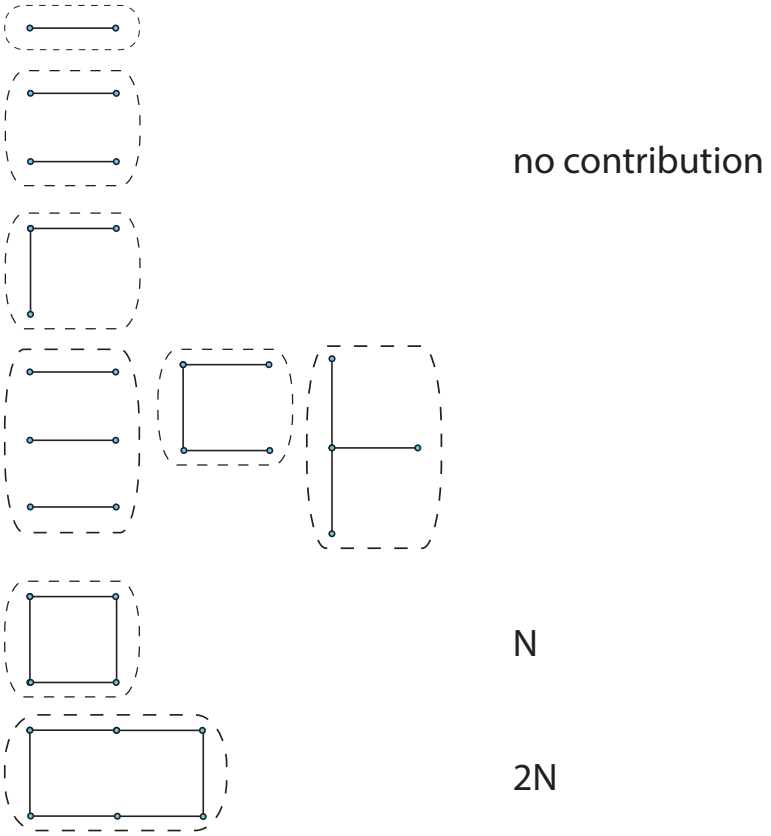


Figure 1.22: **Examples of motifs in high temperature expansion.** Note that motifs with "loose ends" do not contribute, while closed loops do.

We finally end up with the partition function

$$Z = (\cosh \beta J)^{\mathcal{B}} 2^N (1 + Nv^4 + 2Nv^6 + \frac{1}{2}N(N+9)v^8 + 2N(N+6)v^{10} + \dots) \,. \tag{1.107}$$

For simplicity, we introduce $K \equiv \beta J$ in the following. Several transformations are required to get a handy expression for the free energy.[6]

[6]To compute the free energy, one needs to take the logarithm of the partition function, as usual. However, note that this requires to expand in small powers of v: For this purpose, note that, with $\mathcal{B} = 2N$, $\ln(\cosh K)^B = N \ln(\cosh^2 K)$. Further, the square of $\cosh K$ can be expanded as a power series in $v = \tanh K$, i.e.

$$\cosh^2 K = \frac{1}{1-v^2} = 1 + v^2 + v^4 + v^6 + v^8 + v^{10} + \mathcal{O}(v^{12}) \,,$$

which makes it compatible with the final term in Eq. 1.107. Remembering that $\ln x = x - x^2/2 + x^3/3 - \dots$, we get the power series

$$\ln(\cosh^2 K) = \ln(1 + v^2 + v^4 + \dots) = v^2 + \frac{v^4}{2} + \frac{v^6}{3} + \frac{v^8}{4} + \frac{v^{10}}{5} + \mathcal{O}(v^{12}) \,.$$

In this process, it turns out that terms involving powers greater than linear in N drop out, which they should, since the free energy should be extensive (i.e. $\sim N$). The high-temperature free energy finally is

$$\mathcal{F} = -NkT(\ln 2 + v^2 + \frac{3}{2}v^4 + \frac{7}{3}v^6 + \frac{19}{4}v^8 + \frac{61}{5}v^{10} + \mathcal{O}(v^{12})) \; .$$

Mini tutorial: Briefly discuss the v-independent contribution to $\mathcal{F}$.

Some applications. Note a few simple lessons from the high-temperature expansion. Returning to our 1D Ising model (Sec. 1.6), we have a chain of N spins, which we can apply our formalism to. Consider first the case of open boundary conditions, i.e. sites 0 and $N-1$ only have one neighbor each. In that case, no closed loops are possible, and

$$Z_{1D}^{open} = 2^N \cosh^B \beta J \; ,$$

where $B = N-1$, i.e. the number of bonds in the open 1D chain. For the periodic chain ($B = N$), one closed loop of length N is possible, i.e. a single contribution arises from v^N, and

$$Z_{1D}^{closed} = 2^N \cosh^B \beta J(1 + v^N) \; ,$$

where the contribution from v^N scales to zero at any finite T in the thermodynamic limit, meaning that the boundary condition does not alter the result.

Note also, that it is straightforward to compute spin-spin correlation functions $\langle s_m s_n \rangle$ using the same formalism, when considering that the product $s_m s_n$ just acts as an additional factor in any of the products of bonds in the

---

We also need to "process" the final factor in Eq. 1.107, i.e.

$$\ln(1+X)$$

with

$$X \equiv (Nv^4 + 2Nv^6 + \frac{1}{2}N(N+9)v^8 + 2N(N+6)v^{10} + \dots) \; ,$$

and making use of the the power series $\ln(X) = X - X^2/2 + \mathcal{O}(X^3)$ to second order, we have

$$\begin{aligned} \ln(1+X) &= X - \frac{N^2 v^8}{2} + 2N^2 v^{10} + \mathcal{O}(v^{12}) \\ &= Nv^4 + 2Nv^6 + \frac{9}{2}Nv^8 + 12Nv^{10} \; , \end{aligned}$$

where we note that the two terms quadratic in N have dropped out, hence all remaining terms are linear in system size ($\sim N$), as we expect it for physical reasons for the free energy. Putting it all together, we can proceed and write down the free energy.

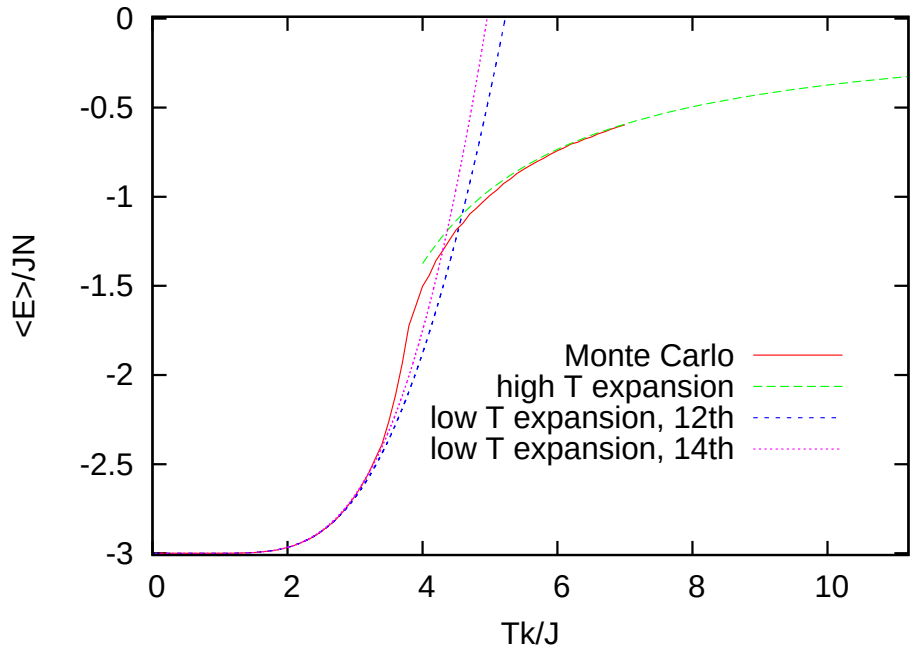


Figure 1.23: **Internal energy for the zero-field triangular lattice.** Comparison of a Monte Carlo simulation (red solid curve) and a high-temperature expansion to 6th order in $v = \tanh \beta J$ (green dashed curve).

series expansion. this means, that

$$\begin{aligned} \langle s_m s_n \rangle &= \frac{\sum_{\{s\}} \left(\prod_{\langle ij \rangle} e^{\beta J s_i s_j} \right) s_m s_n}{Z} \\ &= \frac{2^N \cosh^B \beta J \sum_{\text{graphs w. even powers except at } m \text{ and } n} v^{\#\text{ bonds}}}{2^N \cosh^B \beta J \sum_{\text{all graphs}} v^{\#\text{ bonds}}} \, . \end{aligned}$$

In other words, m and n should be the endpoints of lines on the lattice. In the 1D lattice this is just the path connecting the points m and n and

$$\langle s_m s_n \rangle = v^{|m-n|} = \exp\left(-\frac{|m-n|}{\xi} \right) \, ,$$

with the correlation length $\xi \equiv -1/(\ln \tanh \beta J)$. Hence, correlations decay exponentially in one dimension — a feature that could also be shown using an explicit solution of the 1D Ising model (*refer to* Sec. 1.6 for a derivation using transfer matrices).

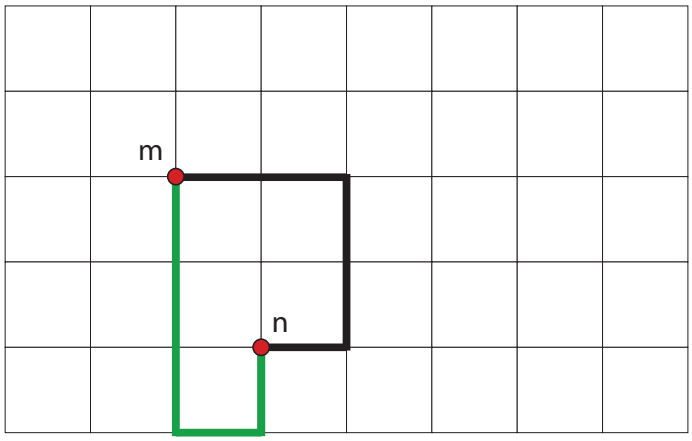


Figure 1.24: **Computation of the correlation function in 2D.** Red points show two point between which the correlation function is to be computed. Black and green path indicate possible graphs contributing to the correlation function, both of order v^5.

1.7.2 Low temperature expansion

At low temperature, it is convenient to order the partition function starting from the ground state energy (given that it is known) and consecutively add excitations of increasing energy. Consider the partition function

$$Z = e^{-E_0/k_BT}\left(1 + \sum_{n=1}^{\infty} \Delta Z_N^{(n)}\right) .$$

E_0 thereby denotes the ground state energy and $\Delta Z_N^{(n)}$ are all Boltzmann factors corresponding to excitations relative to the ground state. The label (n) indicates that n spins were flipped relative to the ground state.

For example, if one bond is anti-aligned, this leads to an energy "cost" of $2J$, yielding a Boltzmann factor $x = e^{-2J/k_BT} = e^{-2K}$. A single spin flip in a 2D square lattice hence requires a factor x^4,

When two spins are flipped, one needs to distinguish two cases: either, these particular spins are neighbors, in that case, six bonds become anti-aligned and the energy cost is $12J$; or the spins are not neighbors, in which case 8 bonds are anti-aligned and the cost is $16J$. Again, one needs to keep track of multiplicities: in the first case, there are $2N$ ways to choose neighboring spins, in the latter, there are N ways to choose the first spin, and $N-5$ to choose the second, so that the two are *not* neighbors. To avoid double counting, an additional factor of $1/2$ needs to be applied, yielding $N(N-5)/2$ configurations.

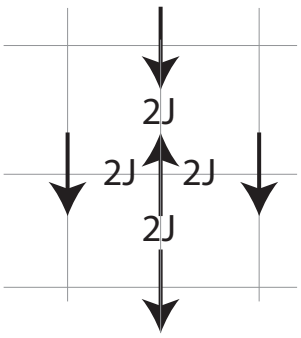


Figure 1.25: **Energy cost relative to the ground state for a single spin flip.** 2D square lattice Ising ferromagnetic without external field. This corresponds to a total Boltzmann factor of $x^4 = e^{-8J/k_BT}$.

Continuing systematically in the fashion for increasing numbers of flipped spins, one obtains an increasingly close approximation of the partition function

$$Z = e^{-E_0/k_BT}\left(1 + Nx^4 + 2Nx^6 + \frac{1}{2}N(N+9)x^8 + 2N(N+6)x^{10} + \mathcal{O}(x^{12})\right) . \tag{1.108}$$

1.7.3 Duality of the 2D square lattice Ising model

Comparing the terms in Eq. 1.107 and Eq. 1.108, it is clear that, in the present example of the 2D square lattice, there is a complete correspondence between $x = \exp(-2K)$ and $v = \tanh K$ for the low and high temperature series. Referring to the final factors in Eqs 1.107 and 1.108 as $g(v)$ and $g(x)$, respectively, one can equate the free energies of the two cases as

$$\frac{-F}{Nk_BT} = \frac{\ln Z}{N} = -\frac{E_0}{Nk_BT} + g(x) = \ln(2\cosh^2 K) + g(v) , \tag{1.109}$$

where $g(x)$ and $g(v)$ are infinite power series in their respective arguments. From other arguments (e.g. a Monte Carlo simulation or the mean field approximation) we might suspect a (single) critical temperature somewhere between the lowest and highest temperatures, i.e. $T = 0$ and $T \to \infty$. If this is so, then the singular contribution, i.e. that which leads to divergences at T_c, should match for the two expansions at hand. Since $\frac{E_0}{Nk_BT}$ and $\ln(2\cosh^2 K)$ are both perfectly "well-behaved" functions for $T > 0$, we are not concerned with these and focus only on the correspondence between $g(x)$ and $g(v)$. Even without knowing all the terms in these functions, it is possible to exploit the topological fact that they both contain the same type of terms. If we can ensure that $g(x) = g(v)$ at some "transition temperature", located in between the

low and high temperature expansion, we have a path to finding T_c. However,

$$v = x \ , \ \text{i.e.} \ \ \tanh K = \exp(-2\tilde{K}) \tag{1.110}$$

does not bring out any symmetry regarding the temperature $T = J/k_B K$. Here we have used different symbols K and $\tilde{K}$ to make clear that we are referring to (presently distinct) temperatures below and above T_c.

We can re-write Eq. 1.110 in a symmetric form by using that

$$\begin{aligned} \sinh 2\tilde{K} &= \frac{1}{2}(\exp(2\tilde{K}) - \exp(-2\tilde{K})) \\ &= \frac{1}{2}\left(\frac{1}{\tanh K} - \tanh K\right) \\ &= \frac{1}{\sinh 2K} \ , \end{aligned}$$

which finally gives the symmetric form

$$\sinh 2\tilde{K} \cdot \sinh 2K = 1 \ .$$

Due to the complete symmetry of $g(x)$ and $g(v)$ the only solution for $K = \tilde{K}$, i.e. in the limit of $T \to T_c$, is that

$$\sinh 2K = 1 = \frac{1}{2}(\exp 2K - \exp(-2K)) \ .$$

Introducing $q \equiv \exp(2K)$ gives a quadratic in q, yielding

$$q_{1/2} = 1 \pm \sqrt{2} \ .$$

Discarding the negative solution for physical reasons, we obtain the critical temperature

$$\frac{k_B T_c}{J} = \frac{2}{\ln(1+\sqrt{2})} \approx 2.27 \ .$$

Mini tutorial: Think about the duality seen in the above derivation. Would such a symmetry-based approach also allow calculating T_c for other lattices, say, the 2D triangular lattice? Give arguments for why you think it would/would not.

1.7.4 Exercises

A Triangular lattice +
By using a high-temperature expansion to sufficient order, obtain the free energy for a 2D triangular lattice. Discuss its scaling with system size and check that it makes physical sense. Consider also the low-temperature expansion. Compare results for internal energy with a variant of your Monte Carlo exercise from Sec. 1.2 by making appropriate plots.

B Triangular lattice

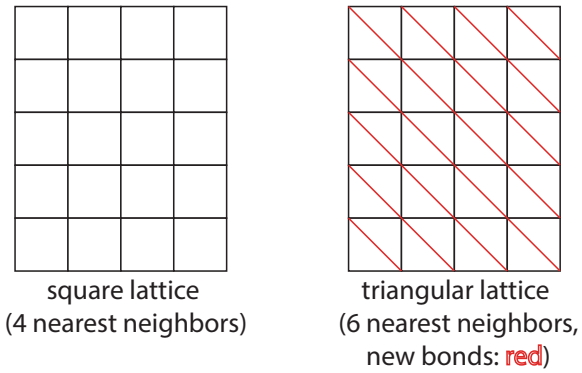


Figure 1.26: **Simple way of converting from a square to a triangular lattice.** By adding the red diagonal links to a square lattice, you can easily obtain a triangular lattice in a Monte Carlo simulation.

1. For a 2D Ising model on a *triangular* lattice, construct the high temperature expansion for the partition function. Continue the expansion to at least sixth order in the expansion parameter v. Compute the free energy F and show that term of higher power than linear in N cancel out — by which it is ensured that F is extensive (scales with system size).

2. Compute also the low temperature expansion for the triangular lattice partition function.

3. Compute the internal energy for the high temperature expansion, either by a derivative of the partition function w.r.t. β or by subtracting TS from the free energy.

4. Use a Monte Carlo simulation to compute the internal energy per site for the 2D triangular lattice (*hint:* figure below). Plot your results for the high and low temperature expansions together with the Monte Carlo result, as a function of temperature.

5. [optional] Construct the low temperature expansion to tenth order in $x \equiv \exp(-2\beta J)$ (there should not be many terms in this case) and plot the free energy along with the other curves. Do the expansions for low and high temperature look symmetric, just as was the case for the square lattice?

6. [optional] Flip the sign of J (hence, now $J < 0$) in your Monte Carlo simulation. Using the ground state energy (which you can find on theoretical grounds by considering first a single triangle) and the free energy near $T = 0$ to estimate the ground state entropy per site of the triangular lattice antiferromagnet. Does its value depend on system size?

1.8 Basic concepts of renormalization

In a renormalization group transformation the original Hamiltonian, usually defined in dimensionless form $\tilde{\mathcal{H}} \equiv \mathcal{H}/k_B T$, is transformed by an operation $\mathcal{R}$ to obtain the modified Hamiltonian $\tilde{\mathcal{H}}'$, i.e.

$$\tilde{\mathcal{H}}' = \mathcal{R}\tilde{\mathcal{H}} \ .$$

In this operation, some of degrees of freedom of $\tilde{\mathcal{H}}$ are removed, i.e.

$$N' = b^{-d} N \ ,$$

where b is the linear rescaling, d is the dimensionality of the lattice and N (N') is the original (transformed) number of lattice sites.

The general idea is, that a *partial trace*, i.e. a summation over a subset of spins (more generally: degrees of freedom) is performed, to obtain a new partition function that has fewer remaining summations to be carried out. In the case of the example in Fig. 1.8, it might be that the even-numbered sites each have a spin-half particle. One then rewrites the partition function to allow these sites to actually take on the available spin values. This will be practically carried out in the subsequent section (Sec. 1.8.2). The art of renormalization is to relate back the resulting Hamiltonian $\tilde{\mathcal{H}}'$ to take on the same functional form as the original $\tilde{\mathcal{H}}$, albeit with "rescaled" coefficients.

In general, the partition function should remain unchanged, i.e.

$$Z_{N'}(\tilde{\mathcal{H}}') = Z_N(\tilde{\mathcal{H}}) \ .$$

Since this also leaves the free energy unchanged, the free energy *per site* in the new Hamiltonian will increase by the rescaling b^d. Similarly, linear lengths are rescaled as b^{-1}, respectively momenta as b.

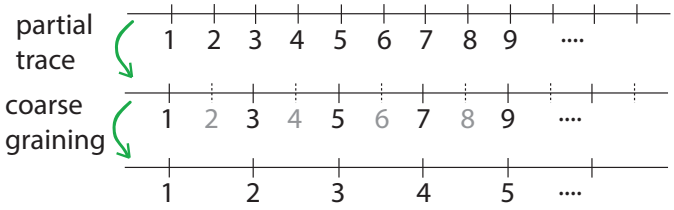


Figure 1.27: **Basic concept behind an RG transformation.** In a first step, a partial trace is carried out, removing a fraction of the original degrees of freedom (gray sites). In a second step, the Hamiltonian is "coarse grained".

The aim is to find the fixed points of the renormalization procedure, i.e. the parameter values where

$$\tilde{\mathcal{H}}' = \tilde{\mathcal{H}} \equiv \tilde{\mathcal{H}}^* \ . \tag{1.111}$$

Mini tutorial: Remind yourself of the condition for linear stability for a fixed point of a dynamical system.

1.8.1 Real-space renormalization for Percolation

You will be thoroughly introduced to the topic of percolation during the second part of Complex Physics (Kim Sneppen), and you will study this problem from approximations such as the Bethe lattice. A system near the percolation transition exhibits self-similarity, a property we will exploit in the following, in order to give an intuitive understanding of renormalization. We here use percolation as perhaps the simplest model that undergoes a "phase transition". The term is placed within quotation marks, as it does not strictly fit the definition of "phase transition" at the start of this chapter (Sec. 1.3), namely that there should be a discontinuity in one of the derivatives of the free energy. However, sometimes the term is also used more loosely for phenomena with a parameter, which can be tuned such that the behavior of the system changes drastically at a certain value — and a measurable or one of its derivatives has a discontinuity at this value of the parameter.

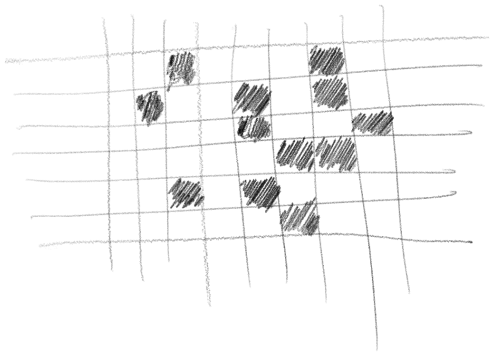


Figure 1.28: **Filling in boxes on a piece of paper.** Here, $p = 12/48 = 1/4$.

To appreciate (site) percolation, for a boxed piece of paper (hence a two-dimensional space), color a fraction p of squares in randomly (Fig. 1.28). If this fraction p is small, it will be unlikely that a path exists along the colored squares from one side to the other. For a reasonably large fraction of colored squares, however, the likelihood will be large. If the paper is infinitely large, the (critical) fraction p_c of squares that minimally need to be colored will be a sharp number. When studying the geometric properties of the emergent patterns, one specific aspect is that of "scale invariance".

The order parameter is now defined as the probability $P_\infty(p)$ of any given site to belong to the "spanning cluster," that is, the cluster that spans from one end to the other. Below the percolation transition, that is, when $p < p_c$, no such path exists, and the order parameter is zero. You can pick any site, and it will never be able to belong to the spanning cluster. As the transition is crossed ($p = p_c$), the probability changes from zero to finite values, as now at least some sites will belong to the spanning cluster. As the number of occupied sites increases further ($p > p_c$), eventually nearly every site will be part of the spanning cluster.

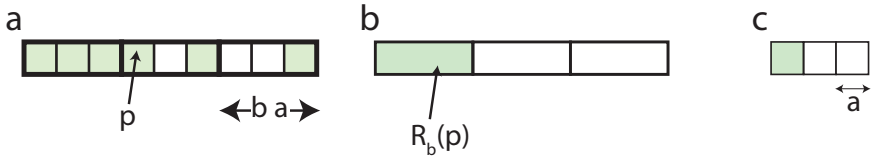


Figure 1.29: **Renormalization for one-dimensional site percolation. a**, Divide the lattice into blocks, which contain b sites each. **b**, Each block is coarse-grained and replaced by a single block which is occupied at probability $R_b(p) = p^b$. **c**, All length scales are reduced by the factor b leading to a renormalized version of the original lattice.

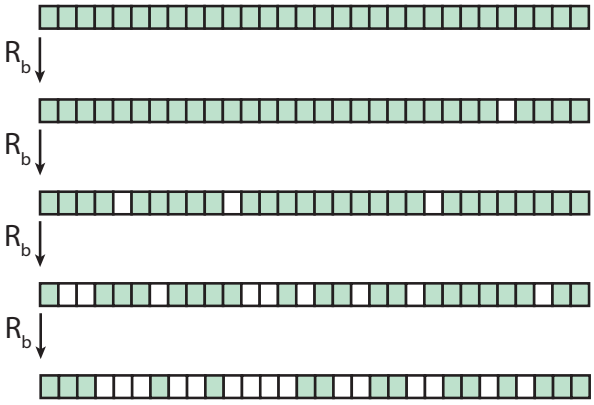


Figure 1.30: **Illustration of repeated renormalization for the one-dimensional lattice.** Each colored (blank) site represents an occupied (empty) site for a system where $p = .98$. Each line represents one additional coarse-graining step. Note that initial imperfections grow and gradually drive the system to the fully unoccupied state.

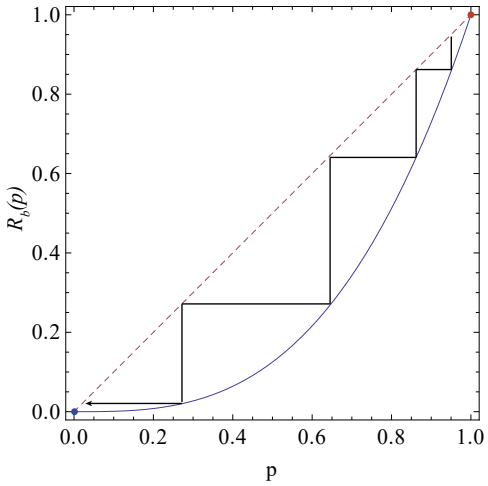


Figure 1.31: **Flow for the one-dimensional renormalization group transformation.** The schematic illustrates successive renormalization steps, where the initial probability p lies at .95.

Renormalization for the one-dimensional lattice. For an arbitrarily large one-dimensional system (a line), it is easy to appreciate that the percolation transition must occur, when each and every site is occupied ($p_c = 1$). If only one site is left unoccupied, there is no spanning cluster from one side to the other. When $p = 1$, each site will be part of the spanning cluster, hence, $P_\infty(p)$ transitions abruptly from zero to one when p goes from $p < 1$ to $p = 1$. The transition is discontinuous.

For dimensions greater than one, as for the Bethe lattice (*see* Sec. 2.6), which could be considered to be infinite-dimensional, a transition occurs at intermediate values $0 < p_c < 1$, and the functional form of the order parameter near p_c is described by a critical exponent β, by

$$P_\infty(p) \propto (p - p_c)^\beta \text{ , when } p \to p_c^+ \text{ .} \tag{1.112}$$

Self-similarity and rescaling. As mentioned, near the percolation transition, the system becomes self-similar, which can be expressed mathematically as the power law relation

$$M_\infty(\infty; l) \propto l^D \text{ ,} \tag{1.113}$$

which says that the mass of the dominant finite cluster contained in a window of size l scales like a fractal. The fractal dimension D thereby is different from the dimension d of the system. For example, if $d = 2$ and $D = 1.9$, the fraction of the finite window occupied by the incipient cluster would vanish, as the side length l increases. Two self-similar configurations are trivial, namely those of the empty lattice ($p = 0$) and the fully occupied lattice ($p = 1$). In both of these cases, since there are no fluctuations from one site to the next, the correlation length ξ (Sec. 1.3.2) vanishes. When p is increased away from zero, the correlation length increases and when $p = p_c$ there are fluctuations of all sizes and $\xi = \infty$, that is, the correlation length diverges.

To qualitatively understand self-similarity, consider a rescaling, where all length scales of the system are reduced by a factor $b > 1$. This rescaling would also affect the correlation length, which is rescaled as $\xi \to \xi/b$. When repeating the scaling, all finite correlation lengths would eventually be reduced to zero, $\xi = 0$ is hence a fixed point of the rescaling transformation, as any further rescaling with $b > 1$ would then leave $\xi = 0$ unchanged. The fixed point $\xi = 0$ can hence be seen as a "trivial" fixed point representing either of the trivially self-similar configurations at $p = 0$ and $p = 1$, whereas the fixed point $\xi = \infty$ corresponds to the self-similar state at the critical occupation probability $p = p_c$.

Formally, we can describe a rescaling as follows: consider the correlation length $\xi(p) = const|p - p_c|^{-\nu}$ for p in the vicinity of p_c. Here, ν is the *correlation length critical exponent* describing the divergence of ξ near the critical occupation probability p_c (*compare*: Eq. 1.49). When a rescaling is applied, $\xi \to \xi/b$, which defines the rescaling transformation $T_b(p)$ as

$$\frac{\xi}{b} = \frac{const\ |p - p_c|^{-\nu}}{b} = const\ |T_b(p) - p_c|^{-\nu}\ , \tag{1.114}$$

hence, $T_b(p)$ acts to map p onto a rescaled occupation probability, corresponding to the smaller correlation length ξ/b. The critical exponent ν is therefore a function of the rescaling transformation $T_b(p)$, namely

$$\nu = \frac{\log b}{\log\left(\frac{|T_b(p)-p_c|}{|p-p_c|}\right)}\ . \tag{1.115}$$

Hence, if one can find the rescaling function $T_b(p)$, one can determine the critical exponent. Since $T_b(p_c) = p_c$, Eq. 1.115 can be written more compactly as

$$\nu = \frac{\log b}{\log\left(\frac{|T_b(p)-T_b(p_c)|}{|p-p_c|}\right)} \tag{1.116}$$

$$= \frac{\log b}{\log\left(\frac{dT_b}{dp}|_{p_c}\right)}\ . \tag{1.117}$$

The above equations simply transfer all complications to the rescaling function $T_b(p)$, which does not automatically simplify the problem, since this function is not easy to determine. To make progress, we here introduce an approximate rescaling function $R_b(p)$, which has the effect of "coarse graining" all fluctuations on small length scales. The fixed point p^* corresponding to $R_b(p)$ ($R_b(p^*) = p^*$) is not necessarily the same as the exact fixed point p_c and there can be many ways to carry out the approximate rescaling $R_b(p)$. The coarse graining has the effect of "smearing out" all fluctuations on length scales less than b and simultaneously reduces the number of degrees of freedom in the system, for a system of dimension d, each rescaling reduces the number of degrees of freedom by b^d, and the original degrees of freedom N are reduced

to N/b^d. Information is lost upon rescaling and the procedure is therefore not invertible.

We return to the question of percolation, where the transition to a "spanning cluster" is characterized by a path existing from one side to the other. A reasonable choice for the rescaling $R_b(p)$ could hence be, to consider, whether a path exists within the "microsystem" of size $b \times b$. In one dimension, the rescaling is very simple: divide the lattice into blocks of size ba, where a is the lattice constant of the lattice before the rescaling operation (Fig. 1.29). All ba sites within each block are then replaced by a single block of size ba. Using the "spanning cluster" rule for the coarsening procedure, the block sites are then occupied with probability $R_b(p) = p^b$. Since the cluster is only spanned if each site within the cluster is occupied, any unoccupied site would break the spanning property. Finally, all length scales reduce by the factor b to make the block size identical to the original lattice spacing. The fixed point equation for the one-dimensional system hence is

$$R_b(p^*) = p^{*b} = p^* \ , \tag{1.118}$$

yielding the two fixed points $p^* = 0$ and $p^* = 1$, which correspond to the empty and entirely occupied lattices. For any initial occupation probability $p < 1$, repeated rescaling will drive the value of p towards the lower fixed point $p^* = 0$. $p^* = 1$ is an unstable fixed point and is non trivial (Fig. 1.30).

For percolation, it is useful to define the correlation function $\Gamma(\mathbf{r}_i, \mathbf{r}_j)$ between two sites $\mathbf{r}_i$ and $\mathbf{r}_j$ as the probability for the two sites to belong to the same finite cluster. For the one-dimensional system, the correlation function $\Gamma(r_i, r_j) = \Gamma(r) = p^r = \exp(r \ln p) = \exp(-r/\xi)$, where $r \equiv |r_i - r_j|$ is simply the distance between the sites at positions r_i and r_j. The result is easy to see, since each site between r_i and r_j must be occupied, and the probability for this to be the case is then p^r. This formula allows us to identify the correlation length as

$$\xi(p) = -\frac{1}{\ln p} \ . \tag{1.119}$$

In the limit of $p \to 1^-$, this can be expanded as[7]

$$\xi(p) = -\frac{1}{\ln p} = -\frac{1}{\ln(1-[1-p])} \to (1-p)^{-1}. \tag{1.120}$$

This expansion allows us to read off the critical exponent $\nu = 1$. Conversely, evaluating $\xi(p)$ for the rescaled lattice

$$\xi(R_b(p)) = -\frac{1}{\ln R_b(p)} \tag{1.121}$$

$$= -\frac{1}{\ln p^b} \tag{1.122}$$

$$= \frac{\xi(p)}{b} \ . \tag{1.123}$$

[7]Using the Taylor expansion $\ln(1-x) \to -x$ for $x \to 0$.

proves that, indeed, the correlation length decreases by the factor b under rescaling, as it should.

To determine ν, we compute the derivative of R_b w.r.t. p near the non-trivial fixed point:

$$\frac{dR_b}{dp}|_{p^*=1} = bp^{b-1}|_{p^*=1} = b \, , \tag{1.124}$$

which yields

$$\nu = \frac{\log b}{\log\left(\frac{dR_b}{dp}|_{p^*=1}\right)} = 1 \, . \tag{1.125}$$

For the one-dimensional case, the real-space renormalization group transformation hence correctly predicts the critical exponent $\nu = 1$.

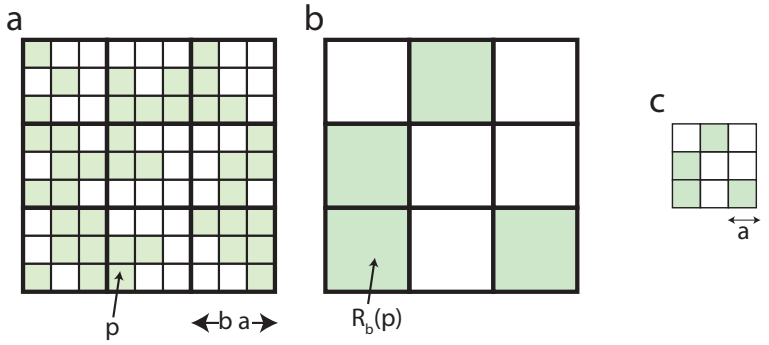


Figure 1.32: **Majority-rule renormalization for two-dimensional site percolation.** Sites of the square lattice have lattice have size a and are occupied with probability p. **a**, the lattice is divided into blocks of sizes ba, containing b^2 sites each. **b**, Each block is coarse-grained and all its sites replaced by one block of size ba that is then occupied with probability $R_b(p)$. **c**, length scales are reduced by the factor b, yielding blocks of the same size as in the original lattice, i.e., the block sites have become the new "elementary" sites. In total, we yield a rescaled version of the original lattice, where the coarsening requires sites to now be occupied at probability $R_b(p)$.

Renormalization in two dimensions requires choices. In two dimensions, there is no exact rule for specifying a possible coarse-graining of probabilities $R_b(p)$, which allows a "zooming-out" from the microscopic to a larger scale. Consider, for example, a possible coarse-graining procedure for the two-dimensional square lattice, termed the *majority rule* (Fig. 1.32). Each block of $b \times b$ sites is replaced by a rescaled block, which is occupied, when the majority of the $b \times b$ sites has been occupied. Such a rule is only approximate, when attempting to identify the critical probability p_c. You will find, that a similar rule can be used, when studying the Monte Carlo simulation of the

two-dimensional lattice Ising model (Sec. 1.2). In the current context of percolation, the actual criterion, which connects more directly to the ability of the lattice to "conduct" a signal from one end to the other, should be more directly reflected in the coarse-graining rule. Consider therefore the following.

Renormalization for the two-dimensional triangular lattice. Again consider that each site is occupied at probability p but that the lattice geometry is triangular, i.e., each site has six nearest neighbors within the plane. To achieve a coarse-graining, we divide up the lattice into sub-blocks of three sites each ($b = 3$). Note that it is possible to tile the entire lattice with such sub-blocks (Fig. 1.33). To arrive again at the original geometry after rescaling, length scales must now be reduced by $\sqrt{b}$. Within each sub-block of three sites, there are $2^3 = 8$ possible states.

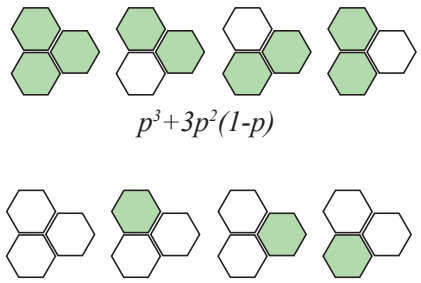


Figure 1.33: **Probability for a spanning cluster in the triangular lattice.** Clusters in the top (bottom) row are considered spanning (non-spanning). Note that the probability for a spanning cluster picks up two terms

If we now again use the spanning-cluster rule without restricting to any particular direction, and consider that "spanning" means that two or more of the sites within the sub-block must be occupied, four of the eight possible configurations allow for spanning — while the others do not (Fig. 1.33). We now evaluate the probability, that any of these four configuration occurs: full occupation occurs at probability p^3 while occupation of two sites occurs at probability $3p^2(1-p)$, leading to the fixed point equation

$$R_b(p^*) = 3{p^*}^2 - 2{p^*}^3 = p^* \,, \tag{1.126}$$

which factorizes to yield

$$p^*(p^* - 1)(2p^* - 1) = 0 \,, \tag{1.127}$$

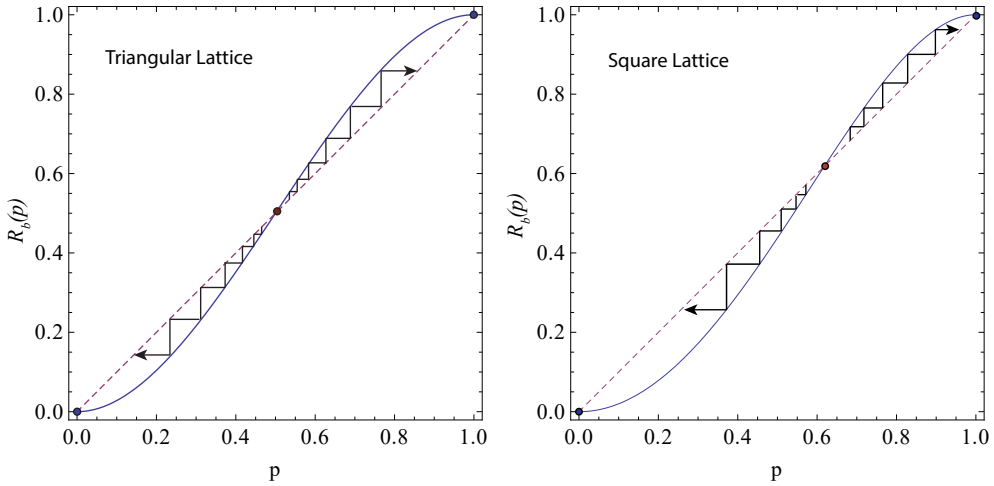


Figure 1.34: **Flow for a two-dimensional renormalization group transformation.** The schematic illustrates two sequences of renormalization steps for the triangular and square lattices (as labeled), where the initial probabilities lie immediately above and below the critical value. Note how, in the two cases, two different stable fixed points (at $p = 0$ and $p = 1$) are approached.

hence implies three fixed points $p^* = 0$, $p^* = 1/2$, and $p^* = 1$ (Fig. 1.34). We check that $p^* = 0$ and $p^* = 1$ are stable fixed points whereas $p^* = 1/2$ is unstable, hence, $p^* = 1/2$ is the non-trivial fixed points representing criticality. We can now again evaluate the derivative in Eq. 1.124 and obtain

$$\frac{dR_b}{dp}|_{p^*=\frac{1}{2}} = (6p^* - 6p^{*2})|_{p^*=\frac{1}{2}} = \frac{3}{2}\,, \tag{1.128}$$

yielding the critical exponent

$$\nu = \frac{\log b}{\log\left(\frac{dR_b}{dp}|_{p^*=\frac{1}{2}}\right)} = \frac{\log\sqrt{3}}{\log(3/2)} \approx 1.355\,. \tag{1.129}$$

Given the crude approximation made, this value lies remarkably close to the exact value $\nu = 4/3$ [9, 10]. However, we may ask, why the renormalization group transformation is not exact. One reason is, that sites that are connected in the original lattice might no longer be connected after renormalizing. This argument also goes the other way around. Further, the renormalization introduces bonds that are actually not there in the original lattice. There is an analogy to the renormalization of the two-dimensional Ising model, discussed in Sec. 1.8.6.

1.8.2 RG for 1D Ising model

The above discussion (Sec. 1.8.1) used the percolation problem as an intuitive introduction to the concept of renormalization, and the aim was, to allow the

probability of spanning to approach a fixed point under rescaling of space. For a Hamiltonian system one aims to again to rescale space, but now allows the parameters of the Hamiltonian to approach a fixed point under this rescaling.

Consider the partition function for the 1D Ising model of N sites. Define $K \equiv \beta J$ and $h = \beta H$, where J and H are the coupling constant and the external magnetic field. The Hamiltonian hence reads

$$\mathcal{H} = -K \sum_i s_i s_{i+1} - h \sum_i s_i \ .$$

When we consider only sites of even index $2i$, then each of these sites is only connected to odd-index neighbors. The goal is to evaluate the possible configurations for the even sites and perform a partial sum. Doing this, in terms of these three constants, the partition function is

$$Z(N,K,h) = \sum_{\{s\}} \left[e^{(K+\frac{h}{2})(S_1+S_3)+h} + e^{(-K+\frac{h}{2})(S_1+S_3)-h} \right] \times \dots \ . \qquad (1.130)$$

The goal is now to put the resulting partition function in a form that resembles that of the original partition function, but where the even sites are left out. To achieve this, the constants N, K, and h are allowed to be *rescaled*.

$$\begin{aligned} Z(N,K,h) &= e^{Ng(K,h)} Z(\frac{N}{2}, K', h') \\ &= e^{Ng} \sum_{\{s\}} e^{-\mathcal{H}'} \ , \end{aligned} \qquad (1.131)$$

where $\mathcal{H}' = -K' \sum_{i \text{ odd}} s_i s_{i+2} - h' \sum_{i \text{ odd}} s_i$. The rescaled Hamiltonian hence indeed represents a "coarse grained" version of the original Hamiltonian, where the lattice spacing has decreased by a factor two by removing half the sites. To accomplish this, the constants have to be rescaled. Matching each of the factors in Eq. 1.130 with the corresponding factors in Eq. 1.131, means that for each value of s_i, $s_{i+2} = \pm 1$

$$e^{(K+\frac{h}{2})(s_i+s_{i+2})+h} + e^{(-K+\frac{h}{2})(s_i+s_{i+2})-h} = e^{K' s_i s_{i+2} + \frac{h'}{2}(s_i+s_{i+2})+2g} \ .$$

Inserting the different combinations of s_i and s_{i+2} gives the three conditions

$$\begin{aligned} e^{2K+2h} + e^{-2K} &= e^{K'+h'+2g} \\ e^{h} + e^{-h} &= e^{-K'+2g} \\ e^{-2K} + e^{2K-2h} &= e^{K'-h'+2g} \ . \end{aligned}$$

These equations can be solved for K', h' and g, yielding

$$\begin{aligned} K' &= \frac{1}{4} \ln \frac{\cosh(2K+h)\cosh(2K-h)}{\cosh^2 h} \\ h' &= h + \frac{1}{2} \ln \frac{\cosh(2K+h)}{\cosh(2K-h)} \\ g &= \frac{1}{8} \ln \left[16 \cosh(2K+h)\cosh(2K-h)\cosh^2 h \right] \ . \end{aligned} \qquad (1.132)$$

The equations in Eq. 1.132 are recursive relations and specify the fixed points and flow diagram of the system. The action of an iteration is to remove half the degrees of freedom, by which the number of sites $N' = N/b$ with $b = 2$. The lattice spacing is increased to $a' = b\ a$. Other quantities depend on the lattice spacing and are correspondingly rescaled, e.g. the correlation length $\xi' = \xi/b$. The spins remaining in the new Hamiltonian interact through the rescaled coupling K' and act under the rescaled field h'. Noticeable, the renormalization for the 1D Ising model is exact, the resulting coarse-grained Hamiltonian looks exactly like the original, in the sense that no new terms are generated, e.g. interactions between three particles. The only thing that is necessary, is to define how the parameters of the system "scale" as the transformation is performed. This exactness is the crucial difference between a 1D and a 2D Ising model. In the latter, such exact mapping is not possible, the transformation always produces additional terms that are not present in the original Hamiltonian. The additional challenge in 2D hence becomes, to discard some of those additional terms, to be able to derive a self-consistent renormalization (Sec. 1.8.6).

1.8.3 Recursion relations

Defining for simplicity, $x \equiv e^{-4K}$, $y \equiv e^{-2h}$, and $z \equiv e^{-8g}$, the recursion relations are

$$\begin{aligned} x' &= x\frac{(1+y)^2}{(x+y)(1+xy)}\ , \\ y' &= y\frac{x+y}{1+xy}\ , \\ z' &= z^2xy^2\frac{1}{(x+y)(1+xy)(1+y)^2}\ . \end{aligned} \tag{1.133}$$

The first two equations do not depend on z, which means that the singular behavior of the free energy does not depend on a shift in energy scale. Investigating the fixed points in the x-y plane it is first seen that $x = 1$ is always a fixed point, irrespective of y, i.e. for any $0 \leq y \leq 1$. These fixed points are infinite temperature sinks.

Notably, the equations in Eq. 1.133 constitute a dynamical system in 3D parameter space. The linear stability in the vicinity of any fixed point X^* can be assessed by the Jacobian

$$J \equiv \frac{\partial q'}{\partial q}|_{X^*}\ , \tag{1.134}$$

where q is any of the three variables and q' represents the primed variables, i.e. the LHS of Eq. 1.133. At the ferromagnetic fixed point $X^* = \{x^*, y^*\} = \{0, 1\}$ (where the continuous phase transition occurs, and we are therefore interested

in), the Jacobian (Eq. 1.134) turns out to already be diagonal, and gives

$$\begin{aligned} \frac{\partial x'(x,y)}{\partial x}|_{X^*} &= 4 \,, \\ \frac{\partial y'(x,y)}{\partial y}|_{X^*} &= 2 \,, \end{aligned} \tag{1.135}$$

hence $x' \sim 4x$ and $\epsilon' \sim 2\epsilon$ near the fixed point X^*, where $\epsilon = y^* - 1$, i.e. a small parameter proportional to the magnetic field near $y = 1$. As the Jacobian J is diagonal, the coefficients in the Eq. 1.135 represent the eigenvalues of the Jacobian. Notably, if we are working in the range where a linearized description of the transformation in Eq. 1.133 is appropriate, i.e. sufficiently close to a fixed point, then a repeated application of the transformation will just lead to an additional scaling of the type in Eq. 1.135. This means, that the eigenvalues of a duplicate application will become

$$\lambda_i(b)\lambda_i(b) = \lambda_i(b^2) \,, \tag{1.136}$$

where b refers to the rescaling of spatial scales accomplished by the renormalization procedure. I.e. the repeated application just gives powers of the eigenvalues, which then represent the eigenvalue corresponding to J^2 (you can check this by just multiplying J with itself and finding the corresponding eigenvalues). But if Eq. 1.136 holds, then λ_i must have the form

$$\lambda_i(b) = b^{y_i} \,, \tag{1.137}$$

where y_i is a coefficient.

Notably, for x and ϵ these coefficients are different: the coupling x scales with $y_1 = 2$ while the field has $y_2 = 1$. Hence, as one zooms out of the lattice, "temperature" increases quadratically with the rescaling, while the "field" goes only linearly.

1.8.4 Broader implications

These consideration have implications for systems in any dimension $d > 1$ (in 1D there it is hard to define the reduced temperature t, as T_c appears in the denominator and $T_c = 0$): given that, under suitable rescaling of the parameters, the renormalization should keep the total partition function unchanged, the free energy per spin, f should increase by a factor of b^d in each step, where b is the rescaling of space. If we write the singular part $f^{(s)}$ of the free energy, i.e. the part that can produce divergences in one of the derivatives, before and after renormalization as a function of the different variables, e.g. temperature t and field h, we have (with Eq. 1.137)

$$f^{(s)}(t, h, \dots) \sim b^{-d} f^{(s)}(b^{y_1} t, b^{y_2} h, \dots)$$

in the limit where t and h approach zero, i.e. near T_c.

Even though the exact form of the free energy might not be known at T_c, knowing the scaling of $f^{(s)}$ and its parameters we should be in a position to analyze its scaling and that of its derivatives, e.g. the specific heat coefficient at zero magnetic field,

$$C \sim \left(\frac{\partial^2 f^{(s)}}{\partial t^2} \right)_{h=0} \equiv f^{(s)}_{tt}(h=0) \sim |t|^{-\alpha} \; .$$

In fact, it is straightforward to write down f_{tt} at zero magnetic field:

$$f^{(s)}_{tt}(t,0) \sim b^{-d+2y_1} f^{(s)}_{tt}(b^{y_1} t, 0) \; . \tag{1.138}$$

1.8.5 Scaling relations

How can we obtain the temperature dependence of the RHS of the equation 1.138? One needs to notice that the rescaling by a factor of b is arbitrary, any number could be chosen for b. So why not choose

$$b = |t|^{-1/y_1} \; ,$$

i.e. make the first argument in $f^{(s)}_{tt}$ become a constant? With this choice, the prefactor in Eq. 1.138 becomes a function of t:

$$f_{tt}(t,0) \sim |t|^{(d-2y_1)/y_1} f_{tt}(\pm 1, 0) \; , \tag{1.139}$$

which allows us to read of the critical exponent of specific heat,

$$\alpha = 2 - d/y_1 \; .$$

Similarly, the exponent β is obtained to be

$$\beta = (d - y_2)/y_1 \; .$$

The exponents corresponding to the magnetic susceptibility $\chi \sim |t|^{-\gamma}$ and that of the critical isotherm $H \sim |M|^{\delta} sgn(M)$ can be computed analogously.

Notably, by this procedure four critical exponents have been expressed in terms of only two variables y_1 and y_2. As a consequence, there must be two relations between these critical exponents. It can be checked that these area

$$\alpha + 2\beta + \gamma = 2 \; ,$$

i.e. the one we obtained from the Rushbrooke inequality in Sec. 1.3, as well as

$$\gamma = \beta(\delta - 1) \; .$$

Another consequence is that the scaling above and below the critical temperature should be the same, which is easily seen by inspecting Eq. 1.139.

When re-considering the pair-correlation function (Eq. 1.33), one obtains

$$\Gamma(\mathbf{r}, t, h, \dots) \sim c^2(b) \Gamma(b^{-1}\mathbf{r}, b^{y_1} t, b^{y_2} h, \dots) \; , \tag{1.140}$$

where it was used that all spatial scales are diminished by the factor b^{-1} at each renormalization step and $c(b)$ is some function of the spatial rescaling only, which however remains to be specified. At zero field ($h = 0$), one can employ a similar "trick" as before, setting $b \sim |t|^{-1/y_1}$ and obtains

$$\Gamma(\mathbf{r}, t) \sim c^2(|t|^{-1/y_1})\Gamma(|t|^{1/y_1}\mathbf{r}, \pm 1) \ ,$$

therefore the critical exponent $\nu = 1/y_1$.

To obtain $c(b)$ one can now set both t and h to zero in Eq. 1.140 and remember the Eq. 1.36, where

$$\Gamma(\mathbf{r}) \sim r^{-(d-2+\eta)} \ . \tag{1.141}$$

Using our previous equation

$$\chi_T \sim N \int \Gamma(r) r^{d-1} dr \tag{1.142}$$

together with Eq. 1.141, we have that

$$\chi_T \sim \xi^{2-\eta} \ . \tag{1.143}$$

Writing both χ_T and Γ in terms of t we have

$$t^{-\gamma} \sim t^{-(2-\eta)\nu} \ , \tag{1.144}$$

which then yields the additional exponent relation

$$\gamma = (2 - \eta)\nu \ . \tag{1.145}$$

In summary, it can be seen that all exponent relations result from the RH scaling. What is — of course — left to do is to obtain the values of y_i, which define the scaling w.r.t. the different variables t, h, etc. Once these y_i are known, all critical exponents can be computed. Note that the dimensionality of space (d) enters naturally in the previous equations, as d relates the different critical exponents to the (linear) rescaling of space b.

1.8.6 RG for 2D Ising model: Triangular lattice*

As mentioned in Sec. 1.8.2, the two-dimensional case presents further complications, not present in 1D. An exact mapping from a given Hamiltonian to one where only parameters are rescaled, is now generally not possible. We here discuss only the simplest 2D case, where the lattice is first broken down into larger clusters of so-called "block spins", which are taken as the "unperturbed" Hamiltonian. The interaction between the block spins is then incorporated perturbatively.

To this end, consider a *triangular lattice* spin-$\frac{1}{2}$ Ising model with a ferromagnetic coupling $J > 0$. The energy hence is

$$\mathcal{H} = K \sum_{\langle ij \rangle} s_i s_j + h \sum_i s_i \ , \tag{1.146}$$

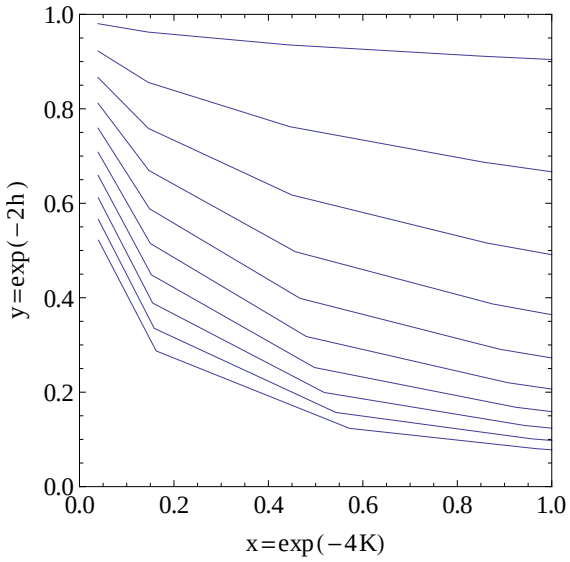


Figure 1.35: **RG flow for the 1D Ising model.** The flow always goes towards $x = 1$, i.e. the limit of $K = \beta J = 0$, hence infinite temperature ($T \to \infty$). The horizontal axis corresponds to $h \to \infty$, hence the flow is towards increasing h. The vertical axis corresponds to $T = 0$.

where $K \equiv -\beta J$ and $h \equiv -\beta H$ and $\langle ij \rangle$ denotes nearest neighbor sites i and j. The lattice is now broken down in triangular blocks of three sites each (Fig. 1.8.6), and we define the *block spin* of each triangle I by a "majority rule"

$$S_I \equiv \text{sign}\{S_1^I + S_2^I + S_3^I\} \; . \tag{1.147}$$

By this definition of block spins the lattice constant has been enlarged by a factor $l = \sqrt{3}$.

As a first step, we want to express the original Hamiltonian by a formally exact Hamiltonian using however the block spins. For this purpose we first define the collection of spins which constitute one triangle I as

$$\sigma_I \equiv \{S_1^I, S_2^I, S_3^I\} \; , \tag{1.148}$$

yielding $2^3 = 8$ possible combinations of spins. Under the majority rule (Eq. 1.147) there are four combinations of spins with $S_I = 1$ and four with $S_I = -1$. In this sense, the total number of degrees of freedom has been preserved.

The coarse grained Hamiltonian is

$$e^{\mathcal{H}'\{S_I\}} = \sum_{\sigma_I} e^{\mathcal{H}\{S_I,\sigma_I\}} \ .$$

The goal is to approximate $\mathcal{H}'$. To this end, we break $\mathcal{H}$ down into the interaction within block spins $\mathcal{H}_0$ and those between block spins V,

$$\mathcal{H} = \mathcal{H}_0 + V \ .$$

The Hamiltonian $\mathcal{H}_0$ is

$$\mathcal{H}_0 = K \sum_I \sum_{i,j \in I} S_i S_j \ ,$$

while the "perturbation", i.e. the interaction between the blocks, is

$$V = K \sum_{I \neq J} \sum_{i \in I, j \in J} S_i S_j \ .$$

We can now write the average of any quantity A with respect to $\mathcal{H}_0$ as

$$\langle A(S_i) \rangle_0 \equiv \frac{\sum_{\sigma_I \rangle} e^{\mathcal{H}_0\{S_I,\sigma_I\}} A(S_I,\sigma_I)}{\sum_{\{\sigma_i\}} e^{\mathcal{H}_0\{S_I,\sigma_I\}}} \ .$$

The equation for the coarse grained Hamiltonian thus becomes

$$e^{\mathcal{H}'\{S_i\}} = \langle e^V \rangle_0 \sum_{\{\sigma_I\}} e^{\mathcal{H}_0(S_I,\sigma_I)} \ .$$

Notably, since for the second factor on the RHS all blocks are independent, this factor can be evaluated to give

$$\sum_{\sigma_I} e^{\mathcal{H}_0\{S_I,\sigma_I\}} = Z_0(K)^M \ ,$$

where $Z_0(K)$ is the partition function for one block

$$Z_0(K) = \sum_{S_1 S_2 S_3} e^{K(S_1^I S_2^I + S_2^I S_3^I + S_3^I S_1^I)} \ .$$

Notably, the value of $Z_0(K)$ does only depends on *bond configurations*, i.e. the relative orientation of neighboring spins. Hence, the overall orientation S_I is irrelevant. There is only one way to obtain a sum of $3K$, but three configuration where two bonds are *frustrated*, i.e. two spins are anti-aligned, each with energy $-K$, hence

$$Z_0(K) = 3e^{-K} + e^{3K} \ .$$

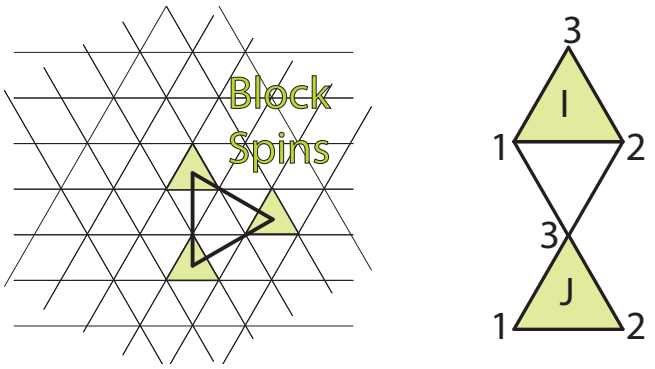


Figure 1.36: **Definition of block spins in 2D triangular lattice.** Block spins (yellow triangles) each consist of three spins and form a "coarse grained" triangular lattice of lattice constant enlarged by a factor $\sqrt{3}$. Schematic (right) shows the numbering of sites for two sub-lattice J and I and possible interactions between I and J.

The problem left to solve is hence

$$e^{\mathcal{H}'\{S_i\}} = \langle e^V \rangle_0 Z_0(K)^M \ ,$$

with M the total number of blocks in the system.

A cumulant expansion of $\langle e^V \rangle_0$ gives

$$\begin{aligned} \langle e^V \rangle_0 &= \langle 1 + V + \frac{V^2}{2} + \ldots \rangle_0 \\ &= 1 + \langle V \rangle_0 + \frac{\langle V^2 \rangle_0}{2} + \ldots \ . \end{aligned}$$

Notably, we consider the "perturbation" V to be small, and in doing so we will neglect higher order terms in V. Using

$$\log(1+x) = x - \frac{x^2}{2} + \mathcal{O}(x^3) \ ,$$

we have

$$\log \langle e^V \rangle_0 = \langle V \rangle_0 + \frac{1}{2}\langle V^2 \rangle_0 - \frac{\langle V \rangle_0^2}{2} + \mathcal{O}(V^3)) \ .$$

Re-exponentiating, we have

$$\langle e^V \rangle_0 = \exp\left(\langle V \rangle_0 + \frac{1}{2}[\langle V^2 \rangle_0 - \langle V \rangle_0^2] + \mathcal{O}(V^3) \right) \ .$$

The Hamiltonian $\mathcal{H}'$ can now be expressed approximately as

$$\mathcal{H}' = M \log Z_0(K) + \langle V \rangle_0 + \frac{1}{2}[\langle V^2 \rangle_0 - \langle V \rangle_0^2] + \mathcal{O}(V^3) \ .$$

$\langle V\rangle_0$ couples nearest neighbor blocks. Explicitly it is with

$$V = \sum_{I\neq J} V_{IJ} \ ,$$

where

$$V_{IJ} = k(S_3^J)(S_1^I + S_2^I) \ ,$$

thus $\langle V_J\rangle_0 = 2K\langle S_3^J S_1^I\rangle_0$ (compare schematic Fig. 1.8.6). Since $\mathcal{H}_0$ does not couple different blocks, i.e. cannot induce any correlations between spins on different blocks, The expectation value factorizes, giving

$$\langle V_{IJ}\rangle_0 = 2K\langle S_3^J\rangle_0\langle S_1^J\rangle_0 \ .$$

But

$$\langle S_3^J\rangle_0 = \frac{1}{Z_0}\sum_{\sigma_J} S_3^J e^{K[S_1^J S_2^J + S_2^J S_3^J + S_3^J S_1^J]} \ .$$

For $S_J = 1$ we have

$$\langle S_3^J\rangle_0 = \frac{e^{3K} + e^{-K}}{e^{3K} + 3e^{-K}} \ ,$$

while for $S_J = -1$ we have

$$\langle S_3^J\rangle_0 = -\frac{e^{3K} + e^{-K}}{e^{3K} + 3e^{-K}} \ ,$$

hence the expectation value of V within the unperturbed Hamiltonian becomes

$$\langle V\rangle_0 = 2K\Phi(K)^2 \sum_{\langle IJ\rangle} S_I S_J$$

with $\Phi(K) \equiv \frac{e^{3K}+e^{-K}}{e^{3K}+3e^{-K}}$. In total, the effective Hamiltonian, to first order in V is

$$\mathcal{H}'\{S_I\} = M\log Z_0(K) + K'\sum_{\langle IJ\rangle} S_I S_J + \mathcal{O}(V^2) \ ,$$

where $K' = 2K\Psi(K)^2$. We have hence achieved the goal of deriving an RG transformation that allows a rough approximation to the recursion relation for the coupling constant K.

Fixed points and critical exponents. What are the fixed points of the recursion relation we just obtained? Fixed points satisfy

$$K^* = 2K^*\Phi(K^*)^2 \ ,$$

which has three solutions

$$\begin{aligned} K^* &= 0 \\ K^* &= \infty \\ \Phi(K^*) &= 1/\sqrt{2} \, . \end{aligned}$$

Using $x \equiv \exp(4K)$, the latter relation can be inverted, giving a non-trivial fixed point

$$K_c = \frac{1}{4}\log(1 + 2\sqrt{2}) \approx 0.34 \ ,$$

whereas the exact result (Onsager) is $K_c = (\log 3)/4 \approx 0.27$.

1.8.7 Exercises

1. Real-Space renormalization group transformation on a square lattice. (*compare:* Fig. 1.34).

1. Define and outline the procedure of real-space renormalization transformation applied to site percolation.

2. Consider site percolation on a square lattice in two dimensions. Using blocks of size 2×2 and adapting the spanning cluster rule (in any direction) to define the real-space renormalization group transformation, show that $R_b(p) = p^4 - 4p^3 + 4p^2$.

3. Find the fixed points for the real-space renormalization group transformation in the above equation and comment on their nature. What are the correlation lengths ξ associated with the respective fixed points? Discuss the concept of flow in p-space associated with the real-space renormalization group transformation R_b.

4. Identify the critical occupation probability p_c, derive the equation used to determine the correlation length exponent ν predicted by the real-space renormalization group transformation, and evaluate ν. Compare the findings to the analytic results and comment on the discrepancies.

5. Discuss the concept of universality in the theory of percolation. Give examples of quantities which are universal and non-universal, respectively.

2. Exponent relations.[8]
Starting from the scaling for the singular part of the free energy

$$f^{(s)}(t,h) \sim b^{-d} f^{(s)}(b^{y_1}t, b^{y_2}h) \tag{1.149}$$

where $t \equiv (T-T_c)/T_c$ and $h = h_0/k_BT$ show that $\alpha = 2-d/y_1$, $\beta = (d-y_2)/y_1$, $\gamma = (2y_2 - d)/y_1$ and $\delta = y_2/(d - y_2)$ and hence confirm that $\alpha + 2\beta + \gamma = 2$, and $\gamma = \beta(\delta - 1)$.

3. Numerical renormalization.
Set $T = T_c$ in your 2D square lattice zero-field Monte Carlo simulation and make lattice size N sufficiently large (say, $N \sim 100 \times 100$). For a snapshot of your simulation near equilibrium, perform a "numerical renormalization", where you apply the following majority rule: For any square consisting of 2×2 sites, color this square by the majority of spins, i.e. if more are pointing up than down, the "block spin" will point up. For a "tie", choose randomly between up and down for the block spin. The resulting lattice will then only have $N^2/4$ sites and represent a zoomed-out version of the original. Repeat this procedure several times and observe the patterns obtained for the various iterations. If you are close to T_c you should observe that patches at different scales remain, even when you rescale several times.

[8] Yeomans, problem 8.2

Now repeat this exercise for a temperature slightly below T_c, e.g. $T = .99T_c$. Now you should find that the resulting patterns "flow" towards one of the polarized extremes, either all spins pointing up or down.

Repeat again for T slightly larger than T_c, say $T = 1.01T_c$. Now the result should be that pattern become random, you will end up with a featureless mix of up and down spins.

Chapter 2

Scaling: Percolation, Self-Organization, Fracture

There is nothing insignificant in the world. It all depends on the point of view.

– Johann Wolfgang von Goethe

2.1 Scaling in context

The previous chapter on the Ising model investigated the classical example of an equilibrium system, in which the energy E_i on each degree of freedom is given by an overall temperature T, and distributed as $exp(-E_i/k_BT)$. For example, if one spin in the 1-dimensional Ising model is aligned with its neighbours has ground state energy 0, it will have the relative probability $exp(-2J/k_BT)$ to be switched in opposite direction. Such systems predict that the probability of finding any particular state to be occupied exponentially decreases with the state's energy (*see* Sec. 1.1.2). Exponential decrease implies that the probability distribution $P(k)$ of some observable k has a typical scale a:

$$P(k) \propto exp(-k/a)\,. \tag{2.1}$$

For example, consider a system where energy increases linearly with the size of the system. In this case, the probability for having size a is

$$\begin{aligned} P(a) &= 1/e \\ P(2a) &= 1/e^2 \rightarrow P(2a)/p(a) = 1/e \\ P(10a) &= 1/e^{10} \\ P(20a) &= 1/e^{20} \rightarrow P(20a)/P(10a) = 1/e^{10} \end{aligned}$$

and the probability for finding a system of size, say, 100 a or 200 a becomes very small — and the latter exceedingly smaller than the first. In equilibrium, this type of distribution arises when the probability to concentrate one more unit of energy on degree of freedom is $1/e$, and this remains true independently

of how much energy one has already concentrated. In any case, the exponential distribution make unusually large concentration of energy essentially impossible.

Exponential distributions and systems with characteristic sizes are indeed often seen in the real world. Think for example about clusters of connected identical spins in the Ising model, which all will be of similar size except when the system is near the critical point. This point is critical in the sense that the exponential associated to the binding energy of the ordered state just exactly balances the exponential associated to entropy of the disordered state (*compare* Fig. 1.6.5 and corresponding exercise). But this is a special situation because the temperature will have to be tuned to be at the critical point.

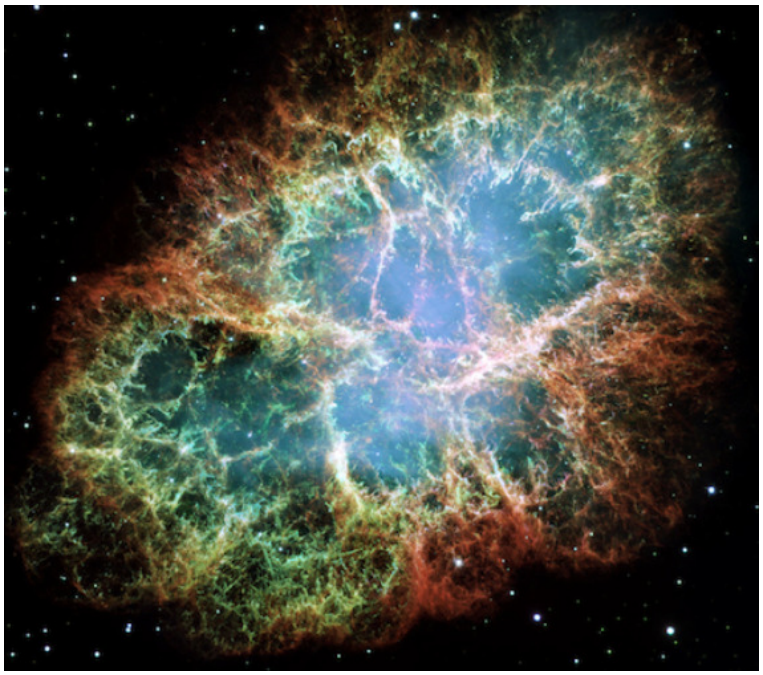


Figure 2.1: **Remnants of a supernova that formed the Crab nebula.** Fractals and power laws are ubiquitous in nature, for example in turbulence or fragmentation. Image: NASA, ESA, J. Hester, A. Loll (ASU) Acknowledgment: Davide De Martin (Skyfactory).

However, there also exist many real world systems that have no characteristic scale or size. Examples include:

- Earthquakes can be very very large, even if most of the recorded earthquakes are so small that we wouldn't even feel them without sensitive equipment.

- Solar flares can be huge, although most solar activity is fairly limited. (The "Carrington event" in 1859 gave rise to aurora all over USA, at a level where the illuminated night sky allowed people to read.) Energy

distribution $P(E) \sim 1/E^{1.52 \to 1.65}$ for solar x-ray bursts [11]. In same data set both duration of solar flares and time between flares scaled with exponent ~ -2.

- Most people have a few 100 followers on their internet activity, but some have millions.
- Gene regulatory networks, where most proteins only associate to few others, but some are central hubs and associate with most others.
- Turbulent liquids can have large vortices with huge velocity gradients, even if most of the liquid is locally laminar.
- Financial crashes can become very big, even if typical day-to-day stock market fluctuations are small.

A visual example with many scales of heterogeneity is shown in Fig. 2.1.

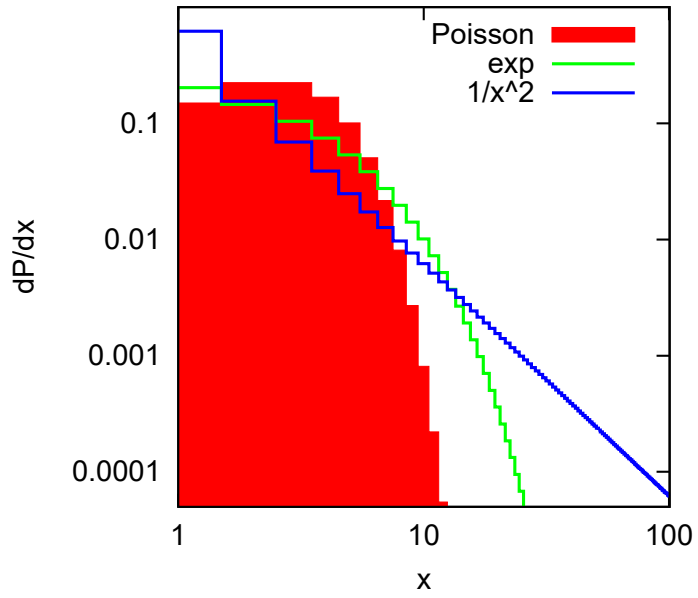


Figure 2.2: *Comparison between a power law with exponent 2, that is,* $\sim 1/k^2$, *the Poisson distribution* ($p(x) \propto a^x/x!$) *and an exponential distribution* $p(x) \propto exp(-x/a)$. *Notice that both the* x *and* y *axes are logarithmic.*

In many of the above cases the distribution $P(k)$ of sizes k of the relevant observable is a power law:

$$P(k) = \frac{1}{k^\gamma} \ , \tag{2.2}$$

with an exponent γ of about two. To give some examples: for energy releases in earthquakes γ is about 1.7; for the number of links in networks $\gamma \sim 2.2$; the largest exponent is found in stock market crashes which is presumably about $\gamma \sim 4.5$). This type of distribution is shown in Fig. 2.2. The exponent $\gamma = 2$ corresponds to the famous Zipf distribution, found for the frequencies

of the distinct words used in books. This is also true for the distribution of number of people in cities: there is half the number of cities above 2 million than numbers of cities with more than 1 million people.

Often the Zipf distribution is plotted in a slightly more complicate way using the rank distribution of sizes in a group/population, see Fig. 2.3. In rank distributions, along the x-axis one plots the rank, where 1 is the largest, 2 the second largest and so forth. On the y-axis one plots the size corresponding to each rank. Thus the rank distribution is the reciprocal of the cumulative distribution (mirror symmetric plot along x=y in the log-log plot). The exponent of the Zipf distribution is one divided by the exponent of the cumulative size distribution (i.e., $= 1/(\gamma - 1)$).

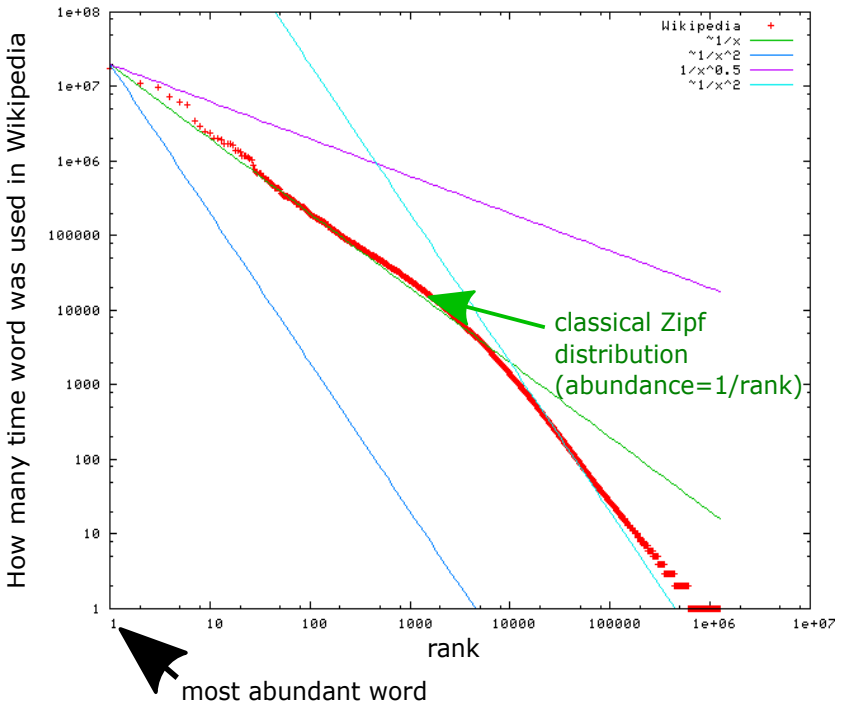


Figure 2.3: **Rank ordered distribution of the number of times different words are used in Wikipedia.** Thus there are ten words that were used more often than 2,000,000 times, and 1,000 words that were used more often than 20,000 times. (identify word of rank 1000 and see how many times this was used. All words with lower rank have been used more). Figure downloaded from Wikipedia. See also Zipf GK (1949) "Human Behavior and the Principle of Least Effort".

Distribution $P(k)$ is scale-free is equivalent to saying that $P(k) \propto k^{\gamma}$:

To prove this, assume first that the distribution is scale free, hence

$$\begin{aligned} P(s\cdot k)/P(k) &= P(s)/P(1) \\ \Rightarrow log(P(s\cdot k)/P(1)) &= log(P(k)/(P(1)) + log(P(s)/P(1)) \\ \Rightarrow f(s\cdot k) &= f(s)+f(k) \end{aligned}$$

where $f(k) = log(P(k)/P(1))$ and $f(1) = 0$. Thus $f(k)$ is a logarithm of k:

$$log(P(k)/P(1)) = \gamma \cdot log(k) \Rightarrow P(k) \propto k^{\gamma} \tag{2.3}$$

Reversely, if one assumes that $n(k) \propto k^{\gamma}$, "scale-freeness" is proven by multiplying the argument k with a factor s, and observing that the frequency n changes with the same factor s^{γ} for all values (scales) of k.

To compare a distribution containing a scale (s), say, an exponential, with a scale-free one, compare:

$$\begin{aligned} P_{scale}(k) = exp(-k/s) \quad &\text{with} \quad P_{power-law}(k) = k^{-\gamma} \\ P_{scale}(s) = exp(-1) \quad &\text{with} \quad P_{power-law}(s) = s^{-\gamma} \\ \frac{P_{scale}(100\cdot s)}{p_{scale}(50\cdot s)} = exp(-50) \quad &\text{with} \quad \frac{P_{power-law}(100\cdot s)}{P_{power-law}(50\cdot s)} = 2^{-\gamma} \end{aligned}$$

Thus, extreme events are much more likely in the case of power laws than when there is a characteristic scale.

In these lectures we will repeatedly return to scale-free behavior, and most often this is caused by some sort of non-equilibrium dynamics. As an introduction this chapter will introduce three very different ways to obtain scale-free behavior:

- Percolation as a static problem with some analogy to the Ising model with power laws close to a critical point.
- Fragmentation, a sudden process that gives power laws (in fragment sizes).
- Self organized critical systems to a critical state that is only maintained when there is a separation of timescales.

In subsequent chapters we will see other mechanisms that generate power laws, including in particular "rich gets richer" and merging processes (both introduced in the network chapter (chapter 3)).

Questions:

2.1) Consider the distribution [12, 13, 14]

$$p(s) \propto 1/s^{\tau}$$

as a distribution for wealth in human society (with s larger or equal to a lower cutoff fortune of unity). Argue that the situation where $\tau \leq 2$ makes for a fundamentally different society than when $\tau > 2$. (Hint: Consider contribution to average wealth) Notice that, as mentioned, $\tau = 2$ is the famous Zipf distribution observed for example for word frequencies in books [15].

Qlesson: Exponent 2 is special, and it is also the most common in nature.

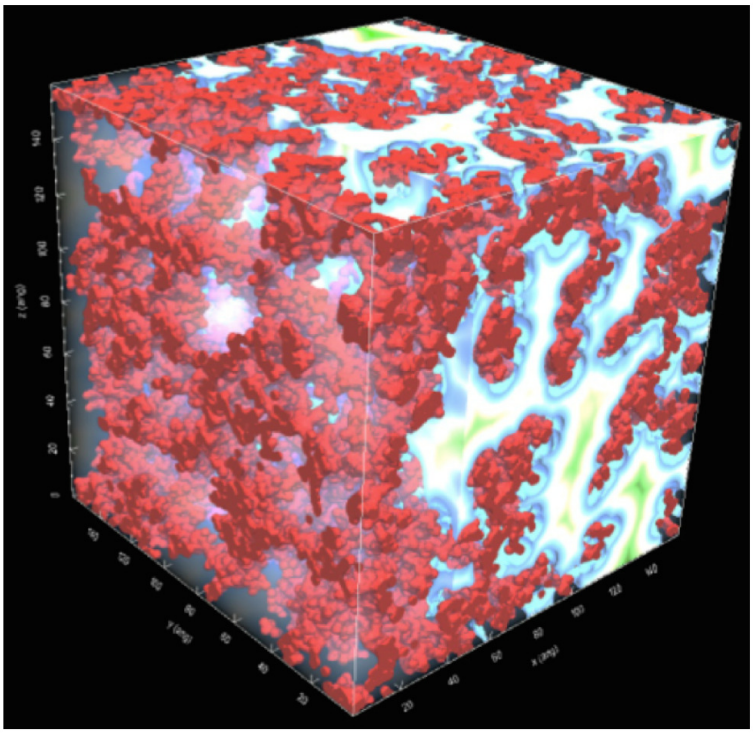


Figure 2.4: **Nanoporous silicate, with color indicating distance to nearest part of solid. From Malte Sørensen's lecture notes on percolation, Springer.**

2.2 Percolation

Percolation deals with the problem of connecting/percolating a path across a heterogeneous material, which can be thought of as partially insulating, partially conducting, and the path must be taken through the conducting part. This type of problem is found within many fields of study, including physics, geology, epidemics and sociology. Imagine a glass jar filled with beads, some of which are made of glass and thus insulating, and some are metal and thus conduct electricity. One may thus ask at which density of metal balls the mixed system will be able to conduct a current. And one may be interested in how the conductivity changes as one approaches this critical density. This and analogous questions are formally addressed by studying percolation.

Let us first consider a simple percolation example on a two-dimensional square lattice (Fig. 2.5). In this simulation we first assign each site a probability p to be conducting and probability $1-p$ to be empty (or insulating). We then allow bonds between all nearest neighbors which are both occupied. This allows us to define clusters, consisting of sites which are directly or indirectly connected by bonds. Each of these clusters are colored with a different color. The cluster size s is defined as the number of sites of equal color. Clearly, for larger values of p the probability of finding larger clusters will increase. In the first exercise session we will repeat the simulation in Fig. 2.5 using python,

which in turn will allow us to gain some intuition for this type of problems.

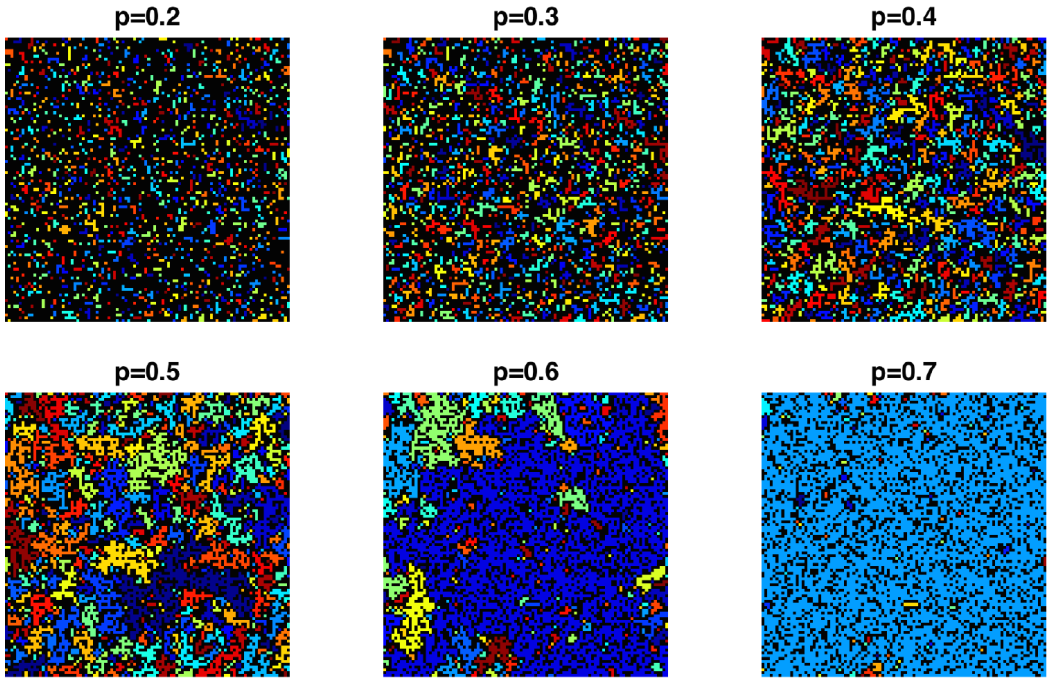


Fig. 1.4 Plot of the clusters in a 100×100 system for various values of p.

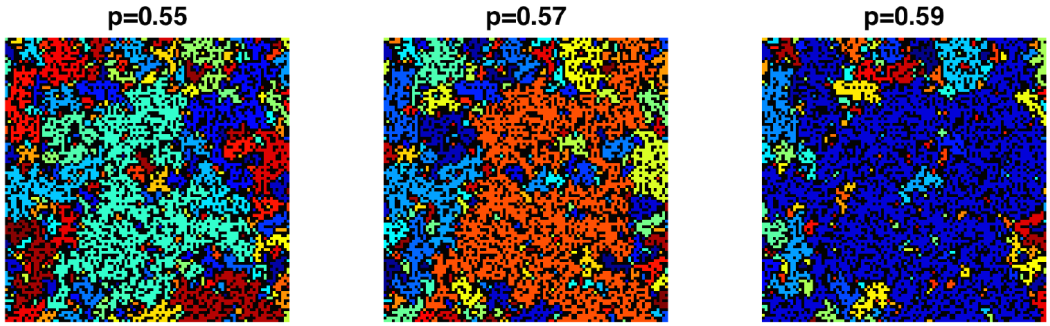


Figure 2.5: **Clusters in the site percolation model.** In the example shown, sites are organized on a two-dimensional square lattice and each site is occupied with probability 0.55. Neighboring sites of equal color belong to the same cluster. From Malte Sørensen's lecture notes on percolation, Springer.

Questions:

2.2) Make a two-dimensional percolation program on a square lattice, identify the critical point p_c, or *percolation threshold*, that is, the probability for conducting sites at which a current could flow across the system. Obtain the cluster size distribution close to this critical point p_c. In matlab a two-dimensional square (matrix) of dimension 100 is generated and plotted by the following sequence of orders: $L = 100$; $r = rand(L, L)$; $p = 0.6$; $z = r < p$; $[lw, num] = bwlabel(z, 4)$; $img = label2rgb(lw)$; $image(img)$. (Hint: $p_c = 0.59275$ and the cluster size distribution at criticality should be $n(s) \propto 1/s^{187/91}$). It can be useful to either use logarithmic

binning, or to plot the cumulative distribution $Cum(s) = \sum_s^\infty n(s)$, where reversely $n(s) = dCum(s)/ds$.

QLesson: Percolation is easy to simulate but it is exceedingly difficult to find analytic expressions. However, it is much simpler in the case of ∞ dimension where branches never meet.

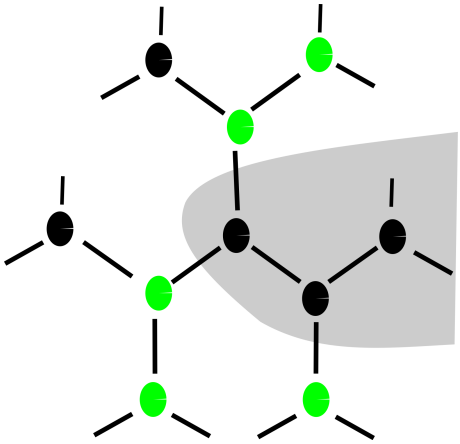


Figure 2.6: **Bethe lattice with occupied (black) and empty sites (green).** The black sites form clusters, that may be percolating to infinity, if the density of the black sites (the probability of being occupied) is sufficiently large (larger than 1/2 in the example shown). The gray-shaded area marks the start of a large cluster. This type of network was introduced by H. A. Bethe "Statistical theory of super-lattices". Proc. Roy. Soc. London Ser A 150: 552–575 (1935).

2.2.1 Percolation on a Bethe-lattice

A Bethe lattice is a tree-like network without any loops, that is, there is only one connection between any two sites. In Bethe lattices each site further has a fixed number z of nearest neighbors (termed *connectivity* or *coordination number*).

The Bethe lattice with coordination number $z = 3$ is illustrated in Fig. 2.6. Each branch originating from a given site contains $z - 1$ (two, in the example) new sub-branches at the $z = 3$ neighboring sites.

How can one think of the dimensionality of the Bethe lattice?

Consider the number $N(x)$ of sites within a distance x from a given site. For the Bethe lattice, $N(x)$ can be computed as

$$N(\leq x) = 1 + 3 \cdot (1 + 2 + 2^2 + ... + 2^{x-1}) = 1 + 3 \cdot \sum_{i=0}^{x-1} 2^i = 3 \cdot 2^x - 2 \ , \quad (2.4)$$

where we have used that $\sum_{i=0}^{x-1} 2^i = 2^x - 1$ (prove that). If one instead considers a lattice in D spatial dimensions, the number of sites within a distance x would scale as

$$N_D(\leq x) \propto x^D \tag{2.5}$$

because $2^x > x^D$ for large enough x, the Bethe lattice will have a larger dimension than any chosen dimension D. Thus, the **Bethe lattice is formally infinite-dimensional**, a feature that it shares with most real world networks.

Mini tutorial:
How many directed paths are there between two points in a beta-lattice?

The critical probability p_c for percolation on a Bethe lattice is the probability p where each new site that belongs to a cluster, on average is connected to one new site that is occupied further out in the next layer of the network. Demand that

$$p_c \cdot (z-1) = 1\,, \quad \text{hence} \quad p = \frac{1}{z-1} = \frac{1}{2} \tag{2.6}$$

Again, inspect the Bethe lattice with $z = 3$ in Fig. 2.6 to confirm that 2 $p_c = 0.5$ would marginally stop a propagating signal/disease.

We here want to characterize the properties of connected clusters of occupied sites close to the percolation threshold p_c. To do this we will first calculate two different exponents, and then use these to determine a third exponent.

Mini tutorial:
When a quantity $\mathcal{Q}$ scales as $\mathcal{Q} \sim (p_c - p)^{-\gamma}$, does a larger value of γ then imply that $\mathcal{Q}$ tends to become relatively larger or smaller when one approaches the critical point p_c, that it, when the limit $\lim_{p \to p_c} \mathcal{Q}$ is approached?

First consider the scaling of the mean cluster size $S(p) \propto |p_c - p|^{-\gamma}$ to which **an occupied site** belongs (scaling for $p \to p_c$, with p in the vicinity of p_c, $p < p_c$). $S(p)$ can be calculated by starting at an occupied site and then adding the contribution from each of the three sub-branches:

$$S(p) = 1 + 3 \cdot T\,. \tag{2.7}$$

Here T is the average contribution from one of the sub-branches. This contribution can be determined self-consistently from

$$T = p \cdot (1 + 2T)\,, \tag{2.8}$$

because T only receives a contribution from the first site of the sub-branch if this is occupied (with probability p), multiplied with the contribution from the two subsequent sub-branches when the first site is occupied.

$$T = \frac{p}{1-2p} \quad \text{for} \quad p < p_c = 1/2\,. \tag{2.9}$$

Inserting into Eq. 2.7, we therefore obtain the mean cluster size as

$$S(p) = 1 + 3T = \frac{1+p}{2(p_c - p)} \propto (p_c - p)^{-1} \ , \tag{2.10}$$

which gives the critical exponent $\gamma = 1$ for the Bethe lattice. Thus, each time the distance to the critical point is halved, the average cluster size $S(p)$ is doubled.

Now we are nearly in a position to calculate an intuitive characteristics of the near percolating system, namely the frequency distribution of cluster sizes, $n_s(p)$, that is, the count of clusters of size s, given a value of p. The philosophy is that we already know the scaling of the average cluster size $S(p)$ and will then supplement this with the cutoff of the cluster sizes. As $S(p)$ also depends on the scaling in the cluster size distribution, we will be able to obtain this scaling exponent.

If we define a perimeter site of a cluster as a nearest neighbor site that is unoccupied, then the number of clusters of size s formally is

$$n_s(p) = \sum_t g_{s,t} p^s (1-p)^t \ , \tag{2.11}$$

where $g_{s,t}$ is the number of different lattice configurations with size s and perimeter number t. The sum runs over all perimeter sizes t and weights each by the corresponding number of different configurations $g_{s,t}$.

For a Bethe lattice with $z = 3$ the perimeter number $t = 2 + s$. This may be seen by induction: start with a cluster of size $s = 1$ that obviously has three perimeter sites ($t = 3$). For each added site, one looses this site's contribution to the perimeter, but gains two new perimeter sites further out in the lattice. Thus, each site added to the cluster will yield a net contribution of one added perimeter site. Thus,

$$t = 2 + s \ . \tag{2.12}$$

Accordingly, the configuration count $g_{s,t}$ in the above sum only contains non-zero values when $t = 2 + s$:

$$n_s(p) = g_{s,2+s} \cdot p^s \cdot (1-p)^{2+s} \ , \tag{2.13}$$

where $g_{s,s+2}$ is the number of cluster configurations of size s. This is a complicated function that we will "scale out". To this end, consider the following ratio, which is independent of $g_{s,s+2}$:

$$\begin{aligned} \frac{n_s(p)}{n_s(p_c)} &= \left(\frac{1-p}{1-p_c}\right)^2 \left(\frac{p}{p_c} \cdot \frac{1-p}{1-p_c}\right)^s \\ &= \left(\frac{1-p}{1-p_c}\right)^2 \exp\left(s \cdot \ln(\frac{p}{p_c} \cdot \frac{1-p}{1-p_c})\right) \\ &= \left(\frac{1-p}{1-p_c}\right)^2 \exp\left(s \cdot \ln(4p - 4p^2)\right) \ . \end{aligned}$$

We here want to use a Taylor expansion around the critical point, $p_c = 1/2$ to obtain the leading-order contribution in the small parameter $p_c - p$ (*compare* Sec 1.3.4). Denote the argument of the natural logarithm in Eq. 2.14 as $f(p) \equiv 4p - 4p^2$. Note that $f(p_c) = 1$, $df/dp(p_c) = 0$ and $d^2f/dp^2 = -8$, giving $f(p) \approx 1 - 4 \cdot (p - p_c)^2$. We therefore get

$$\begin{aligned} \frac{n_s(p)}{n_s(p_c)} &\sim \left(\frac{1-p}{1-p_c}\right)^2 \cdot exp\left(s \cdot ln(1 - 4 \cdot (p - p_c)^2)\right) \\ &\sim \left(\frac{1-p}{1-p_c}\right)^2 \cdot exp(-\frac{s}{s_{p-p_c}}) \end{aligned} \tag{2.14}$$

Thus

$$s_{p-p_c} \equiv \frac{-1}{ln(1 - 4 \cdot (p - p_c)^2)} \sim \frac{1}{4} \cdot (p - p_c)^{-2} \propto |p - p_c|^{-1/\sigma} \; . \tag{2.15}$$

We refer to s_Δ as the "cut-off cluster size" for $p \sim p_c$, as s_Δ sets a scale in the exponential in Eq. 2.14. Cluster sizes in excess of s_Δ will virtually never be observed. The exponent $\sigma = 1/2$ characterizes the scaling of the cluster size cut-off for percolation in the Bethe lattice for p in the vicinity of p_c.

To summarize, we have learned that the mean cluster size $S(p)$ increases proportional to $1/(p_c - p)$ near the critical point, whereas the maximal cluster size s_{p-P-c} increases much faster, namely as $1/(p_c - p)^2$. As $p_c - p$ decreases, the maximal size of clusters deviates more and more from the average cluster size. *This means one have to come up with a distribution that connect a diverging difference between the average and the maximal as p gets closer to p_c. I.e. if the maximal divided by mean is a factor 10 at one value of $p_c - p$, it will be a factor 100 for a ten times smaller $p_c - p$. To bridge this diverging scales we need a scale free distribution, that is we need a scale-free distribution of the cluster sizes.*

Mini tutorial:
How much easier is it to find a cluster of size s than a cluster of size 1?

While we have now computed the maximum cluster size, we have not yet obtained the shape of the cluster size distribution. To make progress, consider therefore the actual distribution of cluster sizes for p very close to $p_c = 1/2$; which we assume takes the form

$$n_s(p) = n_s(p_c) \cdot exp(-s/s_{p_c-p}) \propto s^{-\tau} \cdot exp(-s/s_{p_c-p}) \; . \tag{2.16}$$

We have hence incorporated the cutoff and assume a power law distribution for all cluster sizes below this cutoff.The is, we assume that the only relevant scale is the cutoff scale set by $p - p_c$.

In other words, we assume that $n_s(p_c)$ is proportional to $1/s^\tau$. This power law form is consistent with the fact that clusters can become very large when

$p \to p_c$. Now assuming the above power law with cutoff s_{p_c-p}. We again consider the average size of a cluster starting from a random site. Then the chance to select a cluster of size s is proportional to $sn_s(p)$). When summing over all clusters the average cluster size becomes:

$$\begin{aligned} S(p) &\propto \sum_{s=1}^{\infty} s^2 n_s(p) \propto \int_1^{\infty} s^{2-\tau} \cdot \exp(-s/s_\Delta) \cdot ds \\ &\approx s_\Delta^{3-\tau} \cdot \int_0^{\infty} z^{2-\tau} \cdot \exp(-z) \cdot dz = const \cdot s_\Delta^{3-\tau} \ . \end{aligned} \tag{2.17}$$

where we use the continuous limit because we anyway are concerned by big clusters. Using that we already deduced the cutoff scaling

$$s_\Delta \propto (p - p_c)^{-2}$$

we get that

$$S(p) \propto (p - p_c)^{2\tau - 6}$$

From earlier we know that this should scale as $(p - p_c)^{-1}$ (see eq. 2.10). Therefore

$$2\tau - 6 = -1 \to \tau = \frac{5}{2} \ , \tag{2.18}$$

thus obtaining the cluster size distribution:

$$n_s(p) \propto \frac{1}{s^{5/2}} \cdot \exp(-s \cdot |p - p_c|^2) \ . \tag{2.19}$$

The above procedure for deducing relations between an exponent for cluster sizes τ and the two exponents for respectively the average size ($\gamma = 1$) and the cut-off size ($1/\sigma = 2$) can be generalized to percolation clusters in two or three-dimensional percolation. Then

$$\tau = 3 - \sigma \cdot \gamma \tag{2.20}$$

is obtained. In general the scaling of cluster sizes (eq. 2.19) is one example of a power law distribution augmented by a cut-off function that define the behaviour at and above a certain scale set by the distance from chosen p to the critical p_c The shape of the cutoff function f could be simulated numerically, using:

$$n(s) = s^{-\tau} \cdot f(s \cdot |p - p_c|^\sigma) \tag{2.21}$$

$$\to s^\tau \cdot n(s) = f(x) \ \text{ with } \ x \equiv s \cdot |p - p_c|^\sigma \ . \tag{2.22}$$

Thus one cold plot $s^\tau \cdot n(s)$ as function of $x = x(s, p)$ ($= s \cdot |p - p_c|^\sigma$ in our example) for different guessed values of p_c and σ, until the curves for different p fall on top of each other. This numerical approach is called *data collapse* and allow us to estimate both p_c, σ and in fact also the cutoff function f.

Mini tutorial:
What is cluster size distribution when selecting random points and then counting cluster sizes associated to each of these points? (That is, argue why there is an additional factor s for the hereby selected cluster sizes).

Mini tutorial:
Map the cluster size distribution for the critical Bethe-lattice to a first return of a random walker.

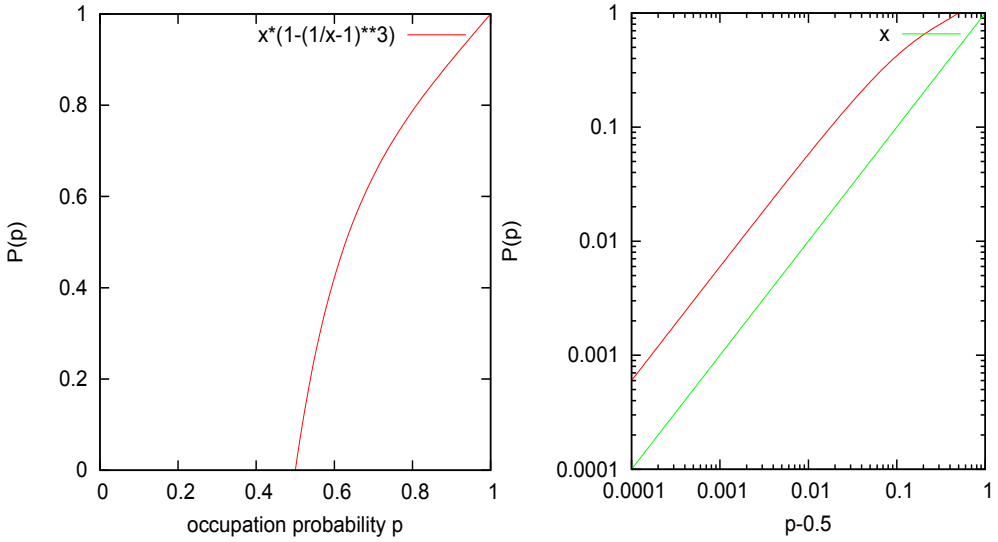


Figure 2.7: **Scaling for the percolation problem.** Strength $P(p)$ of the infinite cluster in Bethe lattice with $z = 3$, where P is the fraction of sites contained in the infinite cluster and p the occupation probability. The right hand side shows a typical scaling plot (note the double-logarithmic axes), allowing one to extract the behavior as one move very close to $p_c = 0.5$ (for question 2.2).

Beyond the critical exponents discussed above, there are further critical exponents in percolation — most importantly the correlation length exponent ν. The correlation function $g(r)$ is the probability that one occupied point is within the same cluster as another another point at linear distance r,

$$g(r) \propto exp(-r/R_{corr}) \ , \tag{2.23}$$

where, for pedagogical reasons, "exp" is used as a name for a function with a scale (in fact the actual cutoff function have another form). The correlation length scales as

$$R_{corr} \propto |p - p_c|^{-\nu}. \tag{2.24}$$

The exponent is $\nu = 4/3$ for two-dimensional percolation, implying that the linear dimension across the largest cluster grows quite fast as one approaches

the critical point from below [1)].

Exponent	Ising 2 dim	Perc. 2-dim.	Perc. 3-dim.	Perc. Bethe lattice
γ	7/4	43/18	1.8	1
β	1/8	5/36	0.41	1
σ		36/91	0.44	1/2
τ		187/91	2.19	5/2
ν	1	4/3	0.88	1/2

Table 2.1: **Exponents for the Ising model in two dimensions and for percolation in two, three and infinite dimensions (Bethe lattice).** Notice that for the Ising model the exponent is relative to varying $|T_c - T|$ whereas $|p_c - p|$ is the variable in percolation. In Ising model the order parameter was the magnetization, in percolation it is the probability to belong to the largest cluster (thereby order parameter in Ising model is finite for $T < T_c$, whereas order parameter in Percolation is > 0 for $p < p_c$). Chin-Kun Hu, 1984 suggested that the Ising model is related to bond percolation with bond probability $p = 1 - exp(-2J/k_BT)$. For Bethe lattice ν is assumed to be the same as for high dimensional percolation, see also argument by [16].

Questions:

2.3) Go through the following argument associated to the effective order parameter $P(p)$ for percolation on a Bethe lattice. That is we are now above p_c and want to explore how the infinite cluster gets denser as we moves into the high density region at $p > p_c$. This would be analogous to explore the order parameter (magnetization) in the Ising model as we lower T below T_c.

The strength of the infinite cluster $P(p)$ is the probability that an arbitrary point belongs to it. The critical exponent β is defined by $P(p) \propto |p - p_c|^\beta$ for p close to but above p_c.

For $p > p_c$ the largest cluster spans across the system and $P(p)$ is finite (that is, larger than zero). We want to calculate how $P(p)$ vanishes as one approaches the critical point p_c from above.

$$\begin{aligned} P(p) &= (Probability\ that\ site\ occupied) \\ &\quad \cdot (Probability\ that\ at\ least\ one\ neighbour\ leads\ to\ infinity) \\ &= p \cdot (1 - Q^3) \end{aligned}$$

[1]Percolation is characterized by a number of exponents, $1/\sigma$, γ, ν, τ and β (see question below). We have here only shown one relation between these exponents, namely the relation between average cluster size and the distribution of cluster sizes: $\gamma = (3-\tau)/\sigma$. In addition, the scaling of the order parameter (density of the largest cluster above p_c) is related to the cluster size distribution by $\beta = (\tau - 2)/\sigma$. Furthermore, the correlation length exponent (that teaches us about the linear dimension of a cluster) is $\nu = (\tau - 1)/(\sigma d) = 1/(\sigma D_f)$. Here d is the dimension of the system of all points and D_f is the dimension of the largest cluster at the critical point (see later for definition of dimensions). This last relation tells us that the extent of the correlation is given by the mass of the largest cluster, corrected by a factor that takes into account that this mass is distributed in more than one dimension.

where Q is the probability that an arbitrary neighbour site is not connected to infinity. For $z = 3$ then:

$$\begin{aligned} Q &= (Probability\ site\ empty) \\ &+ (Probability\ site\ occupied) \cdot (Probability\ no\ subbranch\ leads\ to\ infinity) \\ &= (1-p) + p \cdot Q^2 \rightarrow \\ Q &= \frac{1}{2p} \cdot \left(1 \pm \sqrt{(2p-1)^2}\right) \rightarrow \\ Q &= 1\ or\ \ Q = \frac{1}{p} - 1 \end{aligned} \tag{2.25}$$

Prove this by insertion of the 2 solutions! The strength $P(p)$ is therefore

$$P(p) = p \cdot \left(1 - (\frac{1}{p} - 1)^3\right) = \frac{p^3 - (1-p)^3}{p^2}$$

which is an increasing function of p around $p_c = 1/2$. That is, it can be expanded around the critical point $p = p_c = 1/2$ (where $P(p) = 0$) using

$$\begin{aligned} \frac{dP(p)}{dp} &= \frac{3}{p^2}(p^2 + (1-p)^2) - \frac{2}{p^3}(p^3 - (1-p)^3) \\ &= 3 \cdot (1 + (\frac{1}{p} - 1)^2) - 2 \cdot (1 - (\frac{1}{p} - 1)^3) = 6 > 0 \end{aligned}$$

for $p = p_c = 1/2$ Thus $P(p) \sim 6 \cdot (p - p_c)$ for p above but close to $p_c = 1/2$. Accordingly, the critical exponent $\beta = 1$ for the Bethe-lattice (*see* Fig. 2.5). Thus, it becomes more and more difficult and find the infinite cluster as one approaches p_c (from above), and in fact halving the distance to $p_c = 1/2$ makes it twice as difficult to find one of the points in this infinite tree.

QLesson: At critical conditions it becomes infinitely difficult to find the largest cluster by randomly selecting points. However, this cluster anyway spans the entire system (reaches "from one end to the other").

2.4) Compare exponents in Table 2.1. Discuss the percolation exponents in two and three dimensions relative to those of the Bethe lattice.

Discuss the exponents for the 2d Ising model and 2d percolation.

What does it mean that the mean cluster size exponent is smaller for the Ising model than for corresponding percolation?

What does it mean that the correlation length exponent is smaller for the Ising model than for corresponding percolation?

(QLesson: $p = p_c$ raised to a big exponent means that objects diverge to larger size than if the exponent were small. Also, the cluster size distribution becomes steeper with dimension.)

2.3 Fractal Dimensions

2.3.1 Large objects with zero density

The way the mass (or volume) of an object (a set of points) scales with the object's length can be used to define the object's dimension. More concretely,

we want to introduce the dimension D of a set of points, by

$$M(l) \propto l^D \tag{2.26}$$

as we consider still larger scales l. Here M is the amount of points (=mass) that is within a box of linear dimension l and the equation expresses the extent to which M grows as one considers bigger and bigger parts of the object. When the object is compact lump of matter in three dimensions, the mass within l simply scales with l^3, reflecting a dimension of three. However, a object may often have holes and irregular boundaries, as for example with snow crystals. In that case the *fractal dimension* of the object could be smaller than three (or whatever dimension it is embedded in).

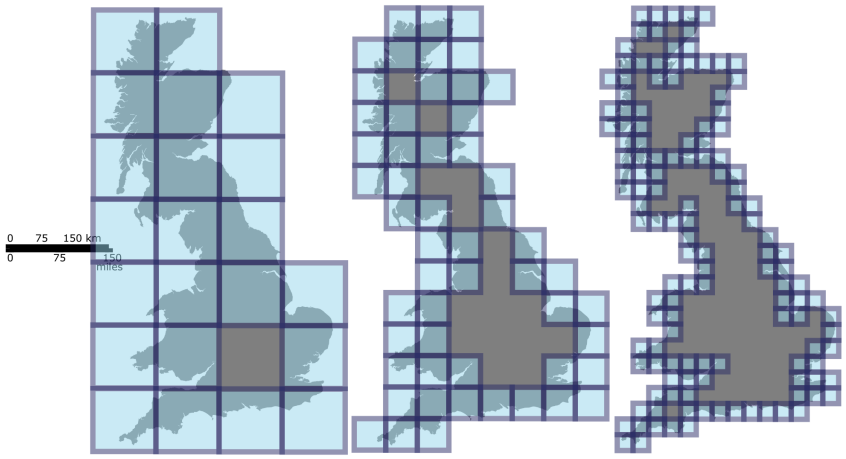


Figure 2.8: **Fractal dimension of Great Britain.** Illustration of the box counting method on the coast of Great Britain, taken from Wikipedia, Prokofiev - Own work, CC BY-SA 3.0, `https://commons.wikimedia.org/w/index.php?curid=12042116`

Let us now return to the percolation problem and the largest cluster, discussed in the previous section. If we were to describe the dimension of this largest cluster, we might investigate larger and larger lattices, and explore the density of the percolating cluster. That is, count how the fraction of lattice sites it occupies decreases as the lattice area is increased (since the dimension of the cluster is smaller than 2, the density will decrease).

In some real world fractals, one measures them with smaller and smaller "measuring stick" $\epsilon \to 0$. If they becomes "large" as the "measuring stick" gets smaller, the object is a fractal. Or more precisely, if the number of boxes needed to cover the object grows with some non-integer power law as function of $1/\epsilon$, then they are fractal.

The fractal dimension D (box dimension) is calculated by the scaling of the number of boxes

$$N(\epsilon) \propto \frac{1}{\epsilon^D} \tag{2.27}$$

needed to cover the object (the set of points) as function of the box size ϵ used. This type of box covering is illustrated in Fig. 2.8. Given $N(\epsilon)$ for a range of different box sizes ϵ, the dimension is then calculated as

$$D = \lim_{\epsilon \to 0} \left(\frac{\log(N(\epsilon))}{\log(1/\epsilon)} \right) \ . \tag{2.28}$$

This can be accomplished empirically by examining the slope of a log-log plot of N as function of $1/\epsilon$. For example, consider a two-dimensional object. As one increases L, the mass within distance L scales as L^2. Reversely, one may subdivide the system into boxes of size ϵ. The number of boxes needed to cover the object then scale as $(L/\epsilon)^2$ where now L is fixed and ϵ is reduced. The coast of Great Britain (Fig. 2.8) is however characterized by dimension $D = 1.2$, whereas the more convoluted coast of Norway needs $D = 1.5$.

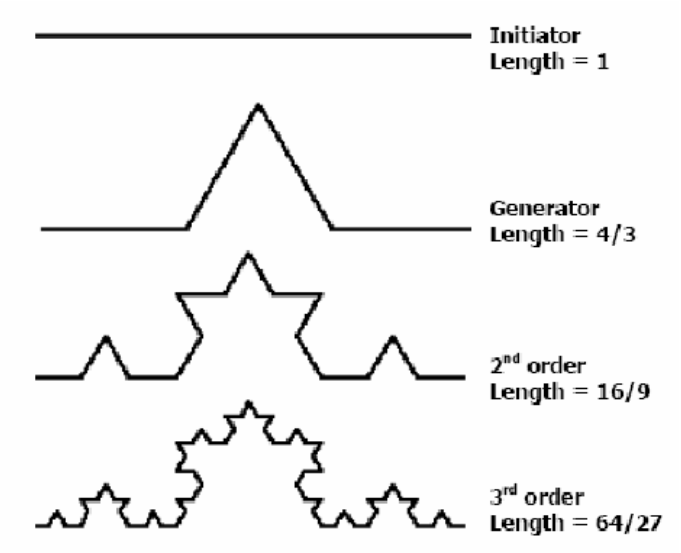


Figure 2.9: **Koch curve.** Illustration of how the length of the line increases as one considers finer and finer detail: When the "measuring stick" is three times smaller, the total length is four times larger. Thus the dimension $D = -\ln(4)/\ln(1/3)$.

Consider now percolation in two dimensions. The mass of the largest cluster becomes larger as we approach p_c, scaling in size with the distance $p_c - p$ to the critical point as

$$M \propto (p_c - p)^{-1/\sigma} \ . \tag{2.29}$$

Similarly, the linear dimension of the largest cluster is given by its correlation length

$$l \sim R \propto (p_c - p)^{-\nu} \Rightarrow (p_c - p) \propto l^{-1/\nu} \, . \tag{2.30}$$

Thus we get the mass of largest cluster in terms of the of its linear extension

$$M \propto l^{1/\nu\sigma} \, . \tag{2.31}$$

The (fractal) dimension of the largest cluster at p_c therefore is

$$D = \frac{1}{\nu\sigma} = \frac{91}{48} < 2 \tag{2.32}$$

at the critical point. Each of the smaller clusters would have the same dimension. I.e. at the critical point there is a biggest cluster, and also some slightly smaller clusters. On smaller scales, one cannot determine whether one is in the biggest or in a somewhat smaller but still big cluster. As a consequence, the dimension of any reasonably sized cluster is $D < 2$. This could be compared with the dimension of all the sites in all the clusters, which is clearly $D_{tot} = d = 2$. This is because the density of points is finite (it is equal to p).

Mini tutorial:
What is the density of the percolating cluster at p_c in an infinitely large two-dimensional system?

The mass of the largest cluster M connects the exponent β to the dimension of the largest cluster. That is, the mass should be calculated for distances up to the size set by the correlation length, with densities deduced from the exponent β. Thus

$$M \sim l^d P(p) \propto l^d (p - p_c)^\beta \propto (correlation\ length)^{d-\beta/\nu} \tag{2.33}$$

where d is the embedding dimension ($d = 2$ for percolation in two dimensions) and where the length l scales as the correlation length as we move p closer to p_c. Thus, the fractal dimension of the percolating cluster is also given by $D = d - \beta/\nu$.

From the two equations $D = 1/(\nu\sigma)$ and $D = 2 - \beta/\nu$ (in two-dimensional percolation) one obtains

$$\nu = (\beta + 1/\sigma)/2 \, . \tag{2.34}$$

Thus, for larger cluster size cutoff $1/\sigma$ one also has larger correlation length ν.

Mini tutorial:
Given that $\beta = 5/36$ and $\sigma = 36/91$ for two-dimensional percolation, calculate ν.

The percolating cluster in two-dimensional lattice had dimension $D = 91/48$. If one instead considered the small sub-part of this cluster that would

carry a current if one applied a voltage drop over the full cluster, this subpart (called "backbone") would have dimension $D_{backbone} = 1.13$ (*see* S. Havlin lecture notes). And an even smaller subset of the largest cluster consists of the sites along this backbone, that if broken, would break the whole infinite cluster up. These are called red bonds, and have dimension less than 1 [2])

First fractal relation: Sometimes it is useful to consider the intersection between a line (or plane) and a fractal object. In general the fractal dimension of the intersection of two independent objects of dimension D_A respectively dimension D_B that both are part of the same space with dimension D_{space} is calculated by

$$D(A \; and \; B) = D(A) + D(B) - D_{space} \tag{2.35}$$

(S. Miyazima and H.E. Stanley, Phys. Rev. B 35, 8898, 1987). This is proven by covering the space with $1/\epsilon^{D_{space}}$ boxes, of which $1/\epsilon^{D(A)}$ cover object A, and $1/\epsilon^{D(B)}$ cover object B (for simplification I just set total space size to 1). Thereby the probability that one box contain something from for example object A is $P(A) = (1/\epsilon^{D_A})/(1/\epsilon^{D_{space}})$. Assuming independent coverage, the intersection fraction is multiplying probabilities:

$$\frac{\epsilon^{D_{space}}}{\epsilon^{D(A \; and \; B)}} = \frac{\epsilon^{D_{space}}}{\epsilon^{D(A)}} \cdot \frac{\epsilon^{D_{space}}}{\epsilon^{D(B)}} \tag{2.36}$$

from which eq. 2.35 is found. Notice that it is important that A and B are independent, and if one chose to large a embedding space D_{space} to large, then A and B will not be independent.

For example, consider a random walk in one dimension. That can be seen as a sequence of random steps along one axis, plotted as function of time along the other axis. The dimension of such a 1d random walk embedded in a two-dimensional space-time plot is $D(RW) = 1.5$ (not proven here). Given this then its intersection with a one-dimension line leaves a fractal dust with dimension $D(dust) = 1.5 + 1 - 2 = 0.5$. As we will see later then the distance between these dust particles are power law distributed[3].

Mini tutorial: Calculate the dimension of the intersection between a line and the infinite cluster in two-dimensional percolation at the critical point.

Mini tutorial: Assume the 1.5 dimensional coastline of Norway comes about as the intersection of a 2 dimensional water surface with a mountain range with

[2]The number of these so-called red bonds scales as $n_{red} \propto \frac{1}{p-p_c} \;\; for \;\; p > p_c$ (Coniglio, 1982) and thus scale with correlation length $R \propto (p-p_c)^{-\nu}$ as $n_{red} \propto R^{1/\nu}$, i.e. $D_{red} = 1/\nu = 3/4$ for two-dimensional percolation.

[3]In general a colored random walk in one dimension with Hurst exponent H (see econophysics chapter) has dimension $D = 2 - H$. Thus, a walk with Hurst exponent 1 correspond to ballistic walk and has dimension $D = 1$. A walk with pink noise ($1/f$ power spectra) have Hurst exponent $H = 0$ and dimension $D = 2$

dimension D. What is D of this "rough" mountain range?

2.3.2 Fragmentation

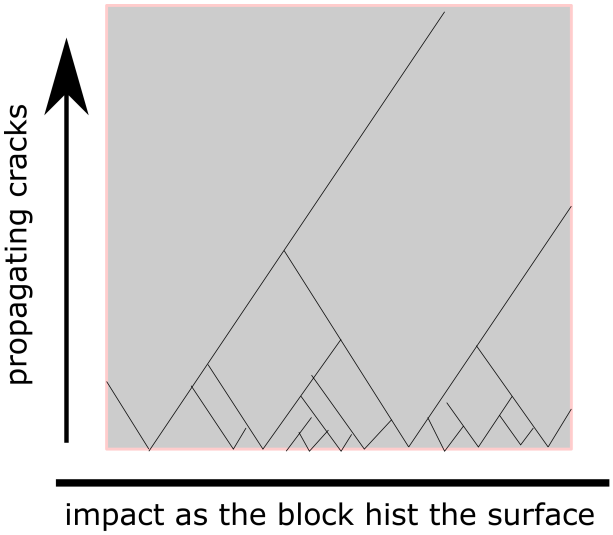


Figure 2.10: **Simple fragmentation model.** Crack propagation on a plane, but created by impact on its side. Cracks coarsen as $1/t^{0.5}$.

Let us now consider the above equations in light of a simple model for fragmentation. In some approximation this can be rephrased as an initial impact with creation of a lot of cracks, followed by subsequent merging of straight cracks, reflecting the fact that when two cracks meet then the first crack stops propagation of the second one. This is illustrated in Fig. 2.10, with similar dynamics in some cellular automata models, see Fig. 2.11.

Imagine that a two-dimensional square object is excited at one of its 1-dimensional surfaces, and cracks spread inwards. That is when a plane of glass hit the floor a lot of cracks start along the impact surface. When two cracks meet, one will have arrived first and the last one will be annihilated (it cannot crack across the other crack, because the glass is disconnected). We now want to calculate the resulting size distribution of fragments.

Second fractal relation: For the fragments created by linear cracks in Fig. 2.10, the dimension of each fragment is $D = 2$, whereas the dimension of all fragments is also $D_{tot} = 2$. But there is a third number that quantifies the spatial distribution, and this is the dimension of fragments when each are counted as one point. Because we assume that one initiate cracks with constant density along the impact surface and each fragment has a point on

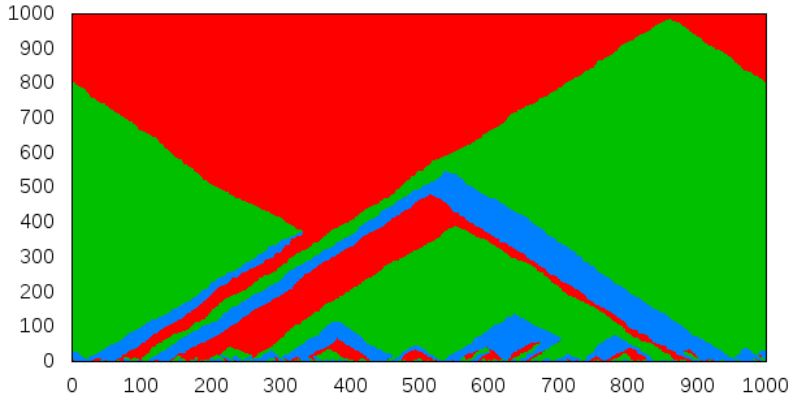


Figure 2.11: **Coarsening in a rock-paper-scissor dynamics in one dimension.** One initially seeds each site with one of the 3 different species types, a "rock", a "scissor" or a "paper". They can grow into each other defined by normal rules of the rock-paper-scissor game. Populations of one species can be separated by other species, forming antagonistic boundaries that moves either left or right. Boundaries between species move left or right, and is annihilated when meeting other boundaries. I.e. when a boundary between a rock and a scissor meets a boundary between a scissor and a rock, they annihilate and an area of pure rock forms. This can then later be eaten by a invading "paper".

this 1-dimensional impact, the number of fragments can be characterized by the dimension $D_{num} = 1$.

From the above three different dimensions one can now calculate the fragment size distributions as done by Greg Huber [17]. I.e. assume that

$$n(s) = \frac{1}{s^\tau} \cdot f(\frac{s}{L^D}) \tag{2.37}$$

where f is some cutoff function that drops to zero when the cluster becomes so big that its linear dimensions is comparable with the whole system L (Because D is the dimension of the cluster, then s feels the boundary exactly when it reaches the size L^D).

Now if we understand $n(s)$ as the probability that a cluster has size s, and the total number of clusters is $L^{D_{num}}$, then

$$\begin{aligned} L^{D_{num}} \cdot \int_1^{L^D} s \cdot n(s) ds &= L^{D_{tot}} \\ L^{D_{num}} \cdot [s^{2-\tau}]_1^{L^D} &= L^{D_{tot}} \\ L^{D_{num}+(2-\tau)D} &= L^{D_{tot}} \\ D_{num} + (2-\tau) \cdot D &= D_{tot} \\ \tau = 2 - \frac{D_{tot} - D_{num}}{D} &, \end{aligned} \tag{2.38}$$

where we assumed that only the upper end of integral counted, i.e. that $\tau < 2$. If τ is larger, one instead has to focus on the lower end of integral.

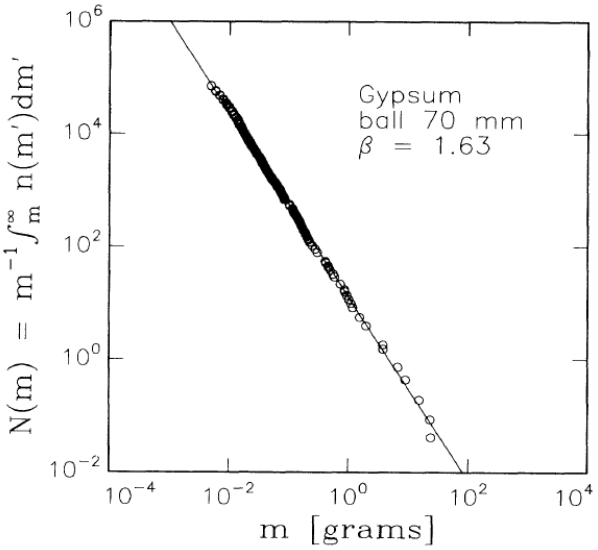


Figure 2.12: **Fragment size distribution.** Measured from fragmentation of gypsum blocks that were dropped on the floor.

For the fragments in Fig.2.10, then $D = 2$, $D_{tot} = 2$ and $D_{num} = 1$ giving fragment size distribution

$$n(s) \propto \frac{1}{s^{3/2}} \tag{2.39}$$

Generalizing to fragmentation of a three-dimensional block, with conical cracks initiated randomly at points on a two-dimensional surface, then $D = 3$, $D_{tot} = 3$ and $D_{num} = 2$. This gives $\tau = 2 - (3-2)/3 = 5/3$, or

$$n(s) \propto \frac{1}{s^{5/3}} \, . \tag{2.40}$$

This is close to the fragments size distribution $1/s^{1.63}$ that was measured by Oddershede et al. [18], see Fig. 2.12.

Noticeably, meteors are distributed with power law of about $1/M^{1.8}$, not far from the above fragmentation exponent. To put this in perspective, then the probability that earth is hit by a meteor larger than mass M scale as $P(> M) \propto 1/M^{0.8}$. The exponent then implies that we should expect a meteor of more than 10% in diameter of the famous 10km diameter meteor from Yucatan every $\sim 64,000,000 \cdot (1/1000)^{0.8} = 250,000$ years or so (assuming that the meteor 64 million years ago was a typical event on that timescale). Notice that the estimate uses the cumulative distribution.

Mini tutorial: Why should one use the cumulative and not the differential estimate above? How does one use the differential distribution?

Mini tutorial: Estimate how often a meteor of diameter larger than 100 m hits earth. And larger than 100 km?

Questions:
2.5) Consider dust on a line, with points distributed with dimension $D_{num} = D_{dust}$. Show that the distribution of length between the dust follows the distribution $n(l) = 1/l^{1+D_{dust}}$.
Qlesson: The larger the dimension of dust, the narrower the distribution of intervals between it. Notice that the average length between dust particles diverges with system size (explain that).
2.6) What is the fractal dimension of the intersection of a line and a two-dimensional percolating cluster at the critical value p_c?
Qlesson: Fractal dimension is obviously (?) smaller than 1. Why?
2.7) Formulate an automaton that would mimic the crack annihilation model above. Simulate it, and calculate the fragment size distribution starting from random crack initiation at one surface. (Hint: Use three numbers, one to give direction.)
Qlesson: Cellular automata can be used in many problems. Can you find a continuum equation that describes the dynamics of the crack propagation? (I could not)
2.8) Consider the dimension equation for $\tau > 2$ where integral in $L^{D_{num}} \cdot \int_1^{L^D} s \cdot n(s)ds$ is dominated by lower limit. Argue for the identity

$$\begin{aligned} L^{D_{tot}} - L^D &= L^{D_{num}} \int s^{1-\tau} f(s/L^D) ds \\ &= L^{D_n um} \cdot (const - L^{D\cdot(2-\tau)}) \end{aligned}$$

where the constant depends on small scale cutoff and the subtracted part comes from using $f = 1 + (f-1)$. Use that $D_{tot} \geq D$ (the dimension of the whole cannot be smaller than the dimension of one fragment), to deduce that when $\tau > 2$ then $D_{tot} = D_{num}$ and

$$\tau = 1 + \frac{D_{num}}{D} \tag{2.41}$$

whereas $D_{tot} = D_{num}$ implies that τ cannot be smaller than 2.
QLesson: The exponent 2 is special, when exponents are larger than 2 then the small clusters contribute to the average whereas the big clusters do not.

2.4 Directed percolation

Directed percolation (DP) is the first dynamical process in these notes. DP can can intuitively be pictured as an infection process. Sites that are close to an infected site may become infected in the subsequent timestep. Importantly, new infections cannot come by themselves: The "dead" or uninfected state is absorbing. As a consequence a site becomes inactive unless it has an active neighbor (or is active itself). When a site have such active neigborhood, the probability to become active, at next time-step is then p. This probability is analogous to our percolation parameter from before.

When p is small all sites will eventually turn to the non infected state. On the other hand, if p is large, then all sites will eventually becomes active (infected) nearly all the time.

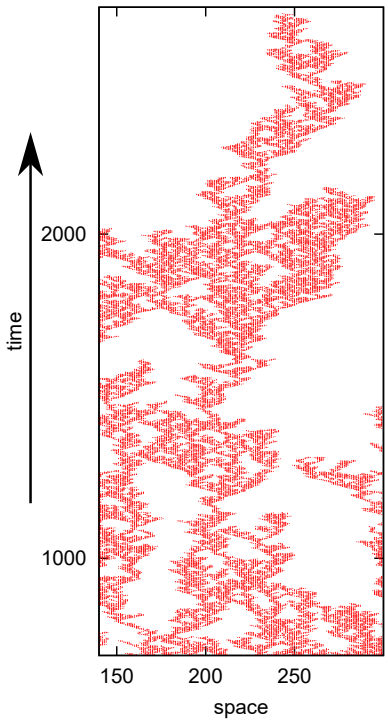


Figure 2.13: **Directed percolation on a square lattice.** Each site can give offspring to itself and its two nearest neighbors. This process has two correlation lengths, one along the time axis $T(p) \propto 1/|p_c - p|^{1.7}$ and a shorter one along the spatial dimension $X(p) \propto 1/|p_c - p|^{1.1}$. In addition, the process is characterized by one density exponent β, with the density of active sites at long times after initiation, $\rho \propto |p - p_c|^{\beta}$, $\beta = 0.27$. For directed percolation the same scaling is also found for the probability that a random point belongs to the infinite cluster.

Directed percolation is also often associated to the following stochastic equation (Reggeon field theory):

$$\frac{dn}{dt} = \frac{d^2 n(x,t)}{dx^2} + b \cdot n(x,t) - d \cdot n(x,t)^2 + \eta(x,t) \tag{2.42}$$

where the rates of birth ($b > 0$) and death ($d > 0$) are both positive and where the noise term $\langle \eta(x,t)\eta(x't') \rangle = n(x,t)\delta(x - x')\delta(t - t')$ is uncorrelated in space and time and only takes values where there is already some active sites $n(x,t) > 0$. Thus there is no noise if all is dead, only life create life.

The above equation has an absorbing state at $n = 0$, a spreading of activity through the diffusion term (d^2/dx^2) and further inhibits replication when local density becomes large, that is, $d \cdot n^2 > b \cdot n$. It will have a transition analogous to directed percolation for a critical value of b (replication rate).

Mini tutorial:

What could be the biological reason for the term $-n^2$ in the above equation?

Directed percolation has a phase transition at a critical $p = p_c$ that depend on the lattice, but the exponents characterizing the phenomenon is independent of such details. Compared with percolation, DP has one additional cutoff exponent, because the space and time axes are highly asymmetric, and correlation extends further in time ($\nu_\parallel$) than in the spatial direction ($\nu_\perp$).

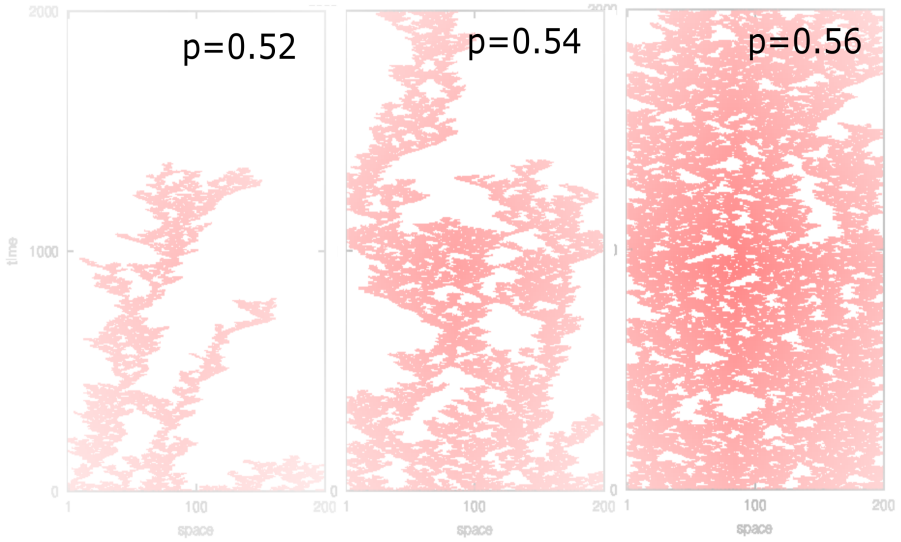


Figure 2.14: *Directed percolation of a square lattice, each site can give offspring to itself and its two nearest neighbors.*

To be more quantitative: If the percolation parameter p is below a critical threshold p_c, the propagation of live sites has a finite lifetime.

If p is above p_c, the propagation of live sites can continue forever. For p just below p_c, the time- like correlation length (lifetime) diverges

$$t_{corr}(p) \propto (p_c - p)^{-\nu_\parallel} \quad with \;\; \nu_\parallel = 1.733 \; , \tag{2.43}$$

The space-like correlation length (width) diverges as

$$x_{corr}(p) \propto (p_c - p)^{-\nu_\perp} \quad with \;\; \nu_\perp = 1.097 \tag{2.44}$$

in the 1+1 dimensional process.

The *order-parameter* exponent β is defined by the density of the infinite cluster (above the threshold p_c) and how this scale with the distance to the threshold

$$\rho \propto (p - p_c)^\beta \tag{2.45}$$

with obtained $\beta = 0.27$ in 1+1 DP, a value that is larger than the $\beta \sim 0.14$ for two-dimensional percolation but much smaller than the $\beta = 1$ for the Bethe lattice (see the table 2.1).

An exponent $\beta = 0.27$ for the infinite cluster, means that when we are at some value $p > p_c$, then the chance that a site is alive within the branching three of directed percolation, scale as $(p - p_c)^{0.27}$. Thus if one move p above a factor 16 times closer to p_c, then the chance that the site is alive is about a factor 2 smaller.

Notice, that directed percolation cannot be solved analytically, and the above scaling exponents was all obtained by extensive numerical simulations.

Mini tutorial:
What does a smaller β mean in terms of the density of the infinite cluster?

For DP above the threshold p_c, $(p - p_c)^\beta$ also counts the fraction of initiated clusters that evolves to infinity. Notice that whether a cluster continues to infinity or dies out is determined before the correlation time $t_{corr} \sim \ell_\parallel \sim \epsilon^{-\nu_\parallel}$. After this time, the cluster have grown sufficiently big to be dominated by the long time bahavior.

Further information is that critical exponents are the same below and above p_c (not proven, just believed). The three exponents are believed necessary and sufficient to completely characterize DP structures and correlations, and their values are listed in table 2.2.

Exponent	Mean-field	1-d DP	2-d DP
β	1	0.276	0.58
$\nu_\perp$	1/2	1.097	0.73
$\nu_\parallel$	1	1.734	1.2
δ	1/2	0.159	0.45

Table 2.2: **Directed percolation critical exponents.** Here, $\delta = \beta/\nu_\parallel$ quantifies survival probability from single site to time t, $P(t) \propto t^{-\delta} g(|p - p_c|t^{1/\nu_{parallel}})$.

One quite remarkable exponent of the DP network is the envelope of living sites with time: $r_{ms}(life) \propto t^\chi$ with $\chi = \nu_\perp/\nu_\parallel = 0.633$. This is seen by considering $\ell_\perp \propto (p_c - p)^{-\nu_\perp} \propto (\ell^{-1/\nu_\parallel})^{-\nu_\perp} = \ell^{\nu_\perp/\nu_\parallel} = 0.63$. This is the exponent for the width of the DP cluster as function of time. Thus the outer edge of the DP cluster makes longer excursions than an ordinary random walk have the smaller exponent $\chi = 1/2$. In case of a time-series this exponent is called a Hurst exponent, see econophysics section.

Directed percolation is an example of a spreading process and most importantly it has a phase transition between extinction and survival. This is illustrated in Fig. 2.14 that shows directed percolation on a 1-dimensional geometry, developing in time as activity spreads and/or dies out. Fig. 2.13 further examines directed percolation close to its critical point by highlighting a small sub part of the spreading process.

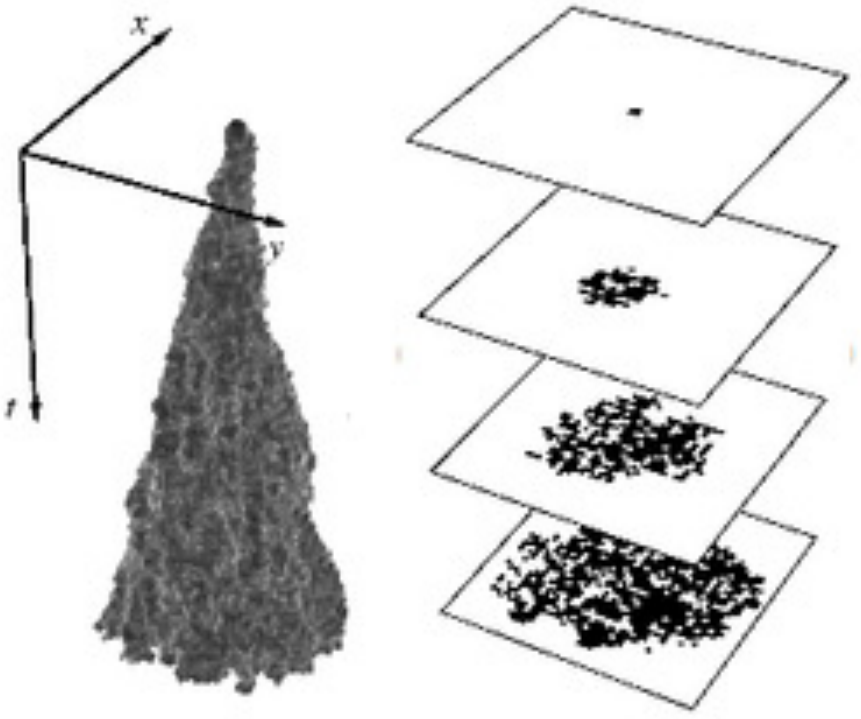


Figure 2.15: **Directed percolation in 2+1 dimensions above the critical point.** From Adam Prugel Bennet and Iain Weaver, University of Southampton.

The scaling properties of directed percolation is found in many other spreading processes, that just fulfil the criteria of being able to spread by diffusion, replicate, and die. A noticeable example is the stochastic version of the so-called rule 18 of cellular automata (Fig. 2.16). Notice that many other cellular automata could be considered. With nearest neighbors in one dimension (three sites input) there is $2^3 = 8$ inputs that each should be defined one of 2 outputs, making $2^8 = 256$ deterministic cellular automata. In 2-dimension there is even more rules, with the most famous being the game of life invented by Conway.

Mini tutorial:
What would the iterated version of rule 255 given in the formalism of Fig. 2.15?

Questions:
2.9) Simulate directed percolation in 1+1 dimensions. Estimate the critical value of $p = p_c$. Consider $p < p_c$ and determine the distributions of the size (number of accumulated life sites) of branching trees starting from a single site at time zero. See how size distribution changes as p becomes closer to p_c.
Qlesson: Life-death processes are also critical, with a cluster size distribution $1/s^\tau$

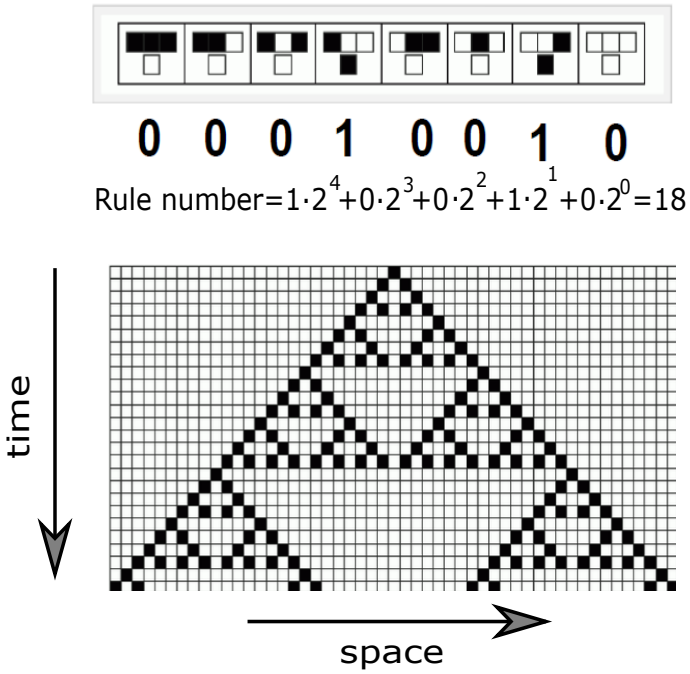


Figure 2.16: **Cellular automaton rules.** Rule 18 is an example of a cellular automaton update depend on itself and two nearest site. In rule 18 a site dies (becomes zero), except if just one of its neighbors are alive at the timestep before. The rule table is outlined in the top of the figure. The rule number is defined by the sequence 00010010, which is understood as a binary number. The deterministic update following a single live site is shown in the bottom panel. The rule can be made stochastic by assigning all site dead except live sites for the 100 and 001 neighborhood with probability p. p can then be fine-tuned to give the same large-scale properties as critical directed percolation in one dimension (same scaling exponents).

that is relatively broad $\tau < 2$.

2.10) Initiating a cluster by one active site in an infinite sea of dead sites, and propagating active sites with probability p close to p_c, one may ask what the distribution of lifetimes of the single cluster is. Convince yourself that the following answer is correct: the density at time $t = \epsilon^{-\nu_\parallel}$ is $\rho \propto \epsilon^{\beta} = t^{-\beta/\nu_\parallel}$. As the density is proportional to survival probability, then the chance that the cluster lives longer than t scales $\propto 1/t^{\beta/\nu_\parallel}$ independent of dimension d.

Qlesson: One can derive scaling of some quantities from others.

2.11) What rule number would standard directed percolation correspond to? (see Fig. 2.16)

Qlesson: Reminder that Directed percolation is a stochastic version of a cellular automata. One could also makes other rules stochastic.

2.12) Consider directed percolation, and convince yourself about the following scal-

ing arguments for some relevant dimensions. The scaling of the mass m of the infinite cluster up to a correlation length $\ell_{\|} \sim \epsilon^{-\nu_{\|}}$: $m \propto \epsilon^{\beta-\nu_{\|}-\nu_{\perp}} \sim \ell^{(\nu_{\|}+\nu_{\perp}-\beta)/\nu_{\|}}$. Thus the dimension counted with a length measured longitudinally is $1-\beta/\nu_{\|}+\chi$. Similarly the transverse dimension measure is $1-\beta/\nu_{\perp}+1/\chi = 2.33$.
Qlesson: This question emphasizes that some exponents can be deduced from the basic correlation length (ν's) and density exponents (β).

Lessons:

- Power laws are a way to quantify the many scale free phenomena in our surrounding world.
- Fractals reflects power distributions between and within spatial objects.
- This chapter presented two possible schemes for obtaining power laws: Fine-tuning $p \to p_c$ in analogy of the Ising model, and then the fast process of fragmentation.

Supplementary reading:

Christensen, Kim, and Nicholas R. Moloney. Complexity and criticality. Vol. 1. World Scientific Publishing Company, 2005.

Stauffer, Dietrich, and Ammon Aharony. Introduction to percolation theory. Taylor & Francis, 2018.

Beautiful example of cellular automata: Alexander Mordvintsev, Ettore Randazzo, Eyvind Niklasson, Michael Levin. https://distill.pub/2020/growing-ca/

Chapter 3

Self Organized Criticality

Thus the sum of things is ever being reviewed, and mortals dependent one upon another. Some nations increase, others diminish, and in a short space the generations of living creatures are changed and like runners pass on the torch of life.

– Lucretius, 94 BC - 55 BC

Figure 3.1: **Snow avalanche.** The central phenomenon in self-organized criticality are avalanches, consisting of causal sequences of toppling events that redistribute and relax local stress.

3.1 Random walks

Imagine a point particle on a one-dimensional lattice with lattice spacing ℓ. If we at each time step τ move the particle one step to the right or left at equal probability, then the position of the particle at time t is

$$x \;=\; \sum_{i=1}^{t/\tau} \eta(i)\;, \tag{3.1}$$

where $\eta(i)$ takes values $\pm\ell$ randomly. Considering the ensemble averaged square of the position (averaged over many copies of a random walker starting at x=0 at time $t = 0$)

$$\begin{aligned}
\langle x^2\rangle &= \langle \sum_{i=1}^{t/\tau}\eta(i)\sum_{j=1}^{t/\tau}\eta(j)\rangle \\
&= \sum_{i,j=1}^{t/\tau}\langle\eta(i)\eta(j)\rangle \\
&= \sum_{i=1}^{t/\tau}\langle\eta(i)\eta(i)\rangle \\
&= \sum_{i=1}^{t/\tau}\ell^2 = (t/\tau)\ell^2 = \left(\ell^2/\tau\right)t = 2\cdot D\cdot t\;,
\end{aligned} \tag{3.2}$$

where we use that steps at different times are uncorrelated (for example $\langle\eta(1)\eta(2)\rangle = ((+1)\cdot(+1)+(+1)\cdot(-1)+(-1)\cdot(+1)+(-1)\cdot(-1))/4 = 0$). Here D represents the diffusion constant, equal to the step size squared, divided with the time τ that it takes to move one step (or a velocity times the length before velocity is randomized). See Fig 3.2 where the walkers are walking up or down along the $x-$axis. Importantly the dimensions of the diffusion constant

$$[D] = \frac{meter^2}{second}\;, \tag{3.3}$$

reflecting that it results from multiplying a velocity with the distance travelled between the times when direction of movement can be changed.

From random walk to diffusion equation

To describe the average behaviour of a random walk one may use the diffusion equation. That is, a large number of random walk particles/ diffusing particle can be described in terms of the development of their density $\rho(x,t)$ where $\rho(x,t)dx$ is the probability that the particle is found between x and $x+dx$ at time t. Consider the current $J(x)$, counted in units of particles per second. The current is given by particles at positions $x-\ell/2$ and $x+\ell/2$ that move

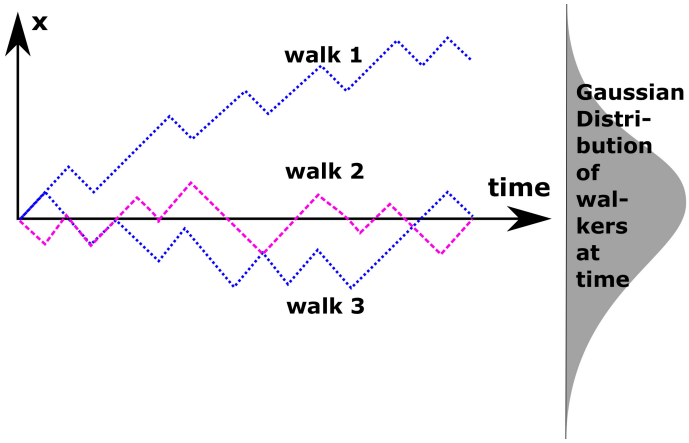


Figure 3.2: **Trajectories of random walkers.** Random walkers starting at position $x = 0$ at time $t = 0$ and ending at some position x. The final distribution of many walkers will be Gaussian in space, with width $\sigma \propto \sqrt{t}$.

across position x during the time τ. The current of particles across position x is

$$\begin{aligned} J &= \frac{1}{2}\left(\rho(x-\ell/2)\frac{\ell}{\tau} - \rho(x+\ell/2)\frac{\ell}{\tau}\right) \\ &= \frac{\ell}{2\tau}\left((\rho(x) - \frac{\ell}{2}\cdot\frac{d\rho(x)}{dx}) - (\rho(x) + \frac{\ell}{2}\frac{d\rho(x)}{dx})\right) \\ &= -D\cdot\frac{d\rho(x)}{dx} \end{aligned}$$

with $D = \ell^2/(2\tau)$. Here the factor $\frac{1}{2}$ is because only half of particles at position $x - \ell/2$ moves forward (and only half of the ones at position $x + \ell/2$ moves backward). The change in density, is subsequently given by the difference between what moves in and what moves out:

$$\frac{d\rho}{dt} = -\frac{dJ}{dx} = \frac{d}{dx}\left(D\frac{d\rho}{dx}\right) \tag{3.4}$$

A particle that starts at $x = 0$ at time $t = 0$, will at time t be found at position x with probability

$$\rho(x,t) = \left(\frac{1}{\sqrt{4\pi t D}}\right)^3 \cdot \exp(-\frac{x^2}{4Dt}) \,, \tag{3.5}$$

with root mean square displacement $\langle x^2\rangle = 2Dt$ [1]. The diffusion equation has the property that a Gaussian stays as a Gaussian, but with an ever increasing width.

[1]The diffusion equation can also be derived directly from considering many non-interacting random walkers, each performing steps of length $\delta l = 1$ during time $\delta t = 1$. Then the density distribution

$$\rho(x,t+1) - \rho(x,t) = \frac{1}{2}\rho(x-1,t) - \frac{1}{2}\rho(x,t) - \frac{1}{2}\rho(x,t) + \frac{1}{2}\rho(x+1,t) \tag{3.6}$$

First passage for random walks

A recurrent theme in many physical models is the distributions of times between zero crossings $x = 0$. For a random walk in one dimension we know that it spreads out in space proportional to the square root of the time t. Thus as time progresses, the walk visits points within this slowly expanding Gaussian several times. During time t a random walker remain within an x interval of size about $\sigma_x = \sqrt{t}$. The number of times it was at the particular position $x = 0$ during time t is therefore $\propto t/\sigma_x = \sqrt{t}$. We however want the distribution of times between subsequent visits to x, i.e. the distribution of first returns to starting point. In particular it should visit the starting point $x = 0$ a number of times proportional to $\sqrt{t}$.

We first consider a simple heuristic argument that basically repeat the thinking associated to the dimensional formula in previous chapter (with $D_{num} = 1/2$, $D = 1$ and $D_{tot} = 1$). We here again assume that the first return distribution is a power law: $P_{first}(t) \propto t^{-\tau}$. We then calculate τ from expressing the total time t as the number of intervals multiplied by the average interval length:

$$t \;=\; \sqrt{t} \int_0^t t' \, P_{first}(t') dt' \tag{3.8}$$

This gives $t = t^{1/2+2-\tau}$, or $P_{first}(t) = t^{-3/2}$. We will prove this below.

Formal proof: Consider the first passage time for a random walker, defined as the time when the first visit to position x occurred, starting at position 0 at $t = 0$. Characterizing the walk with the diffusion constant D, the distribution $P_x(t)$ for the first passage time t to position x is:

$$First_x(t) = \frac{x}{\sqrt{2\pi D} \cdot t^{3/2}} \cdot exp(-\frac{x^2}{4Dt}) \; . \tag{3.9}$$

We here prove this equation based on a derivation that was presented in [19].

Considering the cumulative probability

$$\mathcal{P}_x(t) = \int_0^t First_x(t') dt' \tag{3.10}$$

giving the probability that the random walk reached x at least once before the time t. This probability is also equal to the probability that the walk passed x at some time prior to t. Thus, it is equal to the probability that the *max* excursion of the walk in the time interval $[0; t]$ is larger than x.

The distribution of the "*max*" of the walk up to time t is related to the distribution of the end-point at time t:

---

where we at position x add and subtract contributions according to the exchange of particles with neighbor positions. Thus

$$\frac{\partial \rho}{\partial t} \;=\; D \, \frac{\partial^2 \rho}{\partial x^2} \tag{3.7}$$

with diffusion constant $D = (1/2)\delta l^2/\delta t$ ($D = 1/2$ in the above example with a random walk of unit step and unit time)

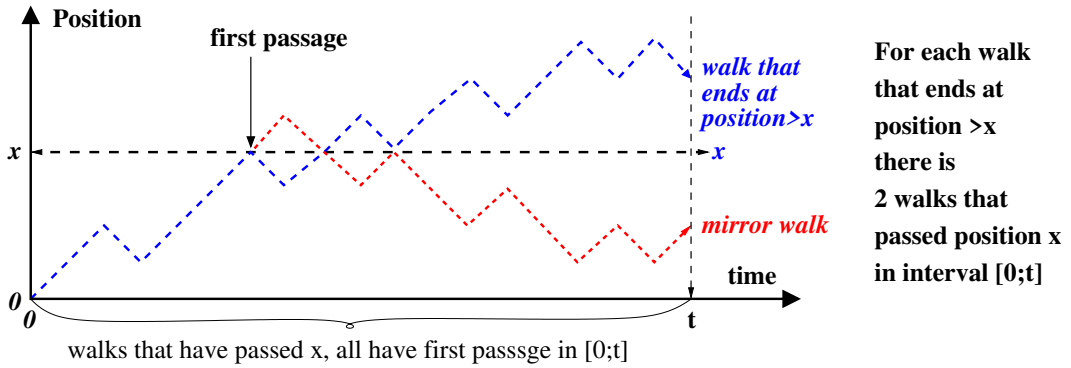


Figure 3.3: **First passage time problem.** Example of a random walker starting at position $x = x_0 = 0$ at time $t = 0$ and ending at some position larger than x. The figure illustrates that for each walk that ends at a position larger than x at time t, there are exactly two walks that have passed x at earlier times.

.

- *For each walk that ends in a point $y > x$, there are exactly two walks that go beyond x at some time before t.*

These two paths are respectively the original path and its *mirror* path, defined as the path that follow the original path to first passage of x, but thereafter is reflected in x. This is illustrated in Fig. 3.3. Also notice that because all paths that passes x have a mirror path, then there is a one -to -one mapping between all paths that end at a position greater than x at time t and all pairs of paths that pass x somewhere before t. There is no paths that passes x that is not included here.

The first of the two path reaches a position greater than x, the second mirror path does not. Therefore

$$\mathcal{P}_x(t) = P(max\ excursion\ in\ [0;t] > x) = 2 \cdot P(end\ position\ is > x)$$

By this argument we have connected the Gaussian distribution at the end of time interval t to the distribution of all passage times before t.

The distribution of positions at time t is given by the normal distribution of the random walk after time t:

$$P(end\ position\ is\ > x) = \frac{1}{\sqrt{4\pi D t}} \int_x^\infty dy\ e^{-y^2/4Dt}\ .$$

Thus the probability that the random walker exceeds x at some time before t is:

$$\mathcal{P}_x(t) = 2\frac{1}{\sqrt{4\pi D t}} \int_x^\infty dy\ .\ e^{-y^2/4Dt} \tag{3.11}$$

The actual probability that it exceeds x for the first time between t and $t + dt$

is given by the differential of this cumulative probability, namely

$$\begin{aligned} First_x(t) &= \frac{d}{dt}\mathcal{P}_x(t) \\ &= 2\frac{1}{\sqrt{4\pi D}}\frac{d}{dt}\left(\frac{1}{\sqrt{t}}\int_x^\infty dy\, e^{-y^2/4Dt}\right) , \end{aligned} \tag{3.12}$$

which is easiest differentiated by bringing $1/\sqrt{t}$ under the integral and substituting. Thus we set a new variable $v = 2Dtx^2/y^2$:

$$dv = -\frac{4Dtx^2}{y^3}\cdot dy \Rightarrow \frac{dy}{\sqrt{t}} = -\frac{y^3\cdot dv}{4Dx^2t^{3/2}} = -\frac{1}{\sqrt{2Dx^2}}\cdot\frac{dv}{v^{3/2}} \propto -\frac{dv}{v^{3/2}} .$$

The minus sign means that the substitution ($y \to v$) leads to a shift between upper and lower boundary. Further, the boundary $y = \infty$ is changed to $v = 0$ whereas the x-boundary is changed from $y = x$ to $v = 2Dt$.

$$First_x(t) \propto 2\frac{d}{dt}\left(\int_0^{2Dt}\frac{dv}{v^{3/2}}\cdot e^{-x^2/2v}\right) \propto \frac{1}{t^{3/2}}\cdot exp(-x^2/4Dt) ,$$

where we differentiated the integral with respect to its upper boundary. Overall, for $x \ll \sqrt{4Dt}$, this provides us with the famous first return scaling that is valid for times large compared to the typical time required to reach x:

$$First_{x\sim 0}(t) \propto \frac{1}{t^{3/2}} , \tag{3.13}$$

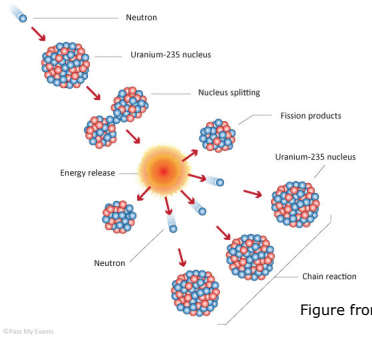
which is also equal to the distribution of times where the random walk first returns to $x = 0$. This is then called the *distribution of first return*.

3.2 Critical branching process

Directed percolation had one active unit possible generating more than one active unit, and leading to a cascade of active unit. An illustrative example of such a cascade dynamics is a chain reaction, like the one depicted in Fig. 3.4. This process is like directed percolation in infinite dimensions (because there is no limitation on where to place the particles produced. This infinite dimensionality is similar to the one we saw for the Bethe lattice (Sec. **??**). The cascade dynamics can be quantified in terms of the number of active states (fission nuclei in Fig. 3.4) as a function of time.

So-called *branching trees* correspond to the directed percolation in infinite spatial dimensions, that is, in so high a dimension that the different branches never overlap. Accordingly, the scaling properties of these trees amounts to the scaling properties of directed percolation in high spatial dimensions.

The simplest way to understand the dynamics of the critical branching process is to map it to a random walk in terms of the number of active states



Super-Critical chain reaction: each reaction give rice to > 1 new reaction

Critical chain reaction: each reaction give rice to one new reaction

Sub-Critical chain reaction: each reaction give rice to < 1 new reaction

Figure from passmyexams.co.uk/GCSE/physics

Figure 3.4: **Chain reaction as known from nuclear reactions.** Each fission leads to the emission of three neutrons, which in principle can lead to three new fission events. This process demands that there is a sufficient number of Uranium nuclei available to capture all three neutrons. If the amount of Uranium is small, then most neutrons escape without causing new reactions, and the process stops. On the other hand, when there is more than a critical amount of uranium, the process will amplify exponentially, leading to a run away effect (explosion/meltdown). The Uranium nucleus has a diameter of 12 fm$= 12 \cdot 10^{-15}$ m, A Uranium atom has a diameter of 3.2 $\mathring{A}$, giving a mean free path of about 10 cm.

as a function of the number of reactions (not time). Each time a reaction occurs, it gives rise to zero, one, two or three more reactions (in fig. 3.4). A critical condition is in place, when the average number of active states does not change with each reaction, that is, the probability that one branch dies is exactly balanced by the probability that it gives rise to more than one new reaction, *see* Fig. 3.5. This process is most simply discussed if each active state can cause zero or two new reactions. In that case each reaction has 50% chance to terminate the local chain, and 50% chance to grow to initiate two new chains. Notably, for the particular chain reaction shown, the mean free path is long compared to the atomic distance, making the cascading event nearly uncorrelated in space.

Let us define the size s of a branching process as the total number of activated sites which are involved at any time during the process. The probability $p(s)$ for having a size s of the tree must fulfil from partitioning the the tree into two sub-trees at the root. Thus we start with one node, $s = 1$, which can then split into a right branch with one node, or with equal probability terminate. In the first case we end with one node, and $p(1) = 1/2$. Summing over all partitions of a tree of size s then

$$p(s) = \sum_{k=1}^{s} p(k) \cdot p(s-k) \ , \tag{3.14}$$

corresponding to all possible sizes k of the left tree, and the additional require-

ment that the corresponding right tree should have the remaining size $s-k$. The top of Fig. 3.5 shows one such possible partition of a total tree. The recursion relation defined by Eq. **??** can be solved using generating functions, and we will do so later (in the end of the Network chapter). For now, we will instead use the mapping to a random walk of the number of live branches, counted in terms of subsequent branch point decisions (see Fig 3.5),

$$p(s) \propto 1/s^{3/2} \tag{3.15}$$

as it simply reflect the first return of a random walker.

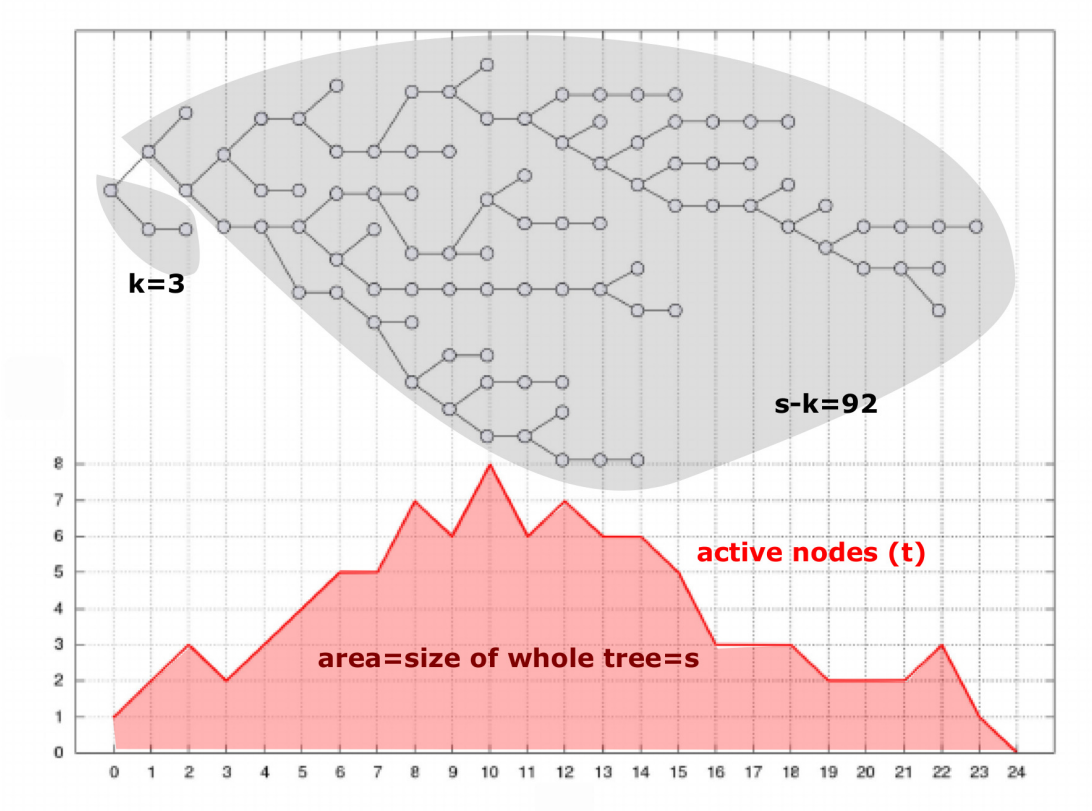


Figure 3.5: **Galton-Walton critical branching process and its relation with a random walk.** At each time update t (horizontal axis) there are several active nodes, and thus the changes tend to be larger when this number is large. Instead one may follow the process as a function of the number of active nodes a by updating it in steps of one node (red curve). Each such node may, with equal probability, either become inactive, or spawn two new active nodes. Note that, whenever one active node is considered, nodes not considered are just replotted at the same time point. As a function of this updating the tree grows or shrinks as a random walk. Figure from Francesc Font-Clos.

Minimal tutorial: Estimate critical thickness of a large uranium plate, i.e. the thickness where just 1/3 of the 3 released neutrons of a fission event are captured in the plate.

Noticeably, one may also consider the distribution of survival times for the critical branching trees in Fig. 3.5. This would be different than the size distribution $1/s^{3/2}$ because there would be several sites that branch at the same time t (see Fig. 3.5). This distribution would be the survival time distribution in directed percolation in high enough dimensions.

3.3 Self Organized Criticality: The Sandpile Paradigm

Mini tutorial:
Consider a critical nuclear chain reaction, where each time a neutron reached a uranium nucleus, it causes emission of 3 subsequent neutrons. How can one modulate this process to obtain critical conditions in a reactor? Why would such critical conditions be desirable?

Previous discussions about critical behavior in the Ising model (Sec. 1.4) and in percolation (Sec. **??**) raise the question why one should bother with properties at a critical point; This is only around the critical point and represent a very small part of the parameter space. However, there is reasons to believe that parts of nature tends to organize towards a critical point by themselves. Some open driven systems tend to be pushed towards larger and larger features, until they just marginally start to break down. The canonical model for this type of phenomenon, which is termed *self organized criticality (SOC)*, was suggested by Bak, Tang and Wiesenfeld in 1987, and is illustrated in Fig. 3.6.

The canonical version of SOC takes place on a two-dimensional square lattice consisting of $N = L \times L$ sites. Each site i can take certain integer values h_i, where $h_i = 0, 1, 2, 3, 4, 5, \ldots$. All sites with $h_i = 4, 5, \ldots$ are considered unstable and topple simultaneously. When they topple they distribute one unit to each of their four nearest neighbors. Thus, the sum $\sum_i h_i$ is conserved when we are away from the boundaries. However, any site that is at the boundary, distributes a unit out to imaginary neighbors outside of the system. These units are lost. When all sites i have $h_i < 4$ a new grain is added at a random site, and the above procedure is repeated.

The model is visualized as the activity in a huge square office, where bureaucrats do exactly nothing, unless they get 4 or more assignments. When they get so many assignments, they get frustrated and push the assignments to their neighbors. Neighbors to windows just throw their assignments out of the window. This version of the model is illustrated in Fig. 3.6.

Mini tutorial:
What would happen to the dynamics if one closes all windows in the bureaucracy model above (i.e. and papers rebounce to the sending bureaucrat at the edge?)

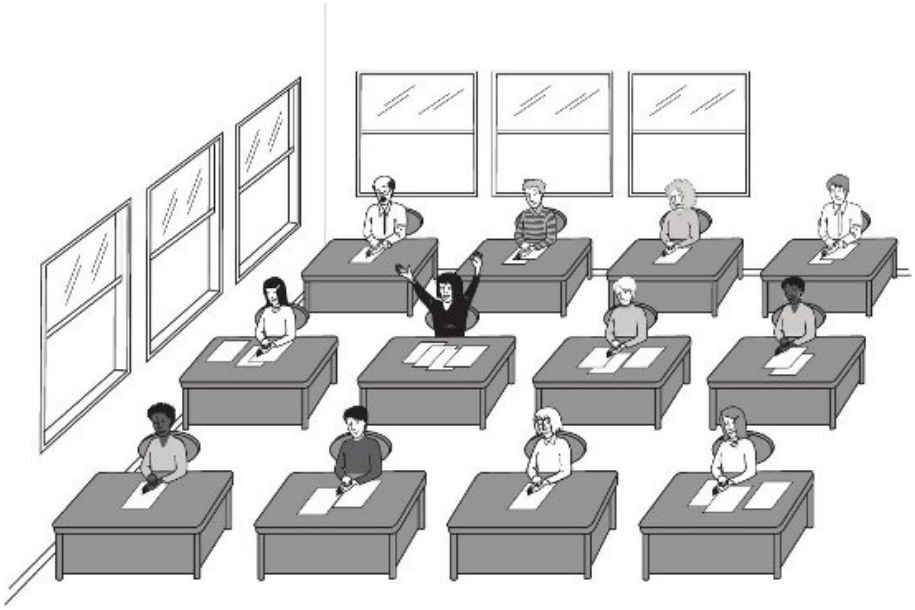


Figure 3.6: **Sandpile model cartoon.** The classical "sandpile model" of Bak, Tang and Wiesenfeld, here re-explained in an office version by Peter Grassberger.

The key observable is the sequence of topplings that take place when one new grain is added until the system is settled (see Fig. 3.7. This constitutes an avalanche, which is measured by the sum of all activity until all sites are below the threshold. The distribution of the sizes of these avalanches turns out to be power law distributed,

$$P(s) \propto \frac{1}{s^{\tau}} \quad , \quad with\ ; 1 < \tau < 3/2\ , \tag{3.16}$$

with the exact value of τ dependent on the dimension. In two dimensions $\tau \sim 1.2$ whereas $\tau = 3/2$ in infinite dimension or a random neighbor or mean field version. Such a power law reflects that the system operate on the critical point. And it does so without any fine tuning. It just needs time to reach it! That is, one should first study the avalanches after many grains have been added, and the system thereby has self-organized to be at the critical point (with examples shown in Fig. 3.8).

Mini tutorial:
Consider a random walker placed in the center of a line of length L. If it steps right or left with equal probability, then how many steps typically pass before it reaches one of the ends?

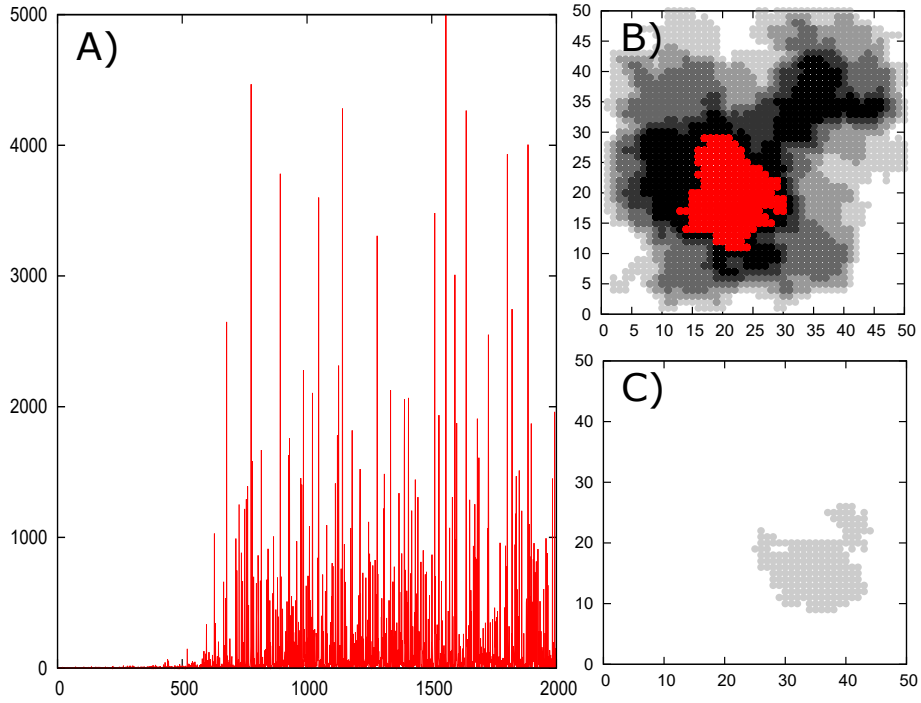


Figure 3.7: **Avalanche dynamics in the sandpile model.** Panel A shows the time series of avalanche sizes as one starts filling the sites in a 50×50 square lattice. After some time (a transient phase), a steady state distribution of avalanches is obtained. In panels B and C we show the spatial extent of two avalanches. The upper one is large, and some sites topple more than once (the red sites toppled more than five times during this large avalanche).

A scaling relation from following a grain of sand.

There exist scaling relations for SOC models. Following Kadanoff we consider the avalanche size distribution in the self-organized critical state:

$$P(s) = \frac{1}{s^\tau} \cdot f\left(\frac{s}{L^D}\right) \sim \frac{1}{s^\tau} \cdot exp\left(-\frac{s}{L^D}\right) \tag{3.17}$$

where the function $f \sim 1$ for $s << L^D$ and decreases very fast (let us for simplicity say exponentially) when the avalanche size exceeds the system size. D is the dimension of the avalanche, i.e. how the number of topplings in the avalanche scales with its horizontal extension. Notice that D can be larger than two because each site can topple multiple times, corresponding to avalanches that get "thicker" as they are horizontally larger.

In average then each time one grain is added, one grain have to leave the system. This statement is true when sampling over many avalanche in the stationary state conditions, and simply means that what comes into the system must also leave the system. This does not mean that one grain has to leave for each avalanche, as many avalanches do not reach the boundary and thus cannot contribute to a loss. However, sometimes avalanches reach the boundary and many grains will leave the system.

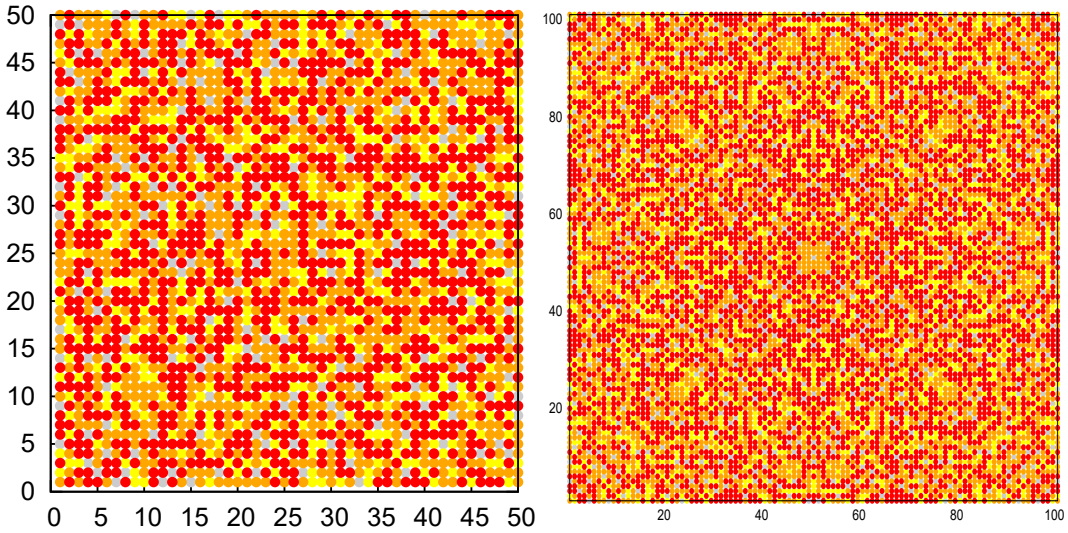


Figure 3.8: **Site configuration in an SOC model.** The model is driven by adding sand grains to random points (left panel), respectively always adding to point in the middle of the 101×101 square lattice used here (right panel). Sites colored in yellow mark $h = 1$, orange ones mark $h = 2$, red ones mark $h = 3$.

The steady state condition means that *on average* one avalanche has to provide sufficiently many topplings to bring one grain out of the system. If grains are added randomly, the average distance to the boundary is $\propto L$, and since grains topple in random directions, the added grain has to participate in $\sim L^2$ steps before reaching the boundary. Therefore, for $\tau < 2$, the steady state implies (inserting the cut-off function in the upper end of the integral)

$$L^2 = \int_0^{\infty} s \cdot P(s) ds \sim \int_0^{L^D} s^{1-\tau} ds = L^{D\cdot(2-\tau)} \tag{3.18}$$

and thus we get

$$2 = D \cdot (2 - \tau) \Rightarrow \tau = 2 - \frac{2}{D} \, . \tag{3.19}$$

This gives us a direct relationship between the cut-off of the avalanche and the distribution of these avalanches far away from the cut-off. For example if $D = 2$, corresponding to flat avalanches avalanches with simple convex boundaries, then $\tau = 1$. Extensive simulations of the sandpile model found that τ are close to 1.25.

Mini tutorial:
Given $\tau = 5/4$, what is D for avalanches in the sandpile model.

Notice that if we instead excited the system by adding grains only at the boundary, then on average it would only take L steps for a grain to leave the

system (instead of L^2). This is because one is much closer to the exit than if grains were added in the bulk. This can be proven by considering the first returns (to the boundary) of added grains:

$$average\ exit\ time\ from\ boundary = \int^{L^2} \frac{t \cdot dt}{t^{3/2}} \propto L \tag{3.20}$$

where the upper boundary is set by the time it takes to cross the system and exit on the other side. When this happens it is not the first return of the random walk, because it has loss on both sides. The excitation at the boundary would give

$$1 = D \cdot (2 - \tau) \Rightarrow \tau = 2 - \frac{1}{D}\ , \tag{3.21}$$

which is substantially steeper than when grains are added in the bulk (large avalanches are less likely).

Directed Sandpile

A solvable version of an SOC model was suggested by Depak Dhar (Physical Review Letters 63 (16), 1659), who assigned a critical threshold of two and distributed units from any position (x, y) in an $L \times L$ lattice to positions $(x-1/2, y+1)$ and position $(x+1/2, y+1)$ (periodic boundaries in x direction, and y effectively acts as a time coordinate). This model and a corresponding avalanche are illustrated in Fig. 3.9.

The critical state of this directed sandpile model is one in which half of the sites has value zero and the other half have value unity and these zeroes and ones are randomly distributed. To see this one first have to realize that each avalanche is compact: Any point inside the avalanche will recieve two grains and thus for certain topples at next step. Also by nature no site will topple more than once during an avalanche. Finally, consider for example the blue edge of the avalanche in the figure. It will expand if the site at the boundary had one grain. It will contract if the grain on the right have zero grain. Thus if the probability to be 1 is exactly 1/2, then the boundary will perform a random walk. This is exactly the criterion for a critical avalanche, with a size distribution that become power law distributed.

Each avalanche, that is, a set of contiguous sites in space and time, will consist of sites that at most topple once, and further the avalanche area (see the figure) will be "compact" in the sense that there are no islands within the avalanche area that do not topple. In fact inside the avalanche each site receives two grains and these are thus certain to topple. As a consequence the size distribution of avalanches is given by the random walk movement of its boundaries (*compare* Fig. 3.9): When these two boundaries merge, the avalanche terminates. Thus the duration of avalanches is given by the point where the two random walks meet each other. As the difference between two

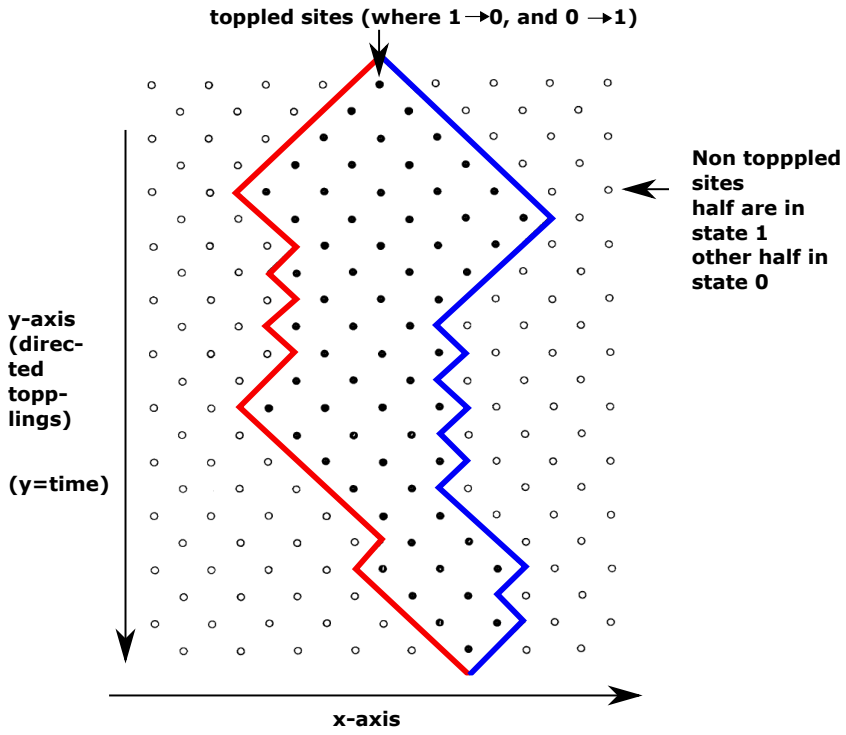


Figure 3.9: **Directed sandpile.** Each site can contain zero, one or more sand grains. If a site contains two or more grains it topples and delivers one grain to each of the two sites below it (downwards on the figure, see green arrows). The critical state contains zero or one with equal probability=1/2 across the entire x-y space. The figure shows an avalanche that involves all sites between the two outer boundaries (solid points). The avalanche has a duration of 22 (layers) and a size of 64. The boundaries of the avalanche perform a random walk, as highlighted by solid lines.

random walks is again a random walk, the avalanche duration will be power law distributed as the first return of a random walk (eq. 3.13):

$$P_{duration}(t) \propto \frac{1}{t^{3/2}} \ , \tag{3.22}$$

and the chance for an avalanche to propagate more that ℓ steps along the y axis would be $P_{duration}(t > \ell) = 1/\ell^{1/2}$. The size of the avalanche is its length (duration) times its width, which for a random walk gives the size $s = \ell \times \sqrt{\ell} = \ell^{3/2}$. Reversely, an avalanche of size s has length $\ell = s^{2/3}$. As a consequence the chance to be larger than size s:

$$P_{size}(> s) = P_{duration}(t > s^{2/3}) = \frac{1}{s^{1/3}} \ , \tag{3.23}$$

yielding the avalanche size distribution $p(s) = dP/ds = 1/s^{\tau}$ with exponent $\tau = 4/3$.

Random Neighbor Sandpile

It is sometimes worthwhile to consider the random neighbor version of a model, as this allows for precise mathematical treatment. This in fact corresponds to avalanches propagating on a Bethe lattice (with the caveat that one then needs to make some sparse holes to get rid of the excitations).

In this case we again explore the random walk feature of critical branching processes. Such a simple SOC model was suggested by (H Flyvbjerg in Physical review letters 76 (6), 940). In this model one considers N sites that each can contain zero, one, or more grains of sand (or papers in the bureaucrat formulation, Fig. 3.6). Any site with more than one grain topples and sends one grain of sand to one random other site, and another grain of sand to another random site. The exception is that any of these grains is lost with probability $1/N$ (corresponding to one open window in the office formulation). The critical state of this model has half of the sites occupied by one grain and the other half of the sites is empty. Each avalanche is started by one addition, that subsequently triggers an avalanche of topplings with an activity that follows a random walk, and terminate as the first return of a random walk, see eq. 3.13. Thus the avalanche size distribution is

$$P(s) \sim \frac{1}{s^{3/2}} \cdot \ exp(-s/N) \ . \tag{3.24}$$

3.4 Evolution as Self Organized Criticality

Evolution of species during the last 540 million years shows signs of large scale cooperative behavior: often during the history of life there have been major "revolutions," where many species have been replaced "nearly" simultaneously. This dynamics is illustrated in Figs 3.11. Spectacular examples include the Cambrian explosion 540 million years ago where a huge variety of life arose within a short time interval, and the Cretaceous-Tertiary boundary, where mammals replaced the dinosaurs as large animals.

In between the major "transitions" there were periods of quiescence, where species seemed to live in "the best of all worlds", with small risk of extinction. However, the pattern of life is more subtle than quiescence versus worldwide on-off transitions. The historical record often exhibits smaller size extinctions in the more quiet periods. Fig. 3.12 shows that ecological events occur on many scales, even cataclysmic events are found, which involve most of the contemporary genera. Furthermore, one sees that larger events are gradually less frequent than smaller events. There is no "bump" or specially enhanced frequency for the largest scale extinctions. In fact the distribution of extinction size s is consistent with a scale free distribution as indicated by the fitted $1/s^{1.5}$ curve. This overall gradual decline of event size distributions indicates:

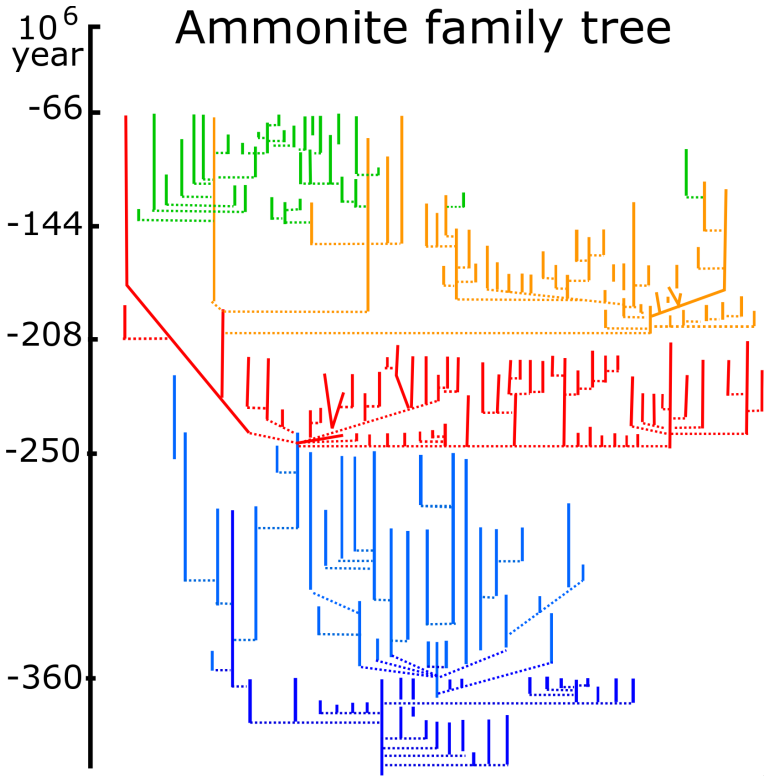


Figure 3.10: **Life & history in term of the Ammonite family tree.** Reproduced from Ref. [20]. Ammonites lived in water, and left a highly diverse fossil record with $\sim$ 7000 species, from 400 to 66 Million years ago. Notice the intermittent dynamics with calm periods interrupted by coherent extinction/speciation events.

- That large and small events may be associated to similar type of underlying dynamics. If extinctions were always externally driven by events like for example asteroid impacts [23] one would expect a peak at the large events.

- The non-Gaussian probability distribution for extinction events shows that the species in the ecosystem do not suffer extinction independently of each other. This is consistent with co-evolution on the scale of the global ecosystem.

To model the observed macro evolutionary pattern we start with units or agents on the size of the main players on this scale. These "agents" model the species of the ecosystem. Or course a species consists of many individual organisms, and dynamics of a species represents the coarse grained view of the dynamics of these entire populations. Thus, whereas population dynamics may be influenced by some sort of fitness, we here assume that species dynamics is governed by their stability against extinctions on evolutionary timescales:

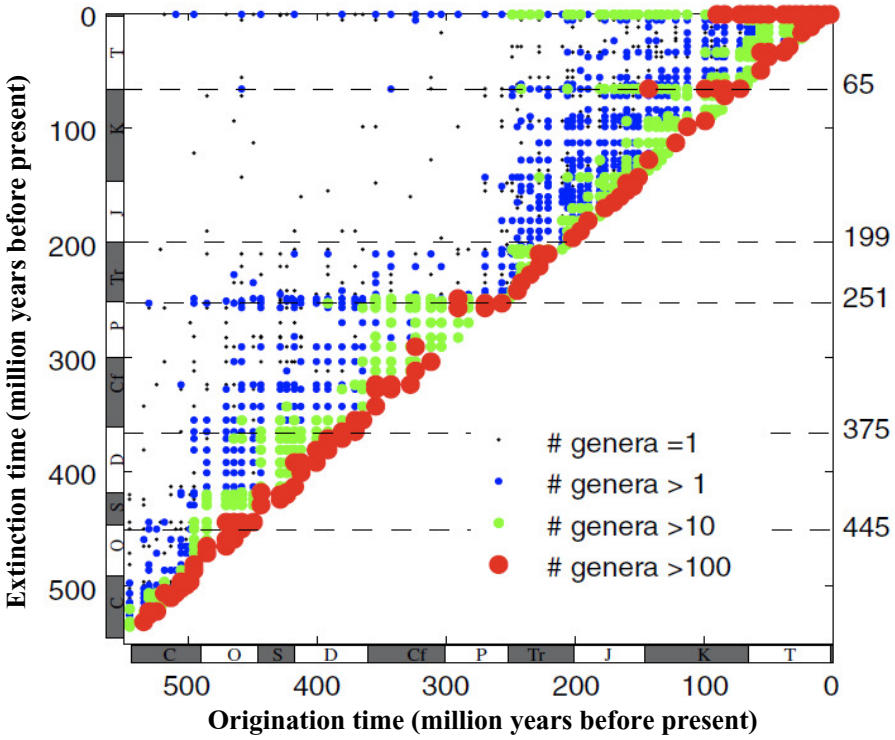


Figure 3.11: **Origination and extinction.** The graph shows times of existence times of 35,000 genera in the Phanerozoic [21] as visualized by [22]. Every event is quantified by number of genera, each defining a group of closely related species. The vertical distance from a point to the diagonal measures the residence time of a species. Notice the many points located close to the diagonal, reflecting the fact that most genera exist less than the overall genera average of about 30 million years. Notice also the division of life before and after the Permian extinction 250 million years before present.

Population dynamics		*Barrier dynamics*
	large time scale $\longrightarrow$	
Survival of fittest		*Evolve the least stable*

Given that the basic evolutionary unit is here defined as a species, we characterize each species using one number B_i. This number characterizes the stability of the species on a time scale much longer than the time needed to amplify to fill its biological potential (to reach the natural population level for that particular species takes short time, while the invention of a new species is rare). An ecosystem of species consists of N numbers B_i that each represent one species. Each of these species is connected to a number of other species, with links that could denote interactions, such as predation, collaboration, or niche maintenance [24].

Mini Tutorial: Draw ten real numbers from a uniform distribution between zero

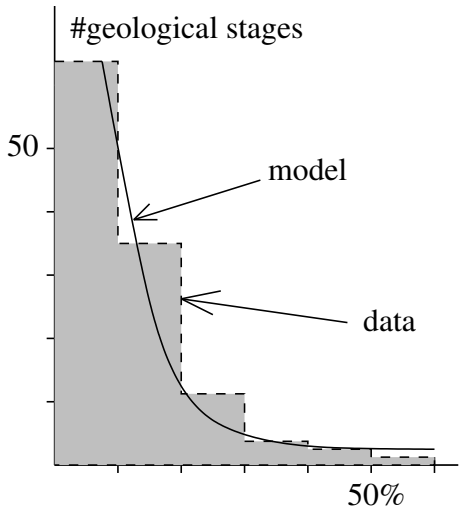


Figure 3.12: **Comparing the Bak-Sneppen model to data.** Histogram of family extinctions in the fossil record as recorded by Raup & Sepkoski [21]. The prediction, $p(s) \propto 1/s^{1.5}$ of the random neighbor version of the Bak-Sneppen model is marked as "model".

and one, Eliminate the smallest and replace it with another number drawn from same distribution. What functional form would the final distribution converge towards? Notice that it in fact does not matter what distribution we draw the numbers from

Bak-Sneppen model: For simplicity, let us first assume that the numbers B_i, $i = 1, 2, ...N$ are placed on a line, mimicking a one-dimensional model ecosystem. At each time step one changes the least stable of these species. As the stability is defined within the context of a given species, the fitness of a given species is a function of the species it interacts with, and accordingly the neighbor species will also change their stabilities (B values). The co-evolutionary updating rule for the agent-based model then reads [25, 26]:

- **At each step, the smallest of the $\{B_i\}_{i=1,N}$ is identified. For this as well as its nearest neighbors one replaces their B_i's by new random numbers in $[0, 1]$.**

The model just described is traditionally referred to as the *Bak-Sneppen (BS) model*. As the system evolves, the smallest of the B_i's in the overall systems are eliminated. After a transient period a statistically stationary distribution of the numbers B_i's is obtained. For the infinite system size limit ($N \to \infty$) this distribution is a step function where the selected minimal B_{min} is always below B_c. As a consequence the distribution of B is constant above B_c For the dimension $d = 1$ discussed, where the two nearest neighbors are updated, one obtains a self-organized threshold $B_c = 0.6670$, *see* Fig. 3.13. The right panel illustrates how the minimal B_i sites changes in the "species space" as

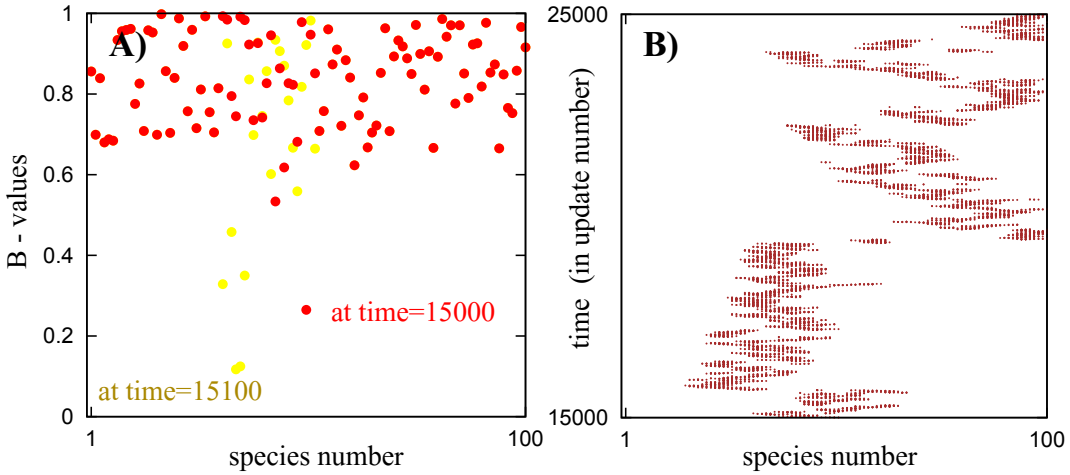


Figure 3.13: **Dynamics of the Bak-Sneppen model.** A) Example of distribution of barriers B_i in space. One observes that $B > B_c = 0.6670$ are distributed randomly in space. In contrast, sites with $B_i < B_c$ are highly correlated and tend to remain within a small region. B) Space-time plot of the activity in steady state where "time" is counted by the number. of updates. At each timestep the site with minimum B_i is highlighted by a "+"-symbol.

the system evolves. One observes highly correlated activity where sites that topples are close to sites that topples in previous timestep.

The right panel of Fig. 3.13 shows a "space-time" map of the minimal B_i sites in the time interval considered. Whenever the lowest barrier is found among the three sites that was updated at and around the previous minimum, then the active site performs a random walk. The figure shows that this type of small steps is what happens most frequently. When the site of lowest barrier value moves by more than one lattice spacing, it most frequently backtracks in subsequent updates. Importantly, activity tends to stay localized and form a sequence of changes in the same region of the model ecosystem. Thereby evolution is reinforced locally, bridging punctuated equilibrium in single species evolution [27, 28] to larger evolution and origination of new taxonomic groups[29, 30].

Punctuated equilibrium is a concept from paleontology coined by Gould and Eldredge, stating that most changes takes place on so fast timescales that one often does not find intermediates between to species where one was evolving from the other. On larger scales, Quantum evolution coined by Simpson, referred to the larger scale punctuations that is observed when one paleontological period terminates, and is replaced by another one with substantial differences in species compositions.

The obtained correlation of evolutionary activity relies on a self-organization that demands a lot of time. The self-organization allows the evolving system to develop towards a dynamical "attractor" that moves among states of the system where the numbers on the lattice are correlated across long distances.

I.e. the distribution of "species" with $B < B_c$ sits on a fractal in both space and time. This "attractor" state is therefore critical, and the algorithm is one of a class of models that let a system self-organize towards such criticality.

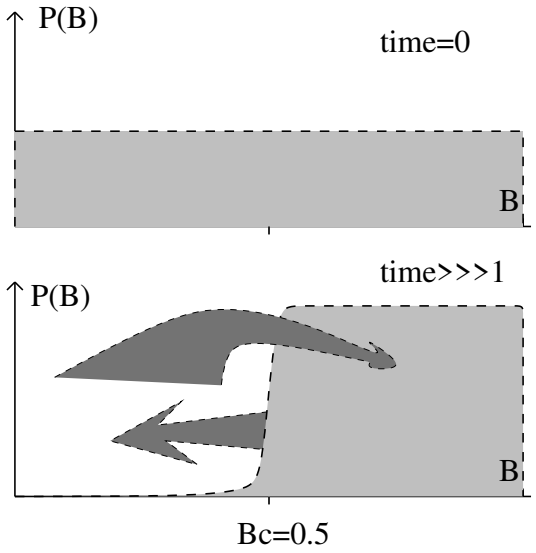


Figure 3.14: **Distribution of barriers/fitnesses in the BS model for the random neighbor version.** At each update one takes the minimum B_i and one random B_i value, and replaces each of them with a new random number within the interval $[0:1]$. The resulting distribution of B_i's reaches a steady state with a threshold value of 0.5. That is, the selected minimum is always below $B_c = 1/2$ for an infinite system, which in turn makes all numbers larger than $1/2$ treated equally. This means that these B_i can only be updated as passive neighbors.

Random neighbor model and its solution: To understand how the threshold in B emerges we consider a simpler random neighbor version of the BS model [31, 32]. Here, at each update one *changes the site with minimal B_i as well as one other randomly selected B_i*, see Fig. 3.14. Because of the absence of spatial correlations between the B_i's this simpler model can be solved analytically [32].

The model with random neighbors proceed like the spatially structured model: Starting a simulation with an initial distribution of $B_i \in [0, 1]$ that is random and uniform, then first the smallest of these B_i gets eliminated. This leads to a systematic depletion of small B_i values. After a transient period one obtains a statistically stationary distribution of B's. For $N \to \infty$ this distribution is a step function where the selected minimal B_{min} is always below or at B_c. This in turn implies that species with $B > B_c$ cannot be selected as the minimum, and therefore are only changed because they was selected as the random neighbor, irrespective of their actual B value. Therefore the distribution of B is constant above B_c.

At each step the dynamics then one B is always selected from below B_c

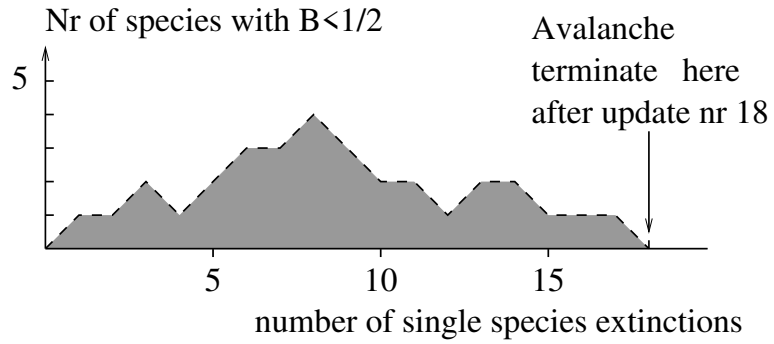


Figure 3.15: **Evolution of the number of sites with B<1/2 in the random neighbor version of the Bak-Sneppen model.** At any timestep there is equal probability to increase or decrease the number of active species a, thus defining a random walk of this number. The first return of the random walk to zero defines an avalanche that terminates when all $B_i > 1/2$. When this occurs the system is so stable, that the subsequent change will occur rarely, but anywhere along the 1-d ecosystem.

and the other B above B_c (in the infinite system size limit where a vanishing fraction of sites are below B_c. As the 2 newly assigned B's are assigned uniform random values in $[0, 1]$, the condition for a statistically stationary distribution of number of species in the interval $[0, B_c]$ is:

$$-1 \; + \; 2\, B_c \; = \; 0 \Rightarrow \; B_c = \frac{1}{2} \tag{3.25}$$

Notice that one in principle could select a B_{min} slightly above B_c, i.e. $B = B_c + 1/N$. But then the chance to again select a subsequent one above this number would smaller that $1/2$. And the chance to select B_{min} substantially above B_c decays exponentially with both system size and distance from B_c.

The time series of the minimal B exhibits correlations. An avalanche is defined as the number of steps s between two subsequent selections of minimal $B > B_t$. The number n of B's below $B_t = B_c$ exhibits a random walk and the size of the avalanche is determined number of updates s before this random walk return to zero: $P(s) \propto s^{-3/2}$. This is the famous distribution of waiting times in the *Gamblers Ruin Problem.*

Time scale separation, Extremal dynamics & Avalanches: The model is defined in terms of updating the site with the global minimum value of B. This selection implicitly assumes a separation of time scales in the dynamics, which in fact also allows us to naturally separate avalanches when all B_i are above the self-organized threshold.

The extremal dynamics (the selection of the global minimum as next active site) can be seen as the $\mu \to 0$ limit of the following local model [33]:

- **At each time step of size dt:**
  Select each of the $\{B_i\}_{i=1,N}$ with probability $\propto e^{-B_i/\mu}\, dt$. This

selection defines a list of active sites. For members in this list, replace them as well as their nearest neighbors by new random numbers in $[0, 1]$**.**

Here μ represents an attempt rate for microscopic evolutionary changes, and is proportional to the mutation rate per year (or perhaps per thousand years).

Noticeably, one may argue that the original choice of selecting a global minimum B_i to some extent violates the more common assumption of agent-based models that the activity of an agent is set by its local properties. However the above formulation indeed illustrates that there exists such a local agent-based formulation of the BS-model. We will see below that the behavior is equivalent on scales that are shorter than a length scale set by μ.

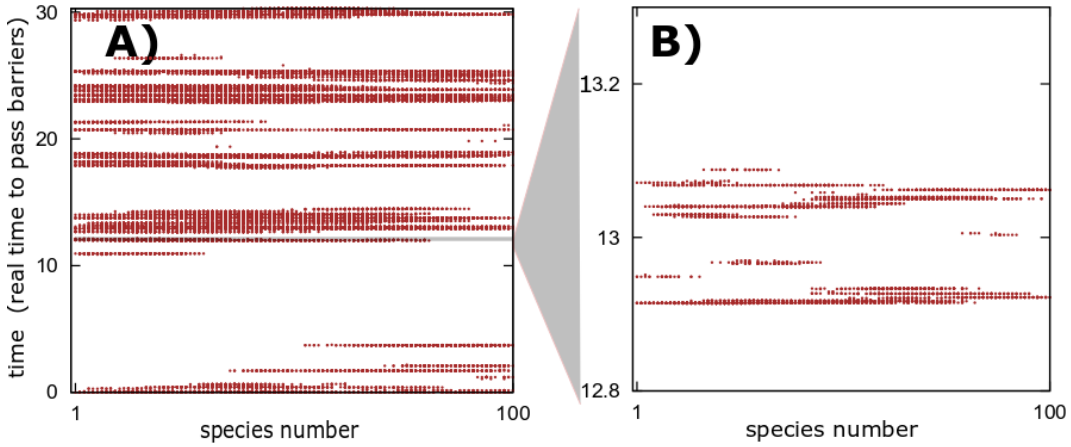


Figure 3.16: **Space-time plot of the activity in the Bak-Sneppen evolution model.** Each update is shown as a black mark. With the coarse time resolution of the plot, the avalanches appear as almost horizontal lines. The magnification on the right shows that there are avalanches within avalanches. The calculation was done at a mutation rate $\mu = 0.005$.

.

For each barrier B_t below the self-organized critical threshold B_c, a "B_t avalanche" starts when a first selected B_{min} is below B_t and terminates when a selected B_{min} is above B_t. In the *local* formulation, all activity within a B_t avalanche occur practically instantly when seen on a time scale of order $\exp(B_t/\mu)$. This statement may be reiterated for the larger avalanches associated to a $B_{t2} > B_t$, thereby defining a hierarchy of avalanches within avalanches. One may view the avalanche-within-avalanche picture as burst-like activity on different time scales (*see* Fig. 3.16). Time scales that may be set by associating each step of the algorithm to a time interval

$$\Delta t \;\propto\; \frac{1}{\sum_i e^{-B_i/\mu}} \quad\sim\quad e^{B_{min}/\mu} \qquad for \quad B_c - B_{min} \gg \mu \; . \tag{3.26}$$

Here, the final approximation uses that the distribution of barriers below B_c is scarce.

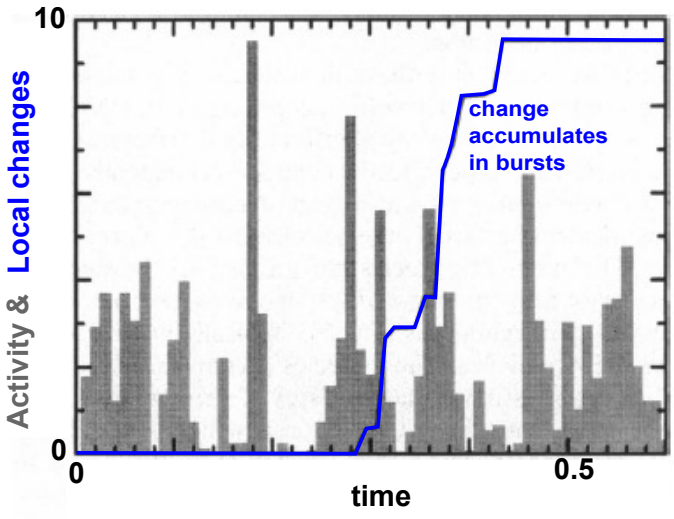


Figure 3.17: **Global activity in the Bak-Sneppen model.** Global activity (grey) plotted together with accumulated change at one position in the model ecosystem, using the finite mutation rate version [26, 33]. The figure illustrates that the individual species change most rapidly, when also the global ecosystem exhibits a lot of activity. Figure reproduced from [26].

For $\mu \to 0$ the evolutionary avalanches may mimic the boundaries between geological periods. Periods where all $B_i > B_c$ and the overall ecosystem is in a quasi-static period until an avalanche is initiated by a spontaneous mutation of one of the species at the threshold of stability. The duration of the stasis-period is then set by B_c, $t \sim e^{B_c/\mu}$, whereas the disturbance at the boundary will cascade as a critical branching process influencing a total number of species s with probability

$$P_0(s) \propto s^{-\tau} \tag{3.27}$$

Here the exponent $\tau = 3/2$ for ecological networks with dimension $d \geq 4$ [34] which compare well with the histogram of extinction events shown in Fig. 3.12.

On the relation to Directed Percolation: Consider the BS model in one dimension. In particular, consider an avalanche initiated in the BS model, and denote all sites with barrier values $B < B_t$ as active. A primary difference between the BS model and Directed Percolation (DP) is that the BS model only updates one site at each timestep, independent of how many active sites there are (sites below B_c). As the avalanche expands, this obviously leads to some delay in its evolution. If we **assume** that the expansion follows that of active sites along the DP network, then after s updates the avalanche has visited r sites where $s = r^D$ with $D = 2.33$ is the dimension of the DP network measured perpendicular to its overall direction (i. e. $= D_\perp + 1/\chi$ from DP). However, numerically the zero'th moment of the activity function for the BS model $r \propto s^{1/D} = s^{0.413}$ differs slightly, but significantly, from that of directed

percolation ($1/D(dp) = 0.43$). Numerical results by P. Grassberger on the BS model suggest that its avalanche exponent $\tau = 1.07 \pm 0.02$ is close, but not equal to, the exponent for DP clusters at the critical point ($\tau(dp) = \frac{\nu_\perp + \nu_\parallel}{\nu_\perp + \nu_\parallel - \beta} = 1.108$). In fact the BS model has only two independent exponents whereas directed percolation has three.

Process:	1-d BS	2-d BS	1-d LI	2-d LI	1-d NLI
D	2.43(1)	2.92(2)	2.23(3)	2.725(20)	1.63(1)
τ	1.07(1)	1.245(10)	1.13(2)	1.29(2)	1.26(1)
γ	2.70(1)	1.245(10)			2.05(5)
π	3.23(2)		2.93(3)	2.89(3)	2.25(5)

Table 3: Review of exponents of various extremal dynamics models, taken from *Paczuski et al.* PRE **53** 414. LI refer to the linear interface model, presumably the same universality class as "Zaitsev model" (which is the BS model with a conservation of the sum of reassigned B_i at each update). NLI refer to the non-linear interface model ("Sneppen model"), which is equivalent to the KPZ equation with quenched noise, driven infinitely slowly. Numbers in brackets indicate uncertainty on last digit. The exponent π describes the distance distribution between subsequent minimal sites, $p(\Delta x) \propto 1/|\Delta x|^\pi$.

Apart from integrating small and large extinction events into one combined framework, the model predicts

I) Each evolutionary avalanche consists of sub-avalanches on smaller scales. Thus when we analyse the fossil data on more fine grained time (and space) levels, we should expect to find each extinction event subdivided into smaller extinction events. Such correlations between extinctions may be examined by more fine grained data [35].
II) The temporal separation between evolutionary events of a given lineage will be power law distributed, with long periods of stasis, that are sometimes broken by a sequence of multiple small jumps.
III) Co-evolution allows for large evolutionary meanderings. Evolutionary barriers that seem impossible to pass at a stasis period gets circumvented by changes in fitness landscapes due to co-evolution adaptations (Fig 3.17).

Questions:
3.1) The sum of N random numbers selected uniformly between 0 and 1 will provide a good fit to a Gaussian (for N sufficiently large). Given that we want a Gaussian with spread 1, what should we select as N. Qlesson: Add 12 such numbers and subtract 6, then you get a Gaussian with mean zero and standard deviation unity. This is a handy and simple way to make it.
3.2) Simulate the standard SOC model for a $N = 50 \times 50$ system, and plot the avalanche size distribution.

Qlesson: See that it actually gives a power law. Notice that it takes time before large avalanches appear, i.e., there is a long transient.

3.3) After reaching the steady state, then restrict additions to one corner of the $N = 50 \times 50$ system and plot the avalanche size distribution.

Qlesson: The avalanche distribution gets steeper when only adding in a corner. Try to explain why.

3.4) Always add grains to position $(x, y) = (25, 25)$ and plot heights on the lattice after a long time.

Qlesson: Organize a fractal pattern, Try possibly larger lattice to get more extended fractal. You can also play with boundary.

3.5) Simulate the Depak Dhar sandpile model on a $N = 100 \times 100$ system. Confirm the scaling exponents mentioned in the text.

Qlesson: Observe that the avalanches are compact. Explain that.

3.6 Simulate a one-dimensional sand pile, with critical height two and a random redistribution rule (Manna model). That is, at each toppling one distributes two grains, but they are randomly put to left or to right neighbor (and sometimes out of system when in site in end or beginning of system topples).

Qlesson: There can be critical behavior in a dynamic model in one dimension. This is not possible in equilibrium models. The 1d ising model would not have that.

3.7) Simulate the evolution model for 100 species placed along a line in a variant of the model where only one of the neighbors is updated at each step. Plot the selected B_{min} as function of time, as well as the max of all previous selected B_{min}'s. How does the minima of B change as time progresses toward steady state (look at envelope defined as max over all B_{min} at earlier times)?

Qlesson: Self organization towards critical attractor is followed by following the maximum of all previous minima.

3.8) Repeat the assessment of the above model, but now simulated for a finite mutation rate $\mu = 0.05$. At each step allow all sites to change with probability $p_i \propto exp(-B_i/\mu)$, and then also update one of the neighbours of each site. Plot the space-time evolution of system. Redo simulation for $\mu = 0.03$. (*Hint*: one may speed up the simulation by using an even driven simulation (Gillespie algorithm (in later chapter)), where one updates one site at a time, selecting the next change as the one with the smallest value of $t_i \propto -ln(ran) \cdot exp(B_i/\mu)$ where ran is a random number between 0 and 1). Qlesson: Extremal-dynamics and self-organized criticality is obtained in the limit of infinitely small mutation rate, corresponding to an extreme separation of timescales.

Lessons:

- Random walkers, including first return of random walkers are a recurrent theme in critical processes, from fine tuned cascade processes to self organized criticality.

- Repetitive dynamics with separated time scales could result in self-organized

criticality, a dynamics where a system converge to a on going dynamics with infinitely long range correlations.

- Analytical expressions can often be obtained in the limit if infinite dimensions, and typical ending with exponents related to the first return of a random walker in one dimension.

Supplementary reading:

Christensen, Kim, and Nicholas R. Moloney. Complexity and criticality. Vol. 1. World Scientific Publishing Company, 2005.
Bak, Per. How nature works: the science of self-organized criticality. Springer Science & Business Media, 2013.

Chapter 4

Networks

Journalist talks with Henry Kissinger:
- Tell me, Mr. Kissinger, you are considered the inventor of the "shuttle diplomacy". Explain what it is, as an example.
– Oh, it's very simple, - says Kissinger, - You want to use shuttle diplomacy to marry Rockefeller's daughter to a simple guy from a Siberian village.
- It's impossible! How would you do that?
- Very simple. I'm going to a Siberian village, find there is a simple peasant and ask: "Do you want to marry an American lady? "
He says: "Why? We've got great girls here! "
And I say: "Yes, but she is Rockefeller's daughter. "
He goes: "Oh! This changes everything."
Then I go to Switzerland to a bank board meeting. I ask them: "Do you want a Siberian peasant to be your bank President? "
And the bank people say: "No way! "
- But what if he is Rockefeller's son-in-law?
- Oh! This changes everything!
So I go to Rockefeller and ask: "Would you like your daughter to marry a Russian peasant? "
- Poof, – says Rockefeller – What are you?
So I go: "But what if he is a president of a Swiss bank? "
- Oh! This changes everything! Susie! Come here, Mr. Kissinger has found a good fiance for you. He's a president of a Swiss bank! "
Susie: "Fu-y! "
I say: "Perhaps, but he is a Siberian man. "
Susie: "Oh! This changes everything! "

4.1 Introduction

4.1.1 When Networks are useful

Networks are a widespread concept in both popular and scientific literature. Networks are used to characterize the organization of a system of *heterogeneous*

components (the *nodes*), each interacting with a small subset of the other components. These interactions are described by *links*. Networks are very much about *history*, as the real life networks evolved through a sequence of events that took place on a much longer time-scale than dynamical processes taking place **on** the network. The concept of networks may accordingly be useful for systems with

- **Heterogeneity:** Systems with distinctly different components
- **History:** A real network does not appear by random assignment of links, but self-organizes over a long time period. Networks are thus useful for **systems with a separation of time scales**, with a dynamics on the network that occurs repeatedly and on a fast timescale. Network rewiring is in contrast much slower.
- **Distribution & containment** of information, energy or material.

Fig. 4.1 illustrates these concepts using the large scale regulatory network of *S. cerevisiae*, a network that consists of many distinctly different proteins that developed with the organism over an evolutionary time scale. This time scale was billions of times larger than that of the dynamics *on* the network, namely the one associated to the two-hour generation of this organisms (the rate at which the organism replicates). Finally, the light colors in Fig. 4.1 show the response of the network to some external perturbation. The response is localized, illustrating that molecular networks not only facilitate information transfer, but also that they confine the information to relevant sub-parts of the system. This can be seen as a "signalling horizon" that is also often reflected in the topology of networks in other complex systems [36].

Mini Tutorial: Why is it often useful to limit information spreading in living systems?

4.1.2 Basic concepts

Figures 4.2-4.6 define basic quantities in network theory. Mathematically, a network is simply a set of nodes and a set of links (connections) between these nodes. Not all nodes need to be connected to other nodes, and a network can well have components that are not connected. These concepts are centered around network connectivity, paths between nodes, how central nodes are in a network, and the characterization of the local neighborhood of a node.

Consider first the simplest model for a random network with N nodes, the Erdös-Rényi (ER) network [38]. This can be constructed by connecting each pair (i, j) of nodes with probability p by a link. The expected number of links in the network is then $L = p \cdot N(N-1)/2$. The average degree (=connectivity), defined as number of neighbors per node, is

$$\langle k \rangle = \frac{2L}{N} = p\,(N-1)\,. \tag{4.1}$$

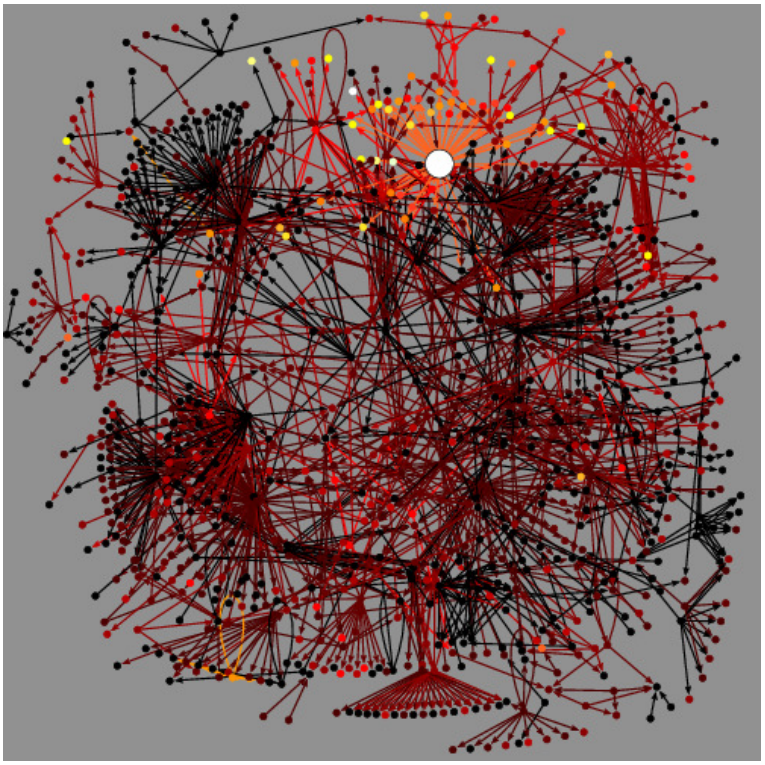


Figure 4.1: **Genetic regulatory network in Saccromyces cerevisiae.** Reproduced from [37]. Notice that the network consists of proteins that are all regulated through the cell nucleus, and does not reflect any "geographical" separation of the proteins. The highlighted nodes reflect genes that change expression in response to a particular external stimulus (amino acid starvation).

The factor 2 comes about because each link has two ends, contributing with 2 to the connectivities. Thus, $2L$ is the number of link ends in the network, distributed among N nodes.

The degree distribution (probability that a given node has k links to the remaining $N-1$ nodes) is the Binomial

$$P(k) \;=\; \frac{(N-1)!}{k!(N-1-k)!}\cdot p^k \cdot (1-p)^{(N-1-k)}\;, \tag{4.2}$$

because each node is connected to k specific nodes with probability p^k, and to none of the $N-1-k$ remaining nodes with probability $(1-p)^{N-1-k}$. The combinatorial pre-factor represent the number of ways to select the k specific nodes. For large N, $P(k)$ approaches the Poisson distribution (see Fig. 4.3)

$$P(k) \;=\; e^{-\langle k\rangle}\frac{\langle k\rangle^k}{k!}\;, \tag{4.3}$$

with an average degree

$$\langle k\rangle = p\cdot(N-1) \tag{4.4}$$

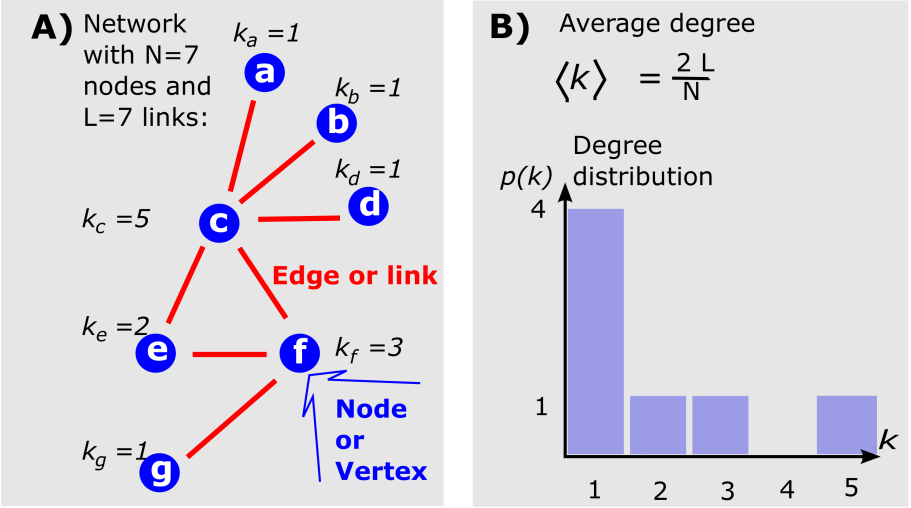


Figure 4.2: **Basic network definitions.** A) Network with nodes and links, and each node characterized by its degree, which is the number of links associated to the node. B) The degree distribution $n(k), k = 1, 2...$ for the network in panel A).

and variance $var = \langle k^2 \rangle - \langle k \rangle^2 = \langle k \rangle$, and thus $\langle k^2 \rangle = \langle k \rangle \cdot (\langle k \rangle + 1)$, an equation that will be useful in discussions of signal amplification.

Note that when the average degree $\langle k \rangle$ is high, thus when the spread $\sigma = \sqrt{var} = \sqrt{\langle k \rangle}$ is much smaller than the mean, then the Poisson distribution will approach a Gaussian distribution with mean and average given by the above equations. If the spread of degree between nodes is comparable to the mean degree, then the Gaussian approximation is poor, in part because it would predict a negative number of links (as it is symmetric w.r.t. to its peak).

A network is said to be *connected* if there exists a path between any pairs of nodes in the network, see Fig. 4.4A). The distance between two nodes in a network is defined as the minimal number of links that connects these nodes [1)]. For a connected network one defines its diameter as the maximum distance between any two nodes. The diameter thus sets an upper scale for distances in the system

Imagine a disease that spreads from a node in a random network where all nodes have equal connectivity k. One step away, $d = 1$, there are k new neighbors. A further step away, each of these newly visited nodes gives access to $k - 1$ new nodes, see Fig. 4.4B). If we ignore that we can ignore overlap where there is link between neighbors of a node, then the neighbors of the

[1]To calculate the distance from a node i to all other nodes in a connected network, one first makes a list of all the neighbor nodes of i and assigns them a distance $d = 1$. Subsequently one adds new layers of nodes to the list from all neighbors of already included nodes, provided that these neighbors are not already in the list. The distance to newly added nodes is calculated from the distance of neighbor nodes that is already present in the list.

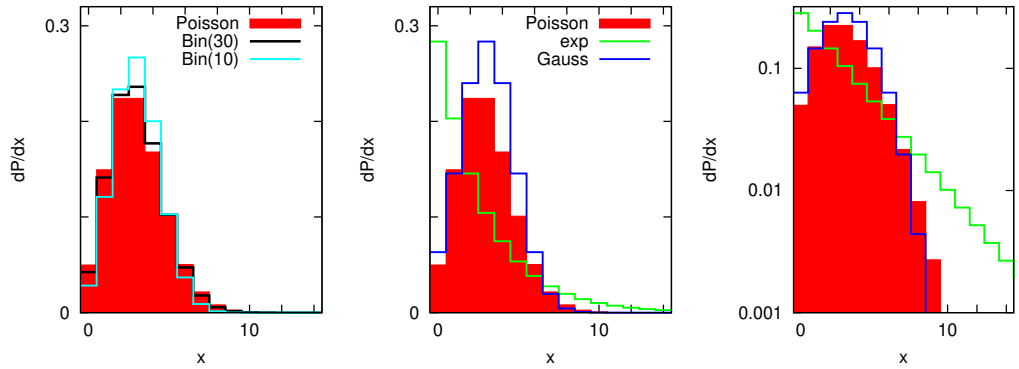


Figure 4.3: **Comparing distributions.** Left panel: Comparison between Poisson and Binomial distributions both with average 3, and Binomial sampled from 10, respectively 30 decisions. Right panels show comparison of Poisson with exponential and Gaussian distribution. The exponential has the same average as the Poisson distribution, and the Gaussian has the same average and the same standard deviation as the Poisson distribution. Note the logarithmic vertical axis. In all plots, the vertical axis we who probability $p(k) = dP(k)/dk$ of a node to have degree k, where $P(k)$ is the probability to for the node to have degree less or equal to k.

neighbors in total reaches $k \cdot (k-1)$ nodes. Assuming further that there is no double counting as we move further out in the network, then the number of visited nodes within distance d from the first node grows as

$$number(nodes\ within\ d) = k \cdot \sum_{d'=1}^{d} (k-1)^{(d'-1)} \sim k \cdot (k-1)^{d-1} . \qquad (4.5)$$

Therefore the number of visited nodes grows exponentially for any $k > 2$, see Fig. 4.4. For a more randomized graph, where the degree k may differ between the nodes, the disease will visit the entire network after a number of iterations d. This number d will be given by the slightly more complicated expression

$$\left(\frac{\langle (k-1)k \rangle}{\langle k \rangle} \right)^{d} \approx N , \qquad (4.6)$$

which we will derive shortly through the "amplification factor" expression below. Importantly, eq. 4.6 takes into account that each subsequent node is selected by a probability proportional to its degree. This is because a node is in end of a link, and when we follow a link we therefore have double the big probability to find a node with say 10 links than a node with 5 links.

We now use the above considerations to estimate the scale at which the signal that amplify and spread on each node will start interfering with each other. This scale is set by the scale at which the signal have reached a big fraction of the network. Assuming that there is no overlap between the different

signaling pathways before this upper scale is reached, we estimate the diameter to be about

$$Diam \sim d \sim \frac{log(N)}{log(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1)} \,. \tag{4.7}$$

The main lesson resulting from eq. 4.7 is that the diameter of a random network only grows very slowly with network size N.

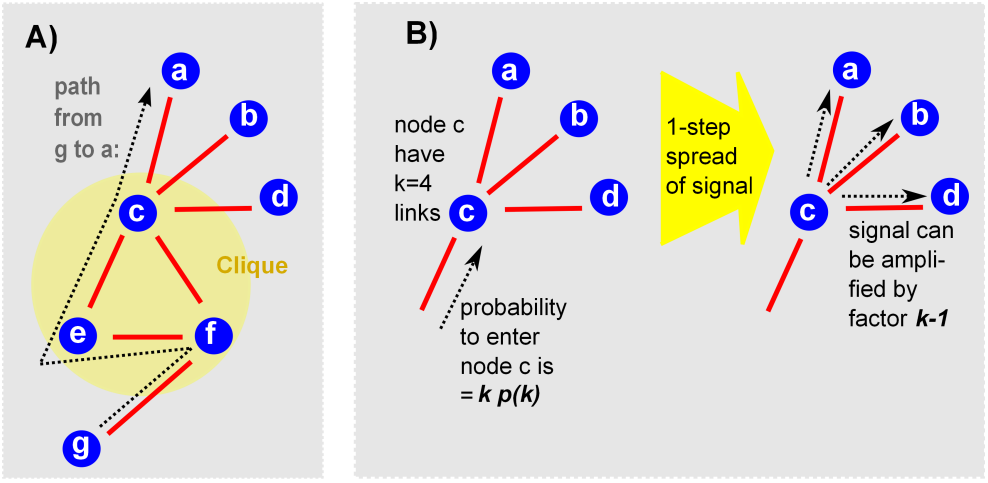


Figure 4.4: **Spreading process on a network.** A) Path from node g to node a in a network. Notice that the path shown is longer than the shortest path (g-f-c-a). B) Amplification of "virus like" signal as it enters and spreads across node c with connectivity $k = k_c = 4$. As such a signal subsequently spreads through many nodes, its activity on the network will be multiplied by the connectivity minus one for each node passed.

Mini Tutorial: Why is the average connectivity of your neighbors typically larger than your own?

4.1.3 Amplification factor

The above considerations on signal spreading can be rephrased in terms of the amplification factor $\mathcal{A}$. Consider a "virus" that enters a node from an incoming link, and assume that it is replicated and transmitted to all new neighbors, see Fig. 4.4B). Thereby, it is amplified by a factor $k - 1$. This means, that the amplification of the virus by passing a node of connectivity k is $k - 1$.

Now we want to average the above amplification as it is transmitted over different nodes in the system. Not all nodes have equal chance to amplify signals, because the probability to enter a node is in itself proportional to its connectivity k. Thereby, the amplification in a network [39] is obtained by the

weighted average

$$\mathcal{A} \;=\; \frac{\int k(k-1)n(k)dk}{\int kn(k)dk} \;=\; \frac{\langle k(k-1)\rangle}{\langle k\rangle} \;=\; \frac{\langle k^2\rangle}{\langle k\rangle} - 1\,. \tag{4.8}$$

Obviously this equation assumes that signals can spread both ways across a link. In case this is not the case, that is where signals only transmit one way along each link, the network is directed. Directed networks are not considered in these notes, but play an important role in for example biological regulation, or in hierarchical organizations.

Equation 4.8 implicitly assumes that there is no correlation between the connectivity of one node and the connectivity of a neighbor node. When $\mathcal{A} > 1$ then "disease like" signals tend to be exponentially amplified, and therefore will spread across the entire network. For $\mathcal{A} = 1$, on the other hand, perturbations will be marginal spreading, where some will spread and others will "die-out".

To have marginal spreading, one input signal on average should lead to one output through a new link. A network with Poisson distributed degrees, and $\langle k\rangle = 1$ will have $\langle k^2\rangle = 2$ and $\mathcal{A} = 1$. Such a network will consist of multiple clusters with the power law distribution of the cluster sizes shown in Fig. 4.12B). We will return to the amplification factor later, as this is interesting in setting the threshold, e.g. the fraction of vaccinations needed to stop an epidemic. Thus $\mathcal{A}$ plays a role in determine the fraction of nodes that need to be susceptible for the a disease to transmit, in analogy to the number 2 for the Bethe lattice where each node has 3 neighbors (percolation threshold would be $1/\mathcal{A}$ instead of $1/2$ for the Bethe lattice.).

Networks can also be characterized by a *cliquishness* quantified by the *clustering coefficient* [40, 41, 42]. A clique is defined as a troangle, i.e. a set of 3 nodes that all are connected to each other. For each node i the clustering coefficient is defined as the fraction of cliques it participate in, in units of max possible numbers of cliques [42] if all its neighbors was connected to each other.

$$C_i = \frac{number\ of\ pairs\ of\ neighbors\ to\ i\ connected\ to\ each\ other}{k_i \cdot (k_i - 1)/2} \tag{4.9}$$

As a measure for the entire network, the *global clustering coefficient* is defined as the average clustering coefficient over all nodes. A large clustering coefficient indicates large locality in the sense that neighbors of a given node tend to be directly connected. A social network with high clustering reflects a society where nearly everybody has common friends. For the Erdös-Rényi network the global clustering coefficient is

$$C = p \sim \frac{\langle k\rangle}{N} \tag{4.10}$$

since the probability that two of a node's neighbors are connected is p. The total number of length 3 cycles in such a network is $N \cdot C = \langle k\rangle$, which is independent on the size N of the network.

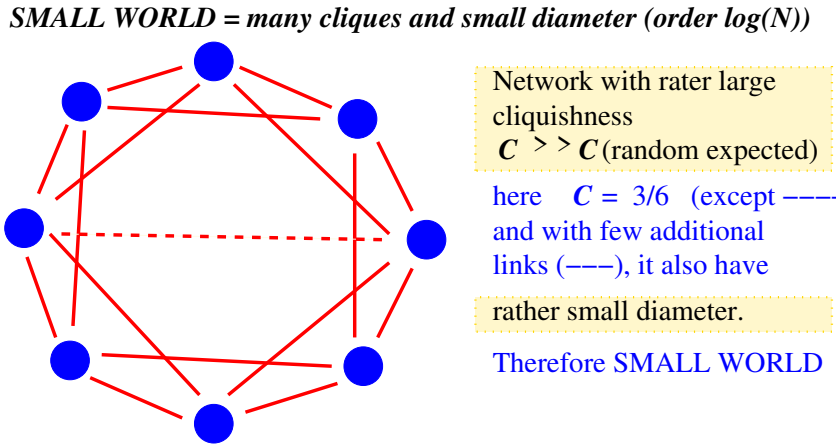


Figure 4.5: **Classical small world network.** Originally introduced by Watts and Strogatz [42]. See figure for explanations.

Mini Tutorial: Can you give a heuristic argument for why the numbers of 3-loops should be independent of network size? (N attempts to make loops, but each attempt only $1/N$ probability to succeed).

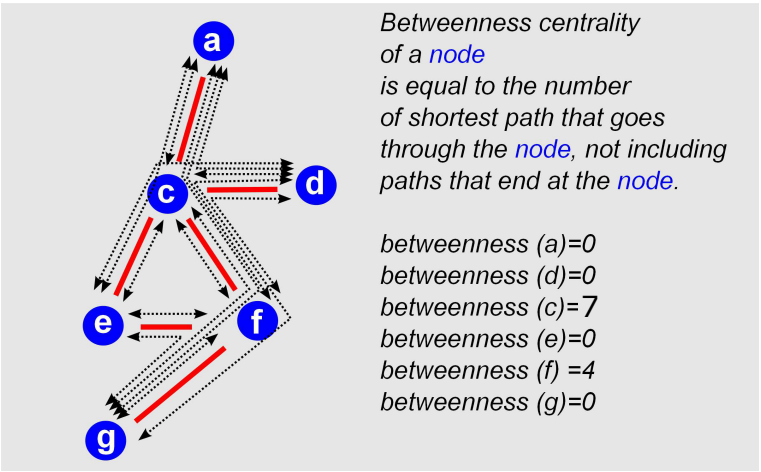


Figure 4.6: **Illustrating the betweenness of a node.** Betweenness measures the communication traffic that runs through the node, assuming that all pairs of nodes send messages to each other and always use the shortest path between them [43, 44].

A network is defined as having small world property, when it has a relatively large cliquishness, while still having a diameter of order $log(N)/log(k)$. Many networks are indeed found to have this interplay between global accessibility and local cliques. This universal tendency was coined to as "small world property" by Watts and Strogatz, see Fig. 4.5.

Another noteworthy concept, which helps characterize network topologies is that of *betweenness centrality* of nodes, illustrated in Fig. 4.6. The measure

ranks nodes according to how centrally they are placed in the network. There are indications that proteins with high betweenness centrality in molecular networks tend to be more important than proteins on more peripheral network locations [45, 46, 47].

4.1.4 Adjacency matrix

It is occasionally useful to represent a network in terms of a matrix, A_{ij} where the existence of a link from node i to node j implies that $A_{ij} = 1$. Absence of a link from i to j implies that $A_{ij} = 0$. A non-directed network is accordingly represented by a symmetric matrix $A_{ij} = A_{ji}$, because a link from node i to j implies that there is also a link from node j to i.

Imagine the population of a disease/virus that is placed on a few of the nodes of a matrix represented by a vector with components v_j, $j = 1, 2, \ldots, N$. The dynamics of disease spreading is a process which allow copying to all neighbor nodes, represented by the update:

$$v_i = \sum_j A_{ij} v_j \,, \tag{4.11}$$

alternatively, in matrix notation, $\mathbf{v}$(t+1)=$\mathbf{A} \cdot \mathbf{v}$(t). Applying the matrix $\mathbf{A}$ multiple times correspond to applying the infection cycle to neighbors, and next nearest neighbors, and back again many times, allowing the disease to present in multiple copies on each node. In the long time limit $\mathbf{A}^t$, $t \to \infty$ is dominated by the eigenvector corresponding to the largest eigenvalue. The size of this eigenvalue reflects the region in the network where self-amplification is strongest.

Another interesting feature can be obtained by applying the matrix $lim_{t\to\infty}$ $\mathbf{A}^t$ to a vector $v_i = \delta(i, j)$, i.e. a vector that is non-zero only at node j. When iteratively applying A on such a single node input one expand the input 1 from node j to first neighbor nodes, and then to next nearest neighbor nodes, and so on. Nodes that are reached by several paths becomes larger than 1, and for example the first node j takes a value equal to its connectivity after 2 applications of A. However, nodes that cannot be reached from node j will never take a finite value. Thus applying A many times will only give non-zero entries for nodes which are directly or indirectly connected to the node j.

Mini Tutorial: If $\mathbf{A}$ is the adjacency matrix for a network without self interactions, then what is in the diagonal of the matrix $A \cdot A = A^2$. What does the trace of A^2 represent for the network??

A network can often be sub-divided in separated clusters, where all nodes within each cluster are directly or indirectly connected, but where no path exist between different clusters. In matrix notation these clusters could be mapped into a matrix $\mathbf{A}$ with a block diagonal form (each block corresponding to a

cluster). A *modular network* is one where one allows a few links between clusters, but where links between pairs of nodes in different clusters are much less likely than for pairs that both lie within the same cluster [48, 49, 50, 51]. A modular network correspond to a nearly block diagonal matrix, where the blocks along the diagonal are supplemented with a few non-zero entries at other places in the matrix.

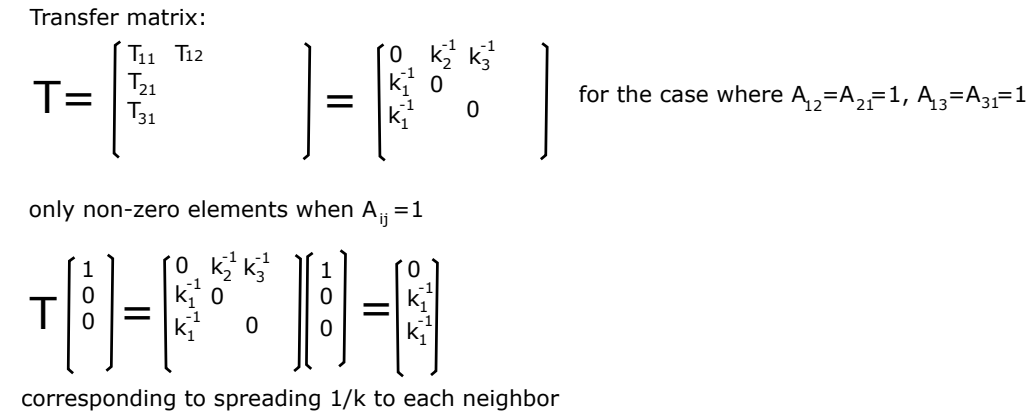


Figure 4.7: **Transfer matrix for a random walk on a network.** For details see illustration.

Notice that the matrix representation opens for simple manipulations. The number of triangles in a non-directed network without self-links is

$$n(\triangle) = \frac{1}{6} \cdot trace\left(\mathbf{A}^3\right) \tag{4.12}$$

where the factor $6 = 2 \cdot 3$ comes from "going" in a clockwise direction, respectively counter-clockwise direction around each triangle (a factor 2), and from the 3 contributions associated to the fact that any of the three nodes in the triangle give a contribution to the count. Please notice that the adjacency matrix in this case only should include links between nodes, and should not include any self-links ($trace(\mathbf{A}) = 0$).

The adjacency matrix $\mathbf{A}$ reflects the replication dynamics of a copying activity of the network. Insight into the network topology may also be gained from other processes on networks. Of these the most popular is the diffusion like process where one follow random walkers on the network [49, 50, 52, 53, 51]. At each time the walker takes a random step along one of the (out) links from its current node. The overall random walk on a network can be described by a transfer matrix $\mathbf{T}$ which has non-zero values at the same matrix elements as the adjacency matrix $\mathbf{A}$, but where each link from a node i is assigned a weight $1/k_i$, where k_i is the total number of links *pointing away from* the node i. Thus a node which has only one link pointing away is will direct all its random walkers that way. In contrast, for a node where there is 10 links pointing away then the probability to walk along a given link is only 1/10. See Fig. 4.7.

If one considers a population of random walkers on a network, the distribution of these walkers approaches a steady state in which the diffusion current flowing from a node i to a node j is exactly balanced by that flowing from j to i. This is satisfied when the average number of walkers on every node i is proportional to its connectivity k_i:

$$\frac{v_i^1}{k_i} = \frac{v_j^1}{k_j} = constant\,, \tag{4.13}$$

that is, the more connected, the more walkers will visit the node. A randomly "test particle" thus tends to visit sites with high (in) degree more often. A variant of this random walk is used for ranking nodes on the famous Google internet site [2)].

Networks can have many features that extend beyond simple connectivity and small loops. One of these is *modules*, or clusters of nodes that are more connected within each other than between the modules. Examples of modules are shown in Fig. 4.8 and 4.9. Fig 4.8 shows a network constructed by using transmission of twitter messages (L. Weng et al, Scientific Reports 2012).

Mini Tutorial: Inspect the CEO network on Fig. 3.8. Is the number of triangle loops larger or smaller than randomly expected?

4.1.5 "Scale free" networks

A common feature of many real networks is that their degree distribution is very broad [54, 55, 39], e.g., networks between genes in a cell (Fig. 4.1), between internet servers (Fig. 4.10) and between members of important social clubs (Fig. 4.9). In all these cases the purpose of the network is to act as a backbone of information transfer, on which work/functions are appended. And in all cases, the distribution of links and connections is far from homogeneous.

In fact, in real networks the number of nodes with connectivity k may often be approximated by a power law,

$$n(k) \;\propto\; \frac{1}{k^\gamma}\,, \tag{4.14}$$

with exponent γ that nearly always is in the narrow range between 2 and 2.5. It is remarkable that one rarely finds network exponents beyond this range. Apparently, the exponent cannot be so little (below 2), that the hubs carries all the links. The exponent 2 is again the famous Zipf exponent, found in many many systems.

[2]Google uses a ranking that is proportional to the probability that a walker that starts at a random site, visits the node within ~ 5 random steps. This probability is calculated from $\mathbf{T}^5 \cdot \mathbf{v}_1$, where the vector $\mathbf{v}_1$ has unity in all entries. This procedure is easily implemented on directed networks, in practice with the addition that a walker in a node without exit link is moved to a random other node.

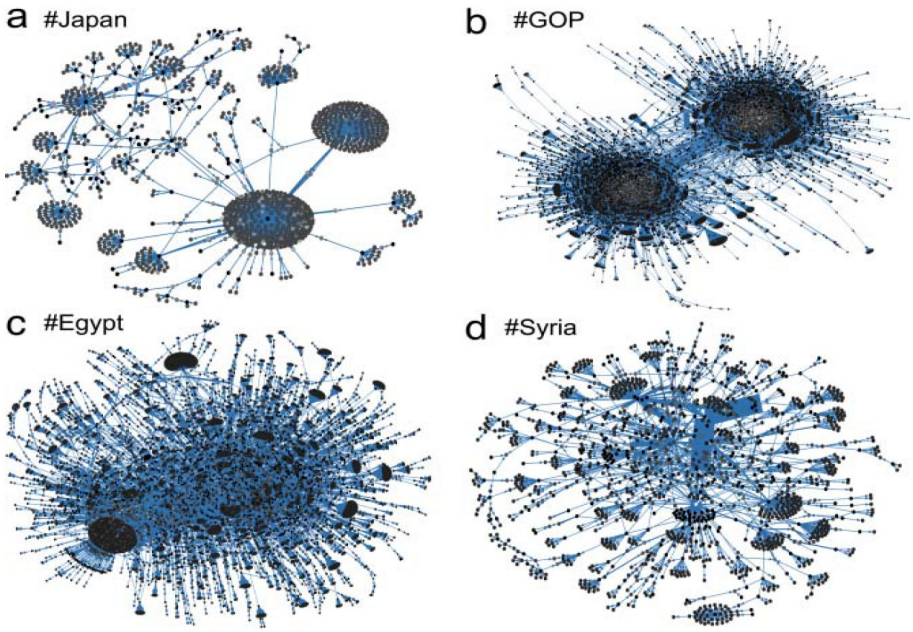


Figure 4.8: **Networks constructed from links formed by retweeting twitter messages with various hashtags.** A) is related to a Japanese earthquake in March 2011, B) Grand Old Party (GOP) is related to the republican party in USA, C),D) are related to the Arab Spring in 2011, focusing on Egypt and Syria, respectively.

Notably, such a scale-free degree distribution is far from the Poisson distribution from the previous sub-chapter (eq. 4.3). That is, if one assigns links completely randomly between nodes, one would in practice never obtain a scale free distribution. Scale free distributions are beyond simple randomness, although the way they appear does involve randomness. We will discuss various history-dependent processes for obtaining scale-free distributions at the end of this chapter.

Mini Tutorial: Which problems would there be in constructing a network with N nodes with exponent $\gamma = 1.8$?

4.1.6 Amplification of "epidemic" signals

One aspect of a broad connectivity distribution is the possibility of a huge amplification $\mathcal{A}$ of disturbances/signals. This feature can be inferred from eq. 4.8. For broad connectivity distributions $\mathcal{A}$ typically depends on the node with the highest connectivity. To see this, assume a scale free network. Then, from

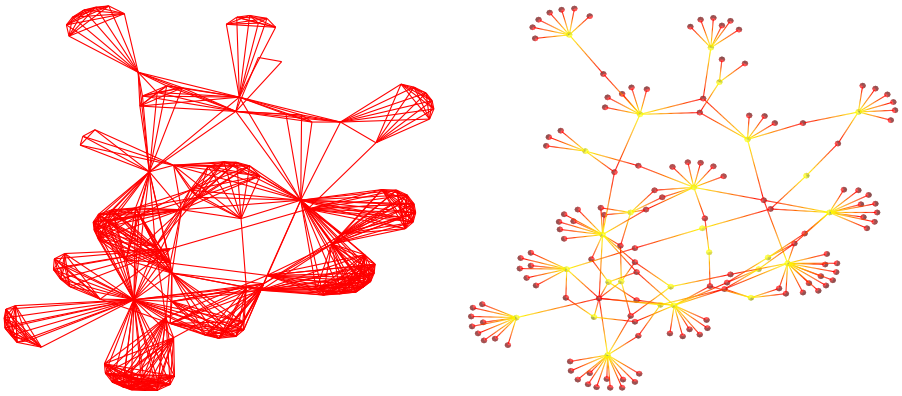


Figure 4.9: **Professional social network.** Small sub-section of the social network of company executives in USA. On left panel links are between the CEO's that sit on same board, the right panel shows the bipartite version of the network, where persons (red) and boards (yellow nodes) are separated.

eq. 4.8,

$$\mathcal{A} = \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 = \frac{\int_1^N \frac{k^2 dk}{k^\gamma}}{\int_1^N \frac{k dk}{k^\gamma}} - 1 \sim N^{3-\gamma}, \tag{4.15}$$

for $\gamma \in]2;3[$. In that case the denominator becomes independent of system size N whereas the numerator increases with N. Thus, for $\gamma < 3$, $\mathcal{A}$ is dependent of the upper cut-off in the integral, which represents the node with highest connectivity.

Mini Tutorial: Why does the denominator become independent of N in the above text?

Robustness and hubs: $\mathcal{A}$ can also be used to estimate the robustness of the overall connectedness of the network against removal of a fraction f of its nodes [3)]. This problem may for example have relevance in disease spreading, where f then would be the fraction of people that are vaccinated against a given potentially epidemic disease [60, 61, 62].

The break-up of the network after removal of a fraction f of the nodes is determined by the value of f at which its $\mathcal{A}$ becomes less than unity. If the initial network has the amplification factor $\mathcal{A}$, then after removal of f nodes, the amplification will be reduced by a factor corresponding to remaining node

[3]In the literature [58] the derivation was first presented using the average connectivity of a nodes in the end of links

$$\kappa = \mathcal{A} + 1 = \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{Var(k)}{\langle k \rangle} + \langle k \rangle$$

and one identifies the percolation threshold with this being = 2.

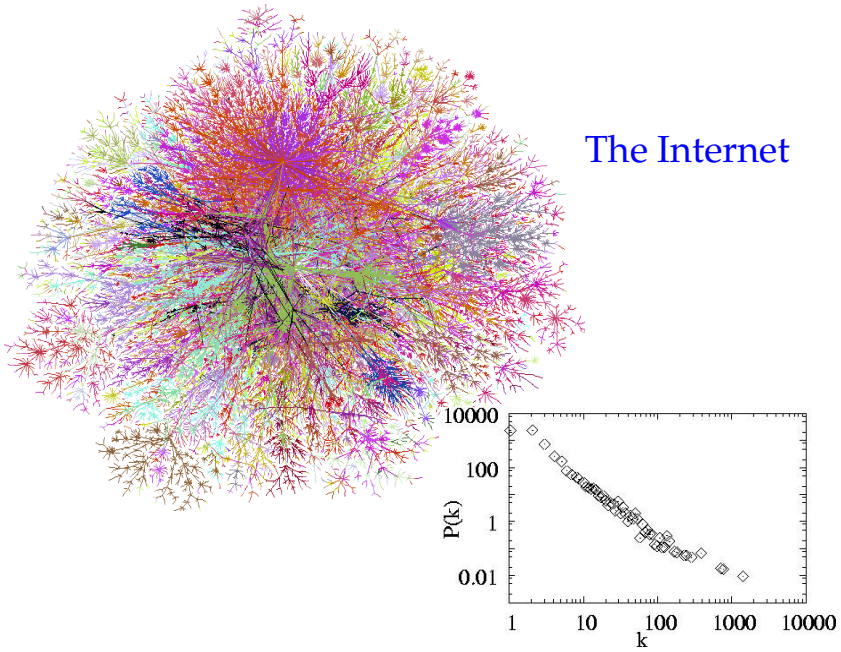


Figure 4.10: **Worldwide Internet:** Degree distribution in lower right corner. Notice the log-log scale, indicating the extremely broad range of degrees. Also notice that the distribution roughly fit a power law $P(k) \sim 1/k^{2.2\pm0.1}$

fraction, $1 - f$:

$$\mathcal{A} \to \mathcal{A}' = \mathcal{A} \cdot (1 - f) \tag{4.16}$$

This comes about because each remaining node will lose each of its links with probability $f' = f$ [4].

The network remains super-critical, when $\mathcal{A} \cdot (1 - f) > 1$ or

$$(1 - f) > \frac{1}{\mathcal{A}} = \frac{1}{\langle k^2 \rangle / \langle k \rangle - 1} \,. \tag{4.17}$$

I.e. the percolation threshold for the network is $1/\mathcal{A}$: This is the fraction of nodes that need to be conducting to make an infinite cluster.

Conversely, the critical fraction for vaccination against disease spreading [58] is

$$f_c = 1 - \frac{1}{\langle k^2 \rangle / \langle k \rangle - 1} \,. \tag{4.18}$$

This threshold is close to unity for scale free networks with degree exponent $\gamma < 3$, as is also seen in Fig. 4.12A. For narrower degree distributions the critical threshold clearly separates from 1, with value $f_c = 1 - 1/\langle k \rangle$ for ER networks when one simply uses the Poisson distribution property $\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle$ [5].

[4] Following a signal that enters into a node, see Fig. 4.4, each of its remaining $k - 1$ links have probability $(1 - f)$ to survive the pruning. Thus the local amplification $(1 - k) \to (1 - k) \cdot (1 - f)$ and the global amplification factor $\mathcal{A}$ is reduced by the factor $(1 - f)$.

[5] After a fraction f is removed from an ER network, the fraction F in the largest cluster obeys $F = (1 - f) \cdot (1 - e^{-\langle k \rangle \cdot F})$ [58]: That is, the probability that a given remaining node is

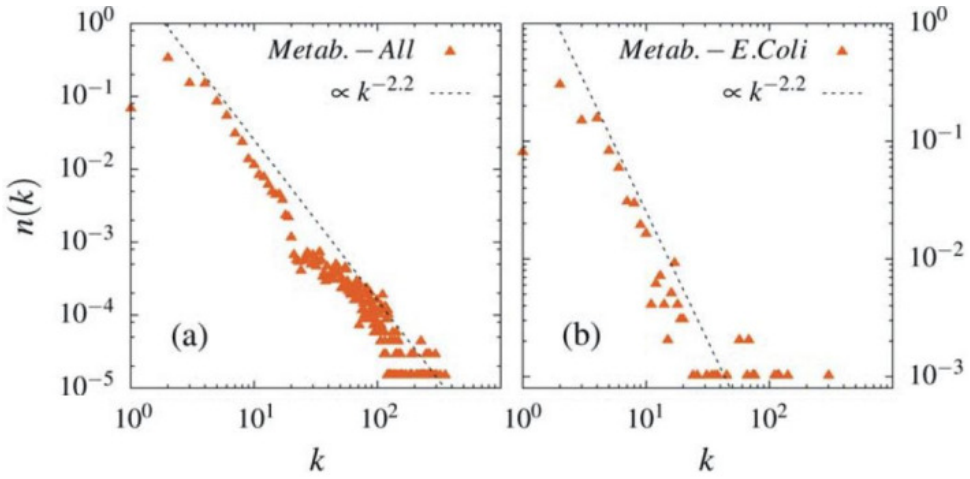


Figure 4.11: **Degree distribution of metabolites in networks.** Two metabolites are linked when one is entering the reaction (a substrate) and the other is a result (a product) of the same reaction. The left panel shows the average degree distribution for 107 organisms in the Ma-Zheng database [56]. The right panel shows the degree distribution from the metabolic network inside the bacteria *Escherichia coli*. Fig. from [57].

Questions

4.1) What is the minimal number of links needed to connect 100 nodes in one large component? (a collection of nodes that is directly or indirectly connected to each other). Hint: Just think, no equations or simulation needed.
Qlesson: Think about the maximum number of links and minimum number of links in a network, and the feature that most networks are closer to the lower limit (matrix with mostly zeroes and a few links, i.e. sparse).
4.2) What is the largest diameter one can have in a network with 100 nodes? Hint: Just think, no equations or simulation needed.
Qlesson: Think about the way to separate nodes from each other while still leaving a path between them. Perhaps networks are not necessarily about maximizing the ease of contact, but also about local protection from nonsense.
4.3) How does the diameter of a network scale with number of nodes N, when these are organized on square/cubic/... lattices in d dimensions? Consider, for example, 4096 nodes, organized in 1-d, 2-d, 3-d lattices. The nodes are thereby placed on the lattice sites, and the links are assigned between nearest neighbors along each of the lattice axes. Determine the diameter of the 4096 node network, when instead organized in a Erdös-Rényi network with an average connectivity of 6 (same number of neighbors as a 3 dimensional cubic lattice). Convince yourself that an Erdös-Rényi network has infinite dimension. Hint: Just think, no equations or simulation needed.
Qlesson: Infinite dimension is easily realized, and nodes are close, but it's easy to get lost anyway.
4.4) Generate Erdös-Rényi networks with $p = 3/(N-1)$ (3 neighbors per node on

not connected to the largest cluster is $exp(-\langle k\rangle \cdot F)$. Therefore, any of the remaining $(1-f)$ nodes will belong to this cluster with probability $(1-e^{-\langle k\rangle \cdot F})$. At critical conditions the largest cluster collapses and conforms to the scaling of the other clusters, see Fig. 4.12B.

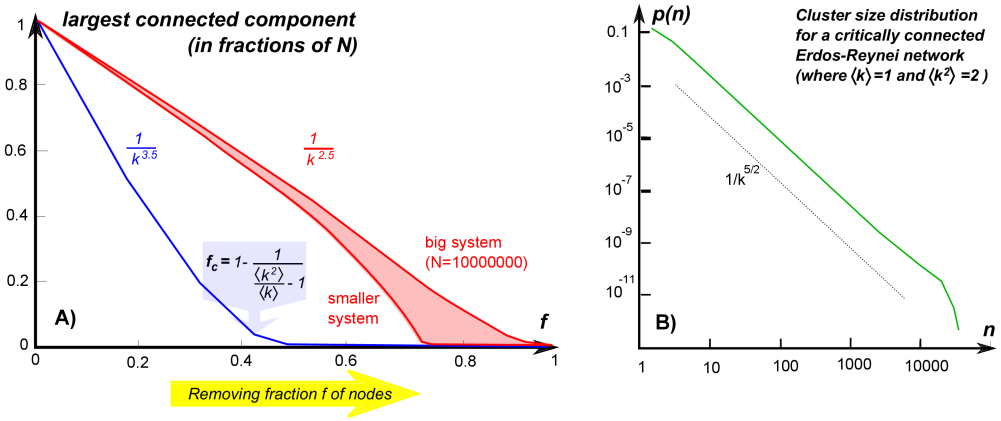


Figure 4.12: **Weight of the largest cluster.** A) The fraction of nodes that is in the largest cluster (F) as a function of the fraction of nodes removed[58]. The blue curve refers to a network with a relatively narrow degree distribution, $1/k^{3.5}$, which exhibits a critical threshold similar to ER networks. The red curves shows behavior of networks where $\langle k^2 \rangle$ is dominated by the largest hub in the network. The latter case is also simulated for different system sizes, demonstrating that very large networks remain connected until nearly all nodes are removed. B) Cluster size distribution in a critical ER network $P(n) \propto 1/n^{2.5}$ [59], implying that a random node will be in a cluster with size distribution $n \cdot P(n) \propto 1/n^{1.5}$. This critical distribution is also obtained as one removes random links from a well-connected ER network until $\mathcal{A} = 1$ and $\langle k \rangle = 1$.

average, and not allowing for self-interactions) for $N = 10$, $N = 100$ and $N = 1000$ and count how many triangles there are at various network sizes. How does the number of triangles change with N for fixed average connectivity (fixed connectivity implies that p decreases with system size)?
Qlesson: Infinite dimension is easily realized, and nodes are close, but it's easy to get lost anyway.
4.5) Visualize the above network for N=100 using for example cytoscape (download from web), the Python package networkx, or the matlab functions $B = graph(A)$ and $plot(B)$ where A is the adjacency matrix. The cytoscape does not read the matrix, but instead a sequence of lines, where each line has two nodes that are connected by a link.
Qlesson: Networks can be visually nice.

4.6) Consider a non-directed network which only consists of one large component. Prove that a random walker after infinitely long time will visit each node with probability that is proportional to its degree. Notice that random walks on networks is at the core of search engines such as Google, where however the walkers also do other moves to deal with properties of directed networks (e.g., to not get "trapped" indefinitely). Hint: Consider a steady state flux between two connected nodes with different degrees.
Qlesson: Application of detailed balance that was introduced in the Metropolis al-

gorithm.

4.7) Construct a network of $N = 100$ nodes subdivided into 10 different classes with 10 nodes in each. Generate a random network where each node has approximately 0.01 links between modules, and nodes within same class have probability 0.5 to be connected (remember that the generated matrices have to be symmetrical, and that diagonal elements have to be zero). Calculate number of loops (=triangles), and compare this with number of triangles when all links are randomized (i.e. when one distribute the about 5+1 link per node randomly across the lattice).

Qlesson: There are much more triangles in the modular network.

4.8) Generate a random network of size $N = 100$ with 150 links (average degree $\langle k \rangle = 3$) and monitor the size of largest component as nodes are removed subsequently. Do the same when removing links subsequently, maintaining all nodes.

4.9) Use the following equation for the fraction F of the nodes that remain in the largest cluster of a Erdös-Rényi network, after a fraction f is removed [58]:

$$F = (1 - f) \cdot (1 - e^{-\langle k \rangle \cdot F}) \, . \tag{4.19}$$

Determine the critical value of f, where $F = 0$, and examine F as f approaches this critical point. Hint: expand the exponential in the above equation using that $\langle k \rangle F \ll 1$ close to the critical point.

Qlesson: Compare the value f obtained with that obtained by using the amplification factor equation 4.15.

4.10) One vaccination strategy is to vaccinate people at the end of links. By vaccinating a fraction f' of the nodes, one removes a fraction

$$f = f' \cdot \langle k^2 \rangle / \langle k \rangle \tag{4.20}$$

of the links. Argue for this equation, and express the vaccination fraction f needed to stop epidemics on a scale free network with $N = 10000$ nodes and degree distribution $n(k) \propto k^{-2.5}$.

Qlesson: Your neighbor has, on average, more link than you do.

4.2 Analyzing Network Topologies

4.2.1 Randomization: Constructing a proper null model

Mini Tutorial: How would you randomize a given network?

In order to identify non-trivial topological features of networks one needs to go beyond the single node property defined by the degree distribution. Such analysis may may help us to understand the function-topology relationship of a particular network: Is there some kind of pattern or motif that are particularly frequent across the network. The key idea in such type of analysis is to compare the network at hand with a properly randomized version of it.

Aiming to pinpoint patterns one step beyond the degree distribution one should compare the network at hand with a random network with exactly the same degree distribution. The best way to generate such random networks is shown in Fig 4.13. The idea is to swap links, pair by pair, multiple times until

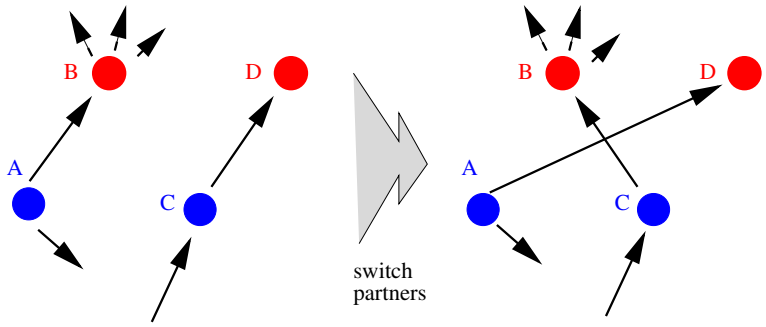


Figure 4.13: **Network randomization.** One step of the network randomization algorithm [63, 64]: A pair of directed links A→B and C→D switches connections in such a way that A becomes linked to D, while C becomes linked to B, provided that none of these resulting links already exist in the network. An independent random network is obtained when this procedure is repeated a large number of times. This algorithm conserves in- and out- connectivity of each individual node.

all nodes in the network are assigned new random links [63, 64]. For a system with L links, then after t link swaps, the probability that a given link is not changed is

$$Fraction(unchanged\ \ links) = (1 - \frac{2}{L})^t \approx e^{-2t/L}\,, \tag{4.21}$$

which becomes insignificant when the number of "swaps" t becomes substantially larger than L. Notice, that one cannot allow all random swaps: If there is already a link between two nodes, then an attempted assignment of a second link should be aborted. Also note, that one may keep a network connected in one big component by simply only allowing "swaps" that maintain overall connectedness.

Given an adequately randomized network, or better, a sample of about 1000 independent random networks, the significance of any quantifiable measure Q is given by the probability that a random network has same value of Q as the real network.

Q could, for example, be the number of short loops, that is, triangles in the network. Alternatively, Q could, for example, be the number of links between nodes with connectivity 10 and nodes with connectivity 20.

The excess ratio of a quantity Q is quantified by

$$Substance = R(pattern) \ = \ \frac{N(Q)}{\langle N_{random}(Q)\rangle}\,, \tag{4.22}$$

whereas the the significance level of Q is quantified by its Z score:

$$Significance = Z_{score} \ = \ \frac{N(Q) - \langle N_{random}(Q)\rangle}{\sigma_{random}(Q)}\,, \tag{4.23}$$

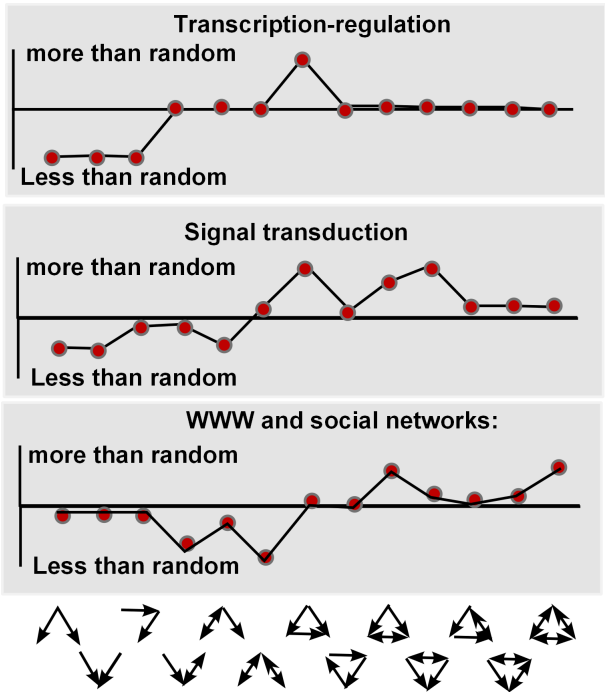


Figure 4.14: **Network motifs and their abundances.** Super-families of motifs as suggested by statistical analysis of different types of networks by Milo et al. [65]. One sees that feed-forward is over-represented in transcription and signaling networks. In contrast, then social networks and the world wide web tends to have over abundance of other types of three-node motifs. In all cases, triangles in some form are favored, reflecting some universal tendency of clustering/locality.

where $N_{random}(Q)$ is the number of times the pattern occurs in the randomized network. Here

$$\sigma^2_{random}(Q) = \langle N_{random}(Q)^2 \rangle \; - \; \langle N_{random}(Q) \rangle^2 \tag{4.24}$$

is the variance among the random networks. The significance amounts to standard hypothesis testing, simply quantifying how many standard deviation the real networks is away from the assumption that the network is formed by random assignment of links. If "Z-score" is $Z_{score} = 2$, the probability is 2.5% to obtain a comparable pattern in a randomly generated network. Notice that a pattern can be significantly over-expressed without really having a substantial excess! I.e, Z could be much larger than 10 with R only 10% in excess ($R = 1.1$).

Mini tutorial: Why is significance and substance not the same?

By considering occurrences of higher order local patters of control in various other networks, ref. [66, 67] found particularly frequent motifs in particular types of networks. Fig. 4.14 show their results in terms of over/under repre-

sentation of motifs in different types of networks. For example for biological regulatory networks ref. [66, 67] report an excess of feed-forward motif and Ref. [68] suggested that these "feed-forward" acts as a noise filters. That is, they would only allows a signal to pass when it is persistent in time. In any case, the abundance of certain recurrent patterns indicates a repeated function, and thus some feature that re-emerges again and again in the formation of some networks.

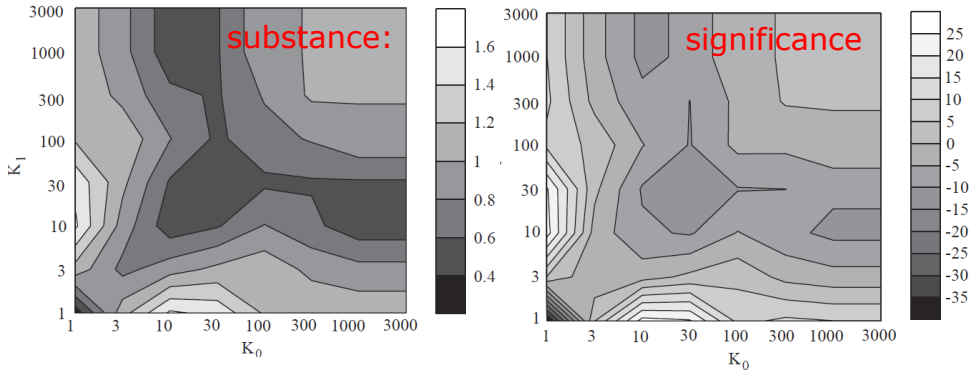


Figure 4.15: **Substance (R), respectively significance (Z) of connections between nodes of different degree in the hardwired internet.** The correlation [63] between internet servers is quantified in terms of probability to be connected in units of randomized expectation (left) and in terms of the Z-score (right). Notice that the absolute effect is about a factor 2, but since the considered internet had 6474 nodes (January 2000), the significance of especially the excess connection between nodes of degree 15 to nodes of degree 1 is huge.

Apart from considering motifs, there is in fact an even simpler correlation that can be considered, namely the extent to which nodes of certain number of links prefer, or tend to avoid, each other. This type of correlation is quantifies in terms of the correlation profile where one compares the extent to which nodes af degree, say between 10 and 15, tend to connect more than expected at random to other nodes of, say, degree between 100 and 150. The correlation profile for the hardwired internet in Fig. 4.15 shows relatively many links between nodes of similar degree (from ref. [63]). This is called an *assortative network*.

Fig. 4.16 elaborate further on the relation between triangle motifs (in non-directed networks the triangles correspond to most of the motifs from Fig. 4.14) and correlation profile. These plots are generating structured networks by repeatedly applying the link swapping move described before, but only requiring moves which optimize some cost function that respectively punishes (left), or favors cycles of length 3. For each potential move one calculates the change in the number of triangles $Diff = N_{after}(\Delta) - N_{before}(\Delta)$. In the left panel one only accepts moves when this number is less than or equal to zero. In the right panel we only accept moves when the number is larger or equal to

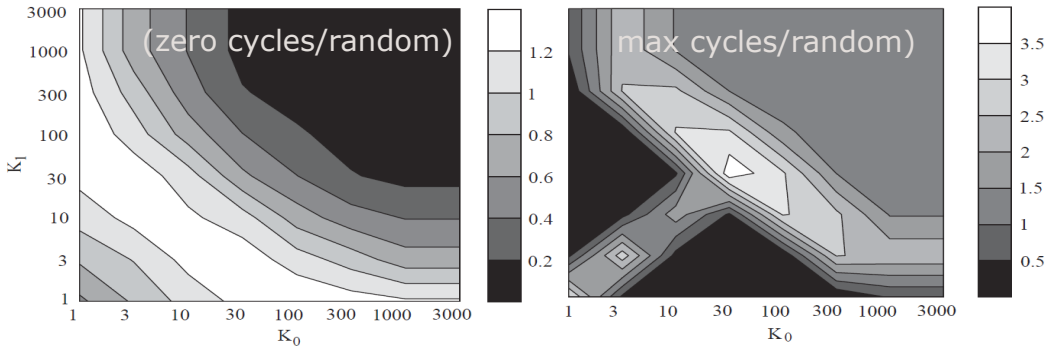


Figure 4.16: **Excess ratio** R**.** (eq. 4.22). The paper analyse some artificial networks. In both cases it quantify correlatioms between connectivities of pairs of nodes that are directly linked to each other. In the left panel we consider a network generated from a rewired version of the hardwired internet (that originally had 6,584 cycles of length three) to a network without triangles. The dark region in upper right corner illustrate that the network disfavour connections between highly connected networks, compared to a randomized version that preserve the degree distribution. In the right panel we rewire the hardwired internet to maximize the number of triangles (obtaining a network with 59,144 cycles of length three). In that case one obtain a network where high degree nodes are more likely to be connected.

zero. Effectively, this corresponds to a zero temperature Monte-Carlo update with an energy that is equal to plus or minus the number of short loops in the network.

In Fig. 4.16, left panel, one sees that the absence of triangles correlates with the absence of links between high degree nodes. In contrast, the right panel shows that the surplus of triangles is associated to many links between intermediate and high degree nodes. In a subsequent sub-chapter we will elaborate more on the extent to which networks are assortative (high degree nodes "like each other," thus forming a topological hierarchy (Fig. 4.18), respectively dis-assortative (nodes of similar degree avoid each other).

Null-models may be extended to take into account more elaborate structures that are already established as important features. That is, we for example freeze the correlation profile, and then randomize the network while only accepting minute variations in this profile. This could again be done in a "Metropolis kind of way" (*compare* Sec. 1.2), assigning an energy to the deviation of observed correlation profile:

$$H = \sum_{K0,K1} \left(N(K_0,K_1) - N_{random}(K_0,K_1)\right)^2 / N(K_0,K_1) \ , \qquad (4.25)$$

where K_0 and K_1 are suitably binned connectivities of nodes and $N()$ is the number of links that connect nodes in the corresponding bins. At each update step one now attempts a link swapping, and calculates the change in H, $\Delta H = H(new) - H(old)$.

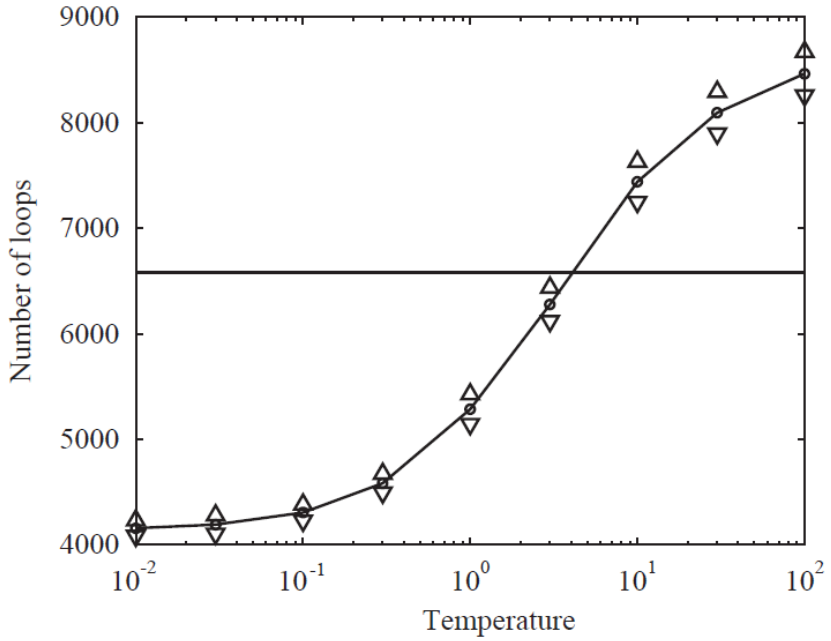


Figure 4.17: **Exemplifying the temperature algorithm.** Number of cycles of length three for rewired internet while fixing correlation profile (having 6,584 cycles of length three). The number of triangles depends on the temperature used. When temperature is zero, this corresponds to the null model where the correlation profile is fixed. When temperature is large, this corresponds to the standard model where only the connectivity distribution is fixed. The horizontal line is the number of loops in the real hardwired internet.

- If $H(new) < H(old)$, the move is accepted.
- If $H(new) > H(old)$: the move is accepted with probability $p = e^{-(H(new)-H(old))/T)}$.

where "T" is some ad hoc assigned effective temperature for our sampling. Fig. shows that the correlation between low and high degree inverses the null expectation for the number of triangles. Thus, the real network has fewer loops than a random network, but a larger number of loops than a random network, where links between high degree nodes are maintained.

4.2.2 Algorithm generating a synthetic scale-free network

In subsequent text we repeatedly explore networks of a given degree distribution, without stating where they come from. This is not entirely trivial to obtain: when the degree distribution is rather flat, $n(k) \propto 1/k^{\gamma}$ with $\gamma \sim 2$, there are several nodes that have many links, and perhaps not enough nodes with small number of links to actually make it possible to assign the correct number of links to every node. We here define a prescription for generating

a random network with whatever degree distribution, which will work if it all possible to find enough partners.

To generate a network of N nodes with a degree distribution $n(k) \propto k^{-\gamma}$, with a maximum of one link between each pair, one first assigns each node i a degree k_i from this distribution. That is, one selects a random number $r \in]0,1[$ and solve for K

$$\frac{\int_K^N dk/k^\gamma}{\int_1^N dk/k^\gamma} = r \ , \tag{4.26}$$

where $K = K_i$ then is the selected number of links for node number i.

After assigning a number to each node we need to link these nodes up, where each node should have its assigned number of K_i links. This is done from the top down, starting with the node which should be assigned the largest number of links. Thus one starts with the node at highest degree and connects it to other nodes, linking it to the node of next largest degree and subsequently connecting lower nodes until all links for this high-degree node are assigned [69]. Subsequently, lower degree nodes are assigned neighbors in the same orderly way, until all nodes have their assigned degree.

The network are now extremely ordered, each nodes are linked to all nodes of higher degree. In fact, such a network where high degree nodes are connected preferentially to high degree nodes is called assortative. Thus the generated network is scale free, but it is not randomly put together.

To obtain a random network the network has to be randomized, using a procedure that accomplishes the pairwise link swapping described in previous section. Importantly one need to make a large number of edge-swapping, of order $L \cdot ln(L)$ where L is the total number of links in the system.

4.2.3 A hierarchy measure of networks

A way to illustrate differences between networks with degree distributions $\sim 1/k^2$ and networks with steeper degree distributions is to consider links, or absence of links, between highly connected nodes. For $\gamma \sim 2$ there are so many links in the system, that these hubs tend to be directly connected. In contrast, the hubs tend separate as γ increases towards 3. This can be quantified in terms of the topological hierarchy [69, 70], which assigns rank proportional to degree, and assigns a network hierarchy measure equal to unity, if the shortest path between any pair of nodes follows a hierarchical path. That is, the shortest path first has go from low to high degree and then from high to low degree. The fraction of pairs in the network that is connected by such hierarchical paths is denoted as $\mathcal{H}$.

Fig. 4.18 shows $\mathcal{H}$ as a function of γ for random scale free networks with different scaling exponent. For degree distribution $P(k) \sim 1/k^{2.2}$ one see a hierarchical structure high degree nodes typically are connected directly to each other. In contrast, then for $P(k) \sim 1/k^{2.8}$ the high degree nodes are

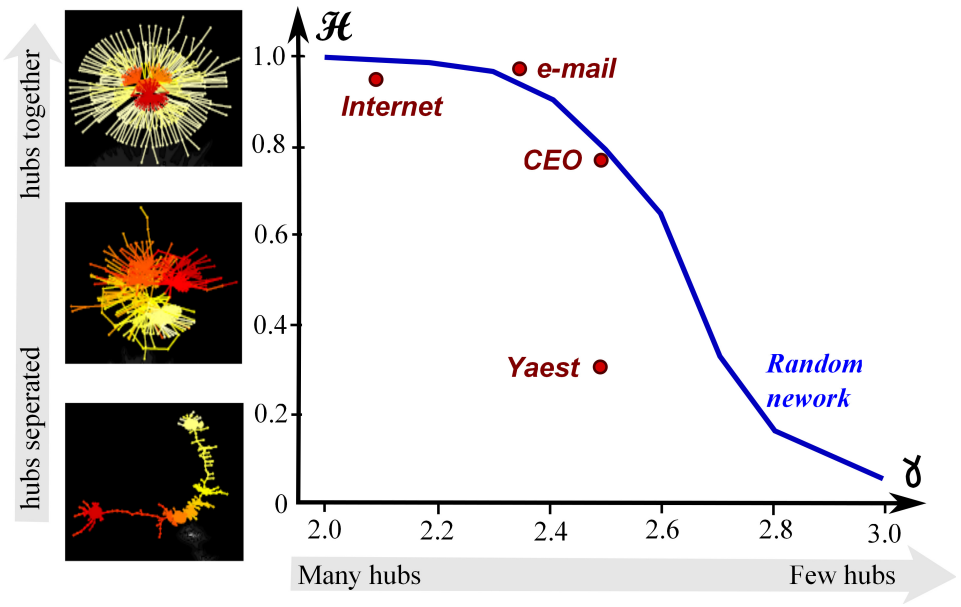


Figure 4.18: **Topological Hierarchy.** Reproduced from [69]. The left part of the figure illustrates maximally hierarchical (top) and anti-hierarchical (bottom) networks of size $N = 400$ with a $1/k^{2.5}$ degree distribution. The main figure shows how $\mathcal{H}$ depends on the degree distribution in random scale-free networks with distribution $f(k) \propto 1/k^{\gamma}$. As the degree distribution narrows, the hubs tend to separate and for $\gamma > 3$ the hubs are distributed along a "stringy" network that is dominated by nodes of low degree. The network examples is the hardwired network of internet routers, an email network, the network of board members in American companies and the yeast protein-protein interaction network.

often not directly connected to each other, and even the random network is anti-hierarchical.

The figure also compares with a few real world networks, leaving us again with the challenge to properly define what a random network actually is. Notice that $\mathcal{H} \sim 1$ for $\gamma \sim 2$, whereas $\mathcal{H}$ decreases to 0 for somewhat larger γ, reflecting that the hubs then have too few links to connect directly to each other. This is also reflected in the behavior of the probability that none of K neighbors have a degree higher than K (see Fig. 4.19):

$$\begin{aligned} P(K \textit{ node is local top}) &= \left(1 - \frac{\int_K^N k^{1-\gamma} dk}{\int_1^N k^{1-\gamma} dk}\right)^K \\ &\sim (1 - K^{2-\gamma})^K \\ &\sim exp(-K^{3-\gamma}) \,, \end{aligned} \tag{4.27}$$

where we use that, when $\gamma > 2$, the integral is dominated by its lower boundary. Subsequently, we use that $1 - K^{2-\gamma} \sim exp(-K^{2-\gamma})$.
Thereby $P(K \textit{ node is local top})$ becomes large for $\gamma \sim 3$, even when K is rather small.

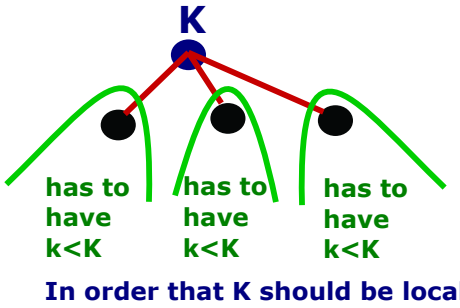


Figure 4.19: **Detecting a hierarchical network.** Calculation of the probability that a node with connectivity K has higher connectivity than all its neighbors. When this happens for nearly all highly-connected nodes, then the network will be characterized by highly-connected hubs that are not linked up to each other: A non-hierarchical network results.

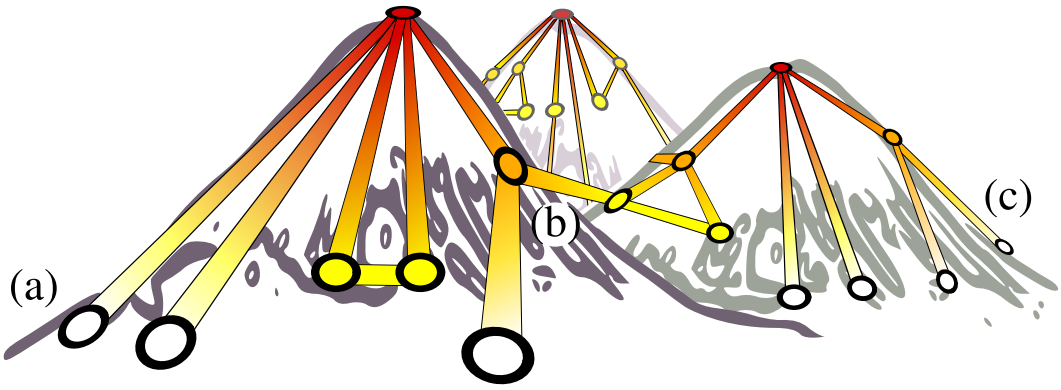


Figure 4.20: **A network as a degree landscape with mountains and valleys.** The "altitude" of a node proportional to its degree. A route over one mountain corresponds to making a degree-hierarchical path ((a) to (b)) while climbing over more than one mountain breaks the degree-hierarchical path ((a) to (c)).

Thus, in the perspective where we imagine the network as a mountain landscape with nodes placed at a height proportional to connectivity, then there will be many small "tops" for steep power laws [69]. That is network with steep power laws is non hierarchical.

If one instead consider networks with $\gamma \sim 2$ then the above equation states that there is no "tops": Any even moderately large K-nodes cannot be a local top. However, this would not be true for the single node with largest connectivity in the system. This then become the only "top" in the system. The system is then perfectly hierarchical: All short distance paths go through the highest connected node in the system.

Emphasizing the degrees of nodes as a key property, the network topology can be visualized using a landscape analogue, with mountains (high degree nodes) and valleys (low degree nodes). Within this interpretation, the internet is one single mountain with first ascending and then descending hierar-

chical paths, whereas biological networks form rough landscapes with several mountains and broken hierarchical paths. To quantify the topology and make it possible to compare different networks, one can measure the typical width of individual mountains and the separation between different mountains (Fig. 4.20).

In particular, in Fig. 4.21 we complement the methods to generate random networks (random one-mountain landscapes) [63] with preserved degree sequences, to generate *ridge landscapes*. In its simplest implementation, we assign a random rank to every node in a network, and organize the nodes hierarchically based on their rank. This method creates non-random networks, distinguished by a separation of hubs (leftmost network). Alternatively one may assign each node a number equal to its rank, and thereby generate the highly centralized system in the rightmost panel.

Thus, Fig. 4.21 shows topologies that all originate from a random scale-free network (shown in Fig. 4.21e) with degree distribution $P(k) \propto k^{-2.5}$ and system size $N = 400$. The extreme networks, the perfect random-rank hierarchy in Fig. 4.21(c) and the perfect degree-rank hierarchy Fig. 4.21(g) ($\varepsilon = 0$), surround the networks with increasing error rate towards the random scale-free network with $\varepsilon = 1$ in the middle (Fig. 4.21(e)).

The intermediate networks are generated by a rewiring rule where links also sometimes, with probability $\varepsilon > 0$ are re-shuffled randomly. Notice, in particular, that when we consider already a small perturbation on the stringy network of left panel, the diameter of the network collapses as seen in Fig. 4.21(d). Note that the color gradient indicates that the random-rank hierarchy is still intact at this stage, and that the hubs ("mountain tops") are separated.

Both when we organize the network according to random numbers, or according to degree, we obtain higher clustering, (meaning: more triangles), than in the completely randomized network (not shown). This clustering is expected, as organization along any coordinate tends to make friends of friends more alike. The effect is stronger in the degree-rank hierarchy, since the clustering automatically increases further, when the hubs with their many links are connected.

Questions

4.11) Generate a random network with $N = 1000$ nodes and with power law distributed connectivity (that is, degree): $n(k) \propto 1/k^{2.5}$, for $k = 1, 2, \ldots, N$. Compute the number of nodes that have no neighbors with higher degree (number of local maxima, i.e., "tops"). Rewire the network such that only moves which lower the degree difference between nodes are allowed. Compute again the number of local "tops". Finally, try to rewire node links, where one only allows moves that increase degree differences, and then compute the number of "tops".
Qlesson: Networks topology is much more than its degree distribution.

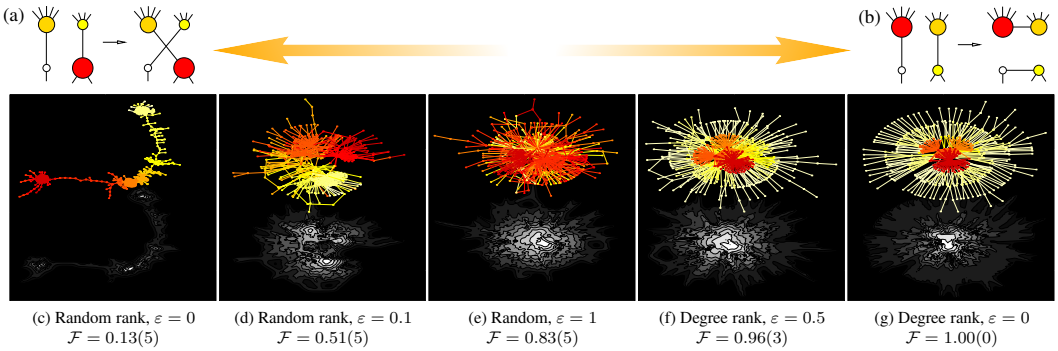


Figure 4.21: **"Degree landscapes" organized from "ridge landscapes".** (c), via random landscapes (e), to peaked one-mountain landscapes (g). The links are swapped pairwise to connect high-ranked nodes and organize the nodes globally according to their rank (color coded from red for high rank, to white for low rank), with random swaps at different rates ε. The rank is set randomly to the nodes, as in the swap example in (a), in (c-d), and proportional to the degree of the nodes, as in the swap example in (b), in (f-g). The random network in (e) corresponds to $\varepsilon = 1$. The corresponding degree landscapes are color coded according to altitude from black (low) to white (high). The networks are scale free with an exponent $\gamma = 2.5$ and of size $N = 400$.

4.3 Models for Scale free networks

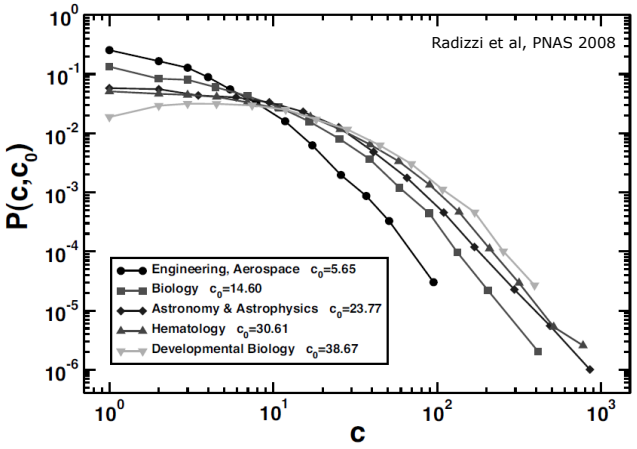


Figure 4.22: **Citations per article:** Notice that as the average of the field changes, so does the tail. The distribution is not really a power law but closer to a log-normal.

Mini Tutorial: Imagine that you jump from article to article, following random entries in the reference list. At a random time you then make a reference to one of the articles you visit. What is then the probability to reference an article as function of its previous number of citations?

We will now introduce two ways that generate close to scale free networks. As you will see, the two ways are fundamentally different and in fact also differ conceptually from any fine tuning to a critical point. Both methods will use time development and certain dynamical rules to obtain interesting networks.

But while one model is closely associated to eternal growth (it becomes boring in steady state), the other obtains its pattern in a ongoing steady state of network re-wirings and in this sense has some similarity with SOC. As nearly all studies of networks are essentially snapshots of only one instance, we at present have no real way to judge the dominating dynamics in any real system. In any case, each of the approaches have their correspondence in other problems from physics, complex systems, and social science.

4.3.1 Preferential attachment

The most famous way to obtain scale free networks is through an agent-based growth model, where nodes are subsequently added to the network, with links attached preferentially to nodes that are highly connected (Price (1976), Barabasi & Alberts (1999)). This process results in a growth model based on minimal information in the sense that each new link is attached to the end of a randomly selected old link. Thus, one connects new nodes with a probability proportional to the degree of the older nodes. Thus, highly connected nodes will grow faster. In other words, it "pays to be popular". After t steps, t nodes have been added and, within the simplest version, also t links.

Let $n(k,t)$ be the number of nodes with connectivity k at time t. The evolution of n is given by (Bornholdt et al. (2001)):

$$\begin{aligned} n(k,t+1) - n(k,t) &= \frac{(k-1)\cdot n(k-1,t) - k\cdot n(k,t)}{\sum_{k'} k' n(k')} \quad \text{for } k>1 \\ \Rightarrow \frac{dn(k,t)}{dt} &= -\frac{1}{\sum k' n(k')} \frac{d}{dk}(k\cdot n(k,t)) \quad \text{for } k>1\,, \end{aligned} \tag{4.28}$$

because the probability to add a link to a specific node of connectivity k is $k/\sum kn(k)$. In the above equation the first term represents the addition of a link to a node of connectivity $k-1$, thereby adding to the number of nodes at connectivity k. The second term represent the addition of a link to a node of connectivity k, thereby reducing the number of nodes with connectivity k by moving one of them to the next connectivity value.

Each added node is associated by one link, which has two ends, one at the new node and one at the node it is attached to. Therefore $\sum_k kn(k,t) = 2t$. Accordingly, the continuum limit:

$$\frac{dn}{dt} = -\frac{1}{2t}\,\frac{d(k\cdot n)}{dk} \tag{4.29}$$

To solve this equation we make the "ansatz":

$$n(k,t) = f(t)\cdot N(k)\,, \tag{4.30}$$

which gives:

$$2t\frac{f'(t)}{f(t)} \;=\; -\frac{1}{N}\,\frac{d(kN)}{dk}\,, \tag{4.31}$$

where $f(t) \propto t$ since $f(t) = 2t/\sum_k kN(k)$. Inserting $f'/f = 1/t$ in the above equation

$$-1 - \frac{dln(N)}{dln(k)} = 2\,. \tag{4.32}$$

From this one obtains $N \propto k^{-3}$:

$$n(k) \propto \frac{1}{k^3}\,. \tag{4.33}$$

Notice the if one instead added two links for each new node addition, each attached to this new node, then nominator and denominator in eq. 4.28 would both double. In the end the scaling behaviour would be the same. If, however, we also added links preferentially without adding new nodes, the result could be different. Thus if such links are assigned to existing nodes preferentially in both ends then they would contribute with two times the nominator (one for each link end), but only contribute with 2 in the denominator. This peculiar growth would make the ratio smaller and change the scaling law $1/k^3$ to $1/k^\gamma$ with $\gamma \in [2;3]$, see Fig. 4.23

In fact, $\gamma = 2/(2-p)+1$ where p is probability to add a new node with a link, and $1-p$ is the probability to add a new link with both ends attached preferentially to already existing nodes. (this is proved by noting that a link with both ends preferentially added give double the nominator as in the above equation. Thus, the numerator is multiplied by $p + 2\cdot(1-p) = 2-p$, whereas the denominator remains $2\cdot t$).

The preferential growth model was originally proposed in an entirely different contexts, relating to modeling of human behavior exhibiting skew distributions in a wide variety of aspects. Yule (Yule, G. U. (1924). Philosophical Transactions of the Royal Society B. 213 (402–410): 21–87.) , Pareto (Pareto, V. (1898). "Cours d'economie politique". Journal of Political Economy. 6) and Zipf observed that the empirical observation of family sizes of taxonomic species, fortunes of humans, and number of times a particular word is used, all tend to be distributed with power laws

$$n(s) \sim 1/s^2;\,. \tag{4.34}$$

Here, n could for example denote the probability to have a word repeated s times. This distribution is marginal in the sense that the average

$$\langle s\rangle \;=\; \int_{min}^{max} \frac{sds}{s^\tau} \tag{4.35}$$

receives a substantial contribution from the upper cut-off of the integral. That is, at power laws wider than $1/s^2$, say $1/s^{1.5}$, a huge fraction of the probability

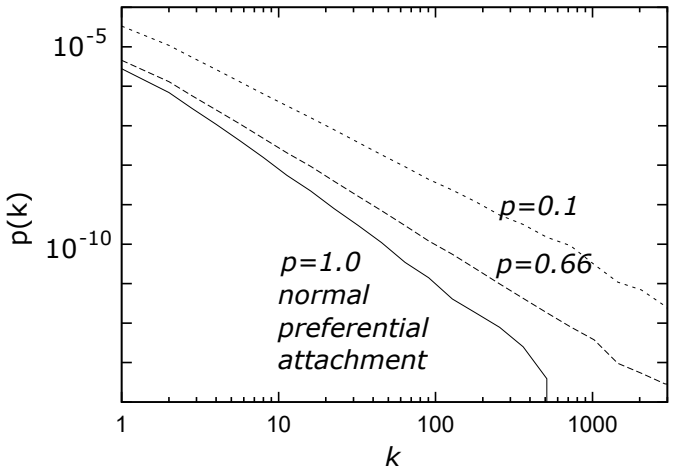


Figure 4.23: *Connectivity distribution in a preferential attachment model for three different values of p. p is a parameter that specifies the fraction of times one adds a node with a link. At probability* $1-p$ *one instead adds links where each end of the link is attached to an existing node with a probability proportional to the connectivity of that node (see also Bornholdt et al., 2001). For all values of p one obtains a power law. The power law obtained depends on the connectivity distribution and approaches* $1/k^3$ *when* $p = 1$.

mass is bound relatively close to the upper cutoff. On the other hand, a narrower scaling like $1/s^{2.5}$ will have an average that is independent of the upper cutoff. Thus, in the case where s denotes resources or money, then social systems should become unstable when the exponent τ becomes less than 2. Popularly speaking, the rich then become so rich that by confiscating their fortunes society could increase the wealth of the rest by a substantial amount.

H. Simon (1955) suggested that the $1/s^{\tau}$ behavior reflected a human tendency to preferentially give to those that already have. As H. Spencer stated already in 1855, the human perception of importance of a particular subject is proportional to how often one has heard about this subject. An observation that relates to the absurdity of much of the public debate.

For networks, a feature of this history dependent mechanism of positive feedback is that the most connected nodes also are the oldest. This property can sometimes be tested, and often fails to be fulfilled. Another feature is that in steady state, supplementing preferential attachment with random elimination of nodes in fact breaks the scale-free degree distribution (because removal of any node preferentially reduces the number of links from the high degree nodes).

Scale-free behavior, obtained by preferential attachment in networks, depends on the ongoing growth process, as seen in Fig. **??**. For networks the removal of small degree nodes has the side effect that it preferentially removes links from the high degree nodes, and thus limits their continued growth. Thus, ongoing growth is often not that realistic.

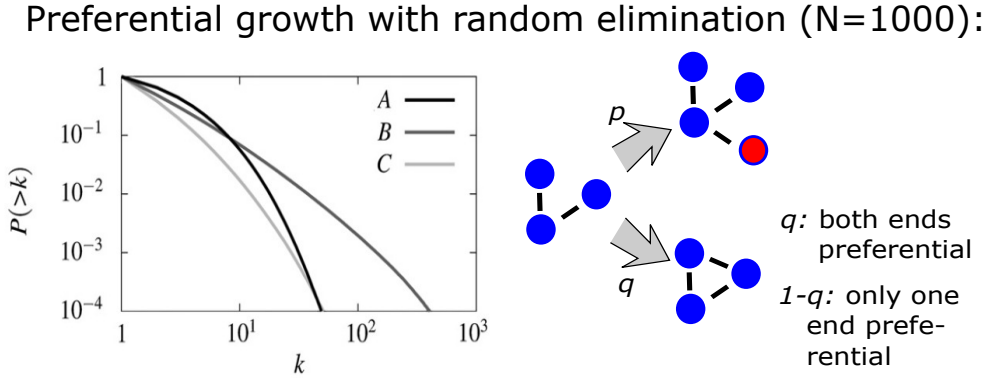


Figure 4.24: **Preferential growth with random elimination of nodes.** Nodes are randomly eliminated, when the total node number exceeds 1,000 (from :Gronlund et al., Physica Scripta 71, 680 (2005)). A: p=0.4, q=0, B) p=0.4, q=1 and C) p=0.8 and q=1 where p is how often one adds a new node with a link and $1-p$ is the alternative where one adds a link without adding a node. $p = 1$ is standard preferential growth. When adding a new link, q is the probability that both link ends are assigned preferentially, whereas $1-q$ is the probability that only one link is assigned preferentially.

Mini Tutorial: Argue why rich-get-richer dynamics could be considered for sizes of families of related species.

4.3.2 Merging and creation

A second scenario for generating scale-free networks is the merging and creation model introduced by Kim et al. (2003). In this model one generates a scale-free network by merging nodes, thereby generating new nodes. The classical model for merging with creation was introduced several times in the literature, for example by Takayashu *et al.* in 1988 (PRA 37,3110):

- Consider a set of N numbers, s_i, $i = 1, 2, \ldots, N$. Initially, all these numbers are set equal to unity. At each time step select two of these numbers, say s_i and s_j, at random and add them, yielding a new number $s_i(new)$: $s_i(new) = s_i + s_j$. Then reset the number at position j to unity: $s_j = 1$.

Mini Tutorial: What would happen if one only merged, and did not inject, any new particles in the above merging and creation model?

The above model persistently injects one new unit into the system, because of the assignment $s_j = 1$. The merging-injection model gives a distribution $p(s,t)$ that develops towards a steady state distribution $p(s)$ as given by the

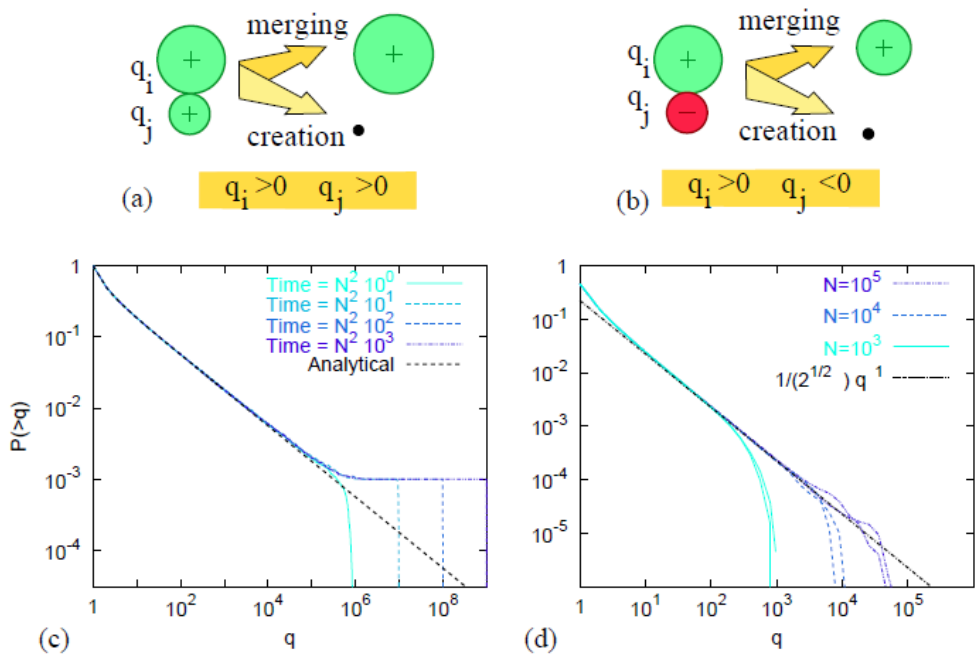


Figure 4.25: **Basic merging and creation process.** (a,c) also implemented while allowing different signs (b,d). Panels c,d show cumulative plots at different times for an $N = 1,000$ system. From Minnhagen *et al.* Physica A 340, 725 (2004).

dynamical equation

$$p(s,t+1) - p(s,t) \;=\; \sum_{u=1}^{s-1} p(s-u,t)\cdot p(u,t) - 2\cdot p(s,t)$$

In the steady state

$$\begin{aligned} p(s,t+1) - p(s,t) \;&=\; 0 \\ &\Rightarrow\; \sum_{u=1}^{s-1} p(s-u)\cdot p(u) = 2\cdot p(s)\;. \end{aligned} \tag{4.36}$$

The first term represents the sum of all combinations of merging that result in a size-s cluster. The second (loss) term $2\,p(s,t)$ comes from selecting two numbers, which each could be of size s with probability $p(s,t)$.

When a number of size s is selected, it will be merged and then become larger that s, thus surely becoming removed from the bin with clusters of size s. The final steady state equation is only true for all the clusters which could be in steady state. The largest cluster could not be in steady state. At any time, the largest cluster $s = s_{max}$ would occasionally be merged with another smaller one, and thus it can only grow. Thus the system rely on a steady injection of small clusters, $s = 1$, and its overall mass will grow as the larges cluster grows and separates from the power law that govern the remaining population of clusters.

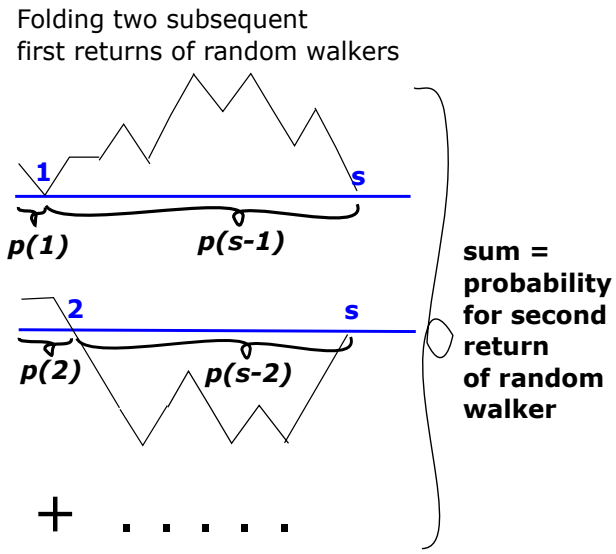


Figure 4.26: **Illustration of random walks.** Emphasis is on the first return distribution $p(s)$ where s is the time axis. The second return distribution of a random walker must fulfil $p_{second}(s) = \sum_{u=1}^{s-1} p(u)p(s-u)$ in terms of first returns. Further, we make the reasonable assumption that second returns must scale proportional to first returns, $p_{second}(s) \propto p(s)$ for large s. Then $\sum p(u)p(s-u) = const \cdot p(s)$. (see also eq. 3.14 and Fig. 3.5). Thus, the first return of random walk should fulfil the recursion relation (apart from a constant that can be absorbed in a pre-factor). First returns of random walkers scale as $p(s) = s^{-3/2}$, a scaling which should therefore also solve the merging-injection model.

In steady state, after long time then the gain of aggregates with size s should be equal to the chance that a cluster at size s is selected and merged with another cluster (the factor 2 comes from the fact that we select two clusters at each time-step). The above steady-state relation is fulfilled by a power law with probability to have a number with size s that scales as

$$p(s) \propto \frac{1}{s^{3/2}} \tag{4.37}$$

($\sum_u u^{-3/2}(s-u)^{-3/2} \approx 5.22 \cdot s^{-3/2}$). Notice that the prefactor of 5.22 can be absorbed by setting $p(s) = 5.22/s^{3/2}$. A further argument for the scaling can be obtained by using that eq. 4.36 is fulfilled for the first return of random walkers is illustrated in Fig. 4.26. This figure illustrate two of the terms in the above sum, with a thin line marking the random walk for the corresponding contribution. The figure illustrate two terms in the sum, the one where first "first return" happens after 1 step, and the one where first "first return" happens after two step. Obviously all second returns at "time" s comes about from walks where the first "first return" is somewhere between 1 and $s-1$.

The power law requires constant injection of "mass" at $s = 1$ (Takayashu *et. al.* 1988), making $p(1)$ a fixed finite number (=1/2, see appendix to this chapter). In the random walk picture this secures that we start the walk. For a simulation see Fig. . As already mentioned, then the model also generates one very cluster (or number) number that constantly grows beyond any size (as the injected mass ultimately ends up in this aggregate).

The merging and creation model was first suggested as a model for aggregation supplemented by on going injection of new dust in some region of the interstellar space (Fields & Saslow (1965)).

Remarkably, if one changes the model slightly by allowing evaporation, i.e. removal, the scaling changes. To accomplish this, at each step select two numbers, merge them, but then remove 1 from the aggregate: $(s_i, s_j) \to (s_i + s_j - 1, r)$ where r now is some random injection with mean $\langle r \rangle = 1$. This model, that amounts to a "mass conserving" version of the above model, predicts a scaling that is markedly steeper,

$$p(s) \propto 1/s^{5/2} \tag{4.38}$$

as was also shown by solving this model analytically (Minnhagen *et al.* Physica A 340, 725, 2004). Hence, objects are merged and a small random quantity is emitted into the list of other objects. In any case, mass conservation will change the power law the steeper exponent 5/2.

Merging in Networks: Now let us return to networks where merging of nodes for example could represent merging of companies and that then combine into a larger company with a their combined business associations. Creation then correspond to startup companies with few customers.

For networks (Minnhagen *et al.* Physica A 340, 725 (2004)) one time step of the merging-injection algorithm consists of selecting a random node,and one of

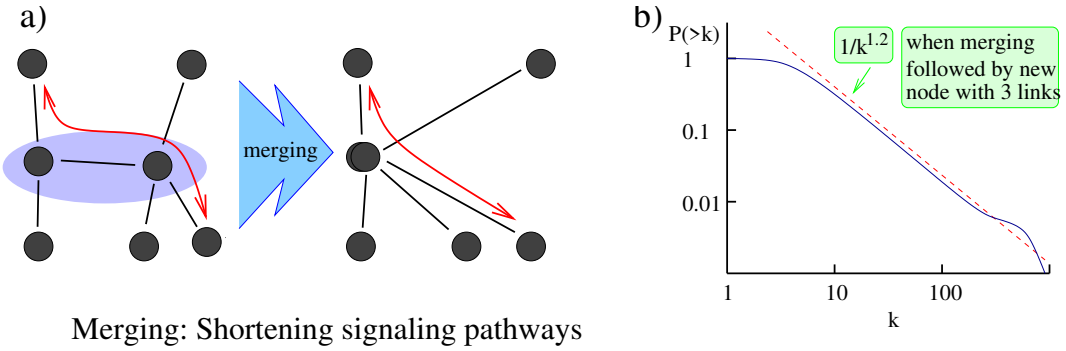


Figure 4.27: *Merging and creation model of Kim et al. (2003). In addition to the merging step shown, a steady state network demands addition of a node for each merging. After a transient this evolutionary algorithm generates networks with scale-free degree distributions, as illustrated in right panel. The scaling exponent for the steady state distribution depends weakly on the average number of links that a new node attaches to the older ones.*

its neighbors. These are merged into one node, see Fig. 4.27a). Subsequently, one adds a new node to the network and links it to a few randomly selected nodes. This merging model partly corresponds to the above merging model with evaporation, thus suggesting an exponent of $1/k^{5/2}$.

For networks, the merging-creation result generates a nearly scale-free network with exponent $\gamma \sim 2.2$, see Fig. 4.27b). On another note, considering companies that may consider to merge with others. The justification for merging two companies could be efficiency, to to shorten communication pathways and increase efficiency. Creation, on the other hand then reflect introduction of new companies and their few start up relations.

In contrast to the preferential attachment model in previous sub-section, the merging/creation model does not demand persistent growth. Instead, it suggests an ongoing dynamics of an evolving network which at any time has a scale free degree distribution.

There is also a merging-creation model of evolving networks that have some potential relevance to solar flare dynamics. Solar flares is associated to eruptions from the solar corona, which in turn is associated to the complex phenomena of turbulence in magneto-hydrodynamics. That is, there is strong magnetic fields on the sun, and the corona consists of charged particle made of protons and electrons. Occasionally magnetic field lines converge into bundles and makes solar spots with magnetic north poles or south poles. These field lines may merge and grow or annihilate each other dependent on directions.

The above "mess" in the solar atmosphere inspire a model with donor ($q > 0$) and acceptor ($q < 0$) nodes to be connected via directed links, see Fig. 4.28(a) and (b). q would then be the number of magnetic field lines in the solar spot, and the sign of q marking its direction , i.e. whether it is a north or a south pole. As links will always be between positive and negative nodes, the network is bi-partite: There is two set of nodes, and there is only links

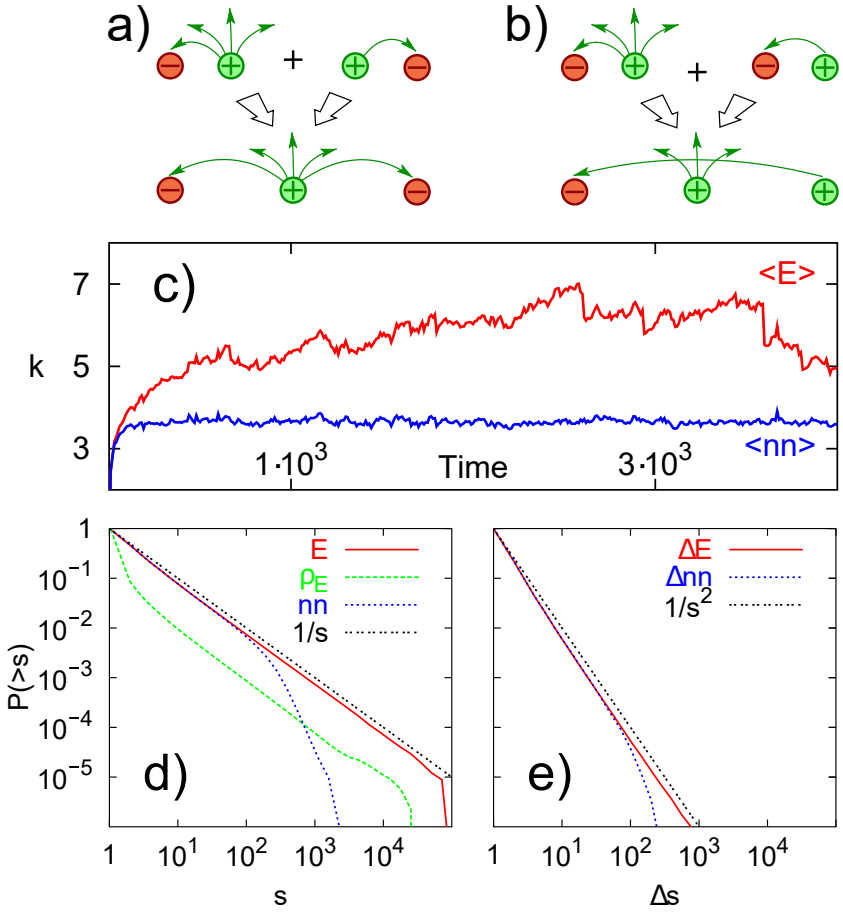


Figure 4.28: *Bi-partite merging/annihilation model with* **a)** *and* **b)** *illustrating merging events. Positive vertices (donor nodes) have outgoing links and negative (acceptor nodes) have incoming links.* **c)** *The dynamics of the average number of links associated to each node, $\langle E \rangle$, (upper curve) and the average number of neighbors $\langle nn \rangle$ (lower curve),* **d)** *The cumulative probability distributions for: number of links incoming or outgoing from a node, E (solid curve); number of neighbors, nn (dotted curve); edge density, ρ_E defined as the number of parallel links connecting two nodes (dashed curve). The distributions for all quantities are scale free $P(>s) \sim 1/s^{\gamma-1}$ with $\gamma = 2$.* **e)** *Cumulative distributions of changes in the number of links due to merging, ΔE and number of neighbors Δnn. The distributions are power laws $P(>\Delta s) \sim \Delta s^{1-\tau}$ with exponent $\tau = 2\gamma - 1 = 3$.*

between the two types; There is no links between nodes that are both in one of the two subset.

Each node may have a different number of links, but at any time a given node cannot be both donor and acceptor. Further, we allow several parallel links between pair of nodes, representing the number of field lines that connect them. Thus we here talk about a network model where some nodes are connected by stronger links than other pair of nodes.

At each time step, two nodes i and j are chosen at random. The update is then:

- **1)** Merge i and j. There are now two possibilities:
  - **a)** If i and j have the same sign, all the links from i and j are assigned to the merged node. Thereby, the merged node has the same neighbors as i and j had together prior the merging, see Fig. 4.28(a).
  - **b)** If i and j are of opposite sign, the resulting vertex is assigned the sign of the sum $q_i + q_j$. Thereby, a number $max\{|q_i|, |q_j|\} - |q_i + q_j|$ of links are annihilated in such a way that only the two merging nodes change their number of links. This is done by reconnecting donor nodes of incoming links to acceptor nodes of outgoing links, see Fig. 4.28(b).
- **2)** One new vertex is created of random sign, with one edge being connected to a randomly chosen vertex.

This bipartite network model predicts power laws associated to the dynamics of reconnections between the nodes (This power laws and a more geometric version of the above model was first studied in the solar flare model of Hughes et al. (2003)). In this regard it is interesting that the number of links per node, k, is distributed with scaling $P(k) \propto 1/k^2$. This was also obtained for the "number of loops at foot-point" in Hughes et al. In addition, the distribution of re-connection events Δk counted as the reduction of k when two nodes of different sign merge is distributed as $P_\Delta(\Delta k) \propto 1/\Delta k^3$. This is in fact similar to the distribution of "flare energies" in the solar flare model of Hughes et al.

This suggest a simple perspective on solar corona dynamics. Perhaps ongoing merging is a main reason for scale-free behavior for magnetic activity in the solar atmosphere. The bipartite model has its analogy in a scalar model, with matter/antimatter as studied in the early 1990s by Krapivsky (1993).

Questions:

4.12) Simulate preferential growth in its original "rich gets richer" version solved by Herbert Simon in 1955. That is, at each step add 1\$. With probability p this amount is added to an already existing person, with probability proportional to the wealth of this person. In case this is not happening, that is with probability $1 - p$, introduce a new person with a fortune of 1. Explore the behavior for p close to

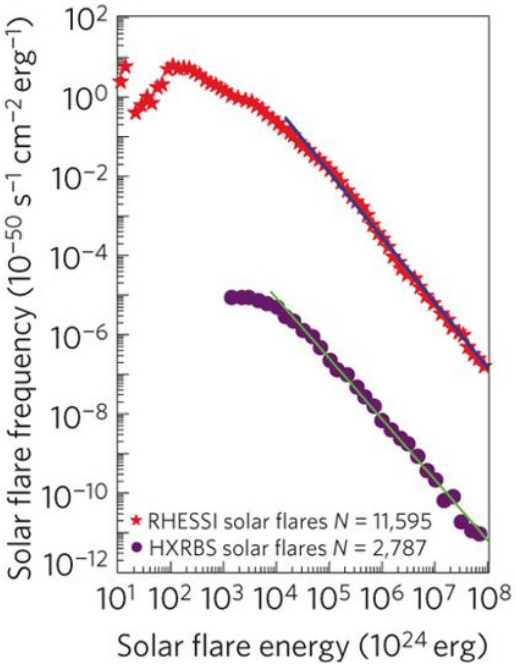


Figure 4.29: **Energy distribution in solar flares**. Reproduced from [11], suggesting that $P(E) \propto 1/E^{1.65}$

1. That is in the limit where one very rarely gives money to the people who have nothing.
Qlesson: Preferential growth indeed gives scaling. $p = 1/2$ correspond to the preferential attachment growth model for networks.
4.13) Repeat the above, but now supplemented by the rule of removing a random person each time the number of persons in the system exceeds 1,000.
Qlesson: Here, the rich gets richer dynamics remains robust to removal. This is because the rich are not affected by other people's elimination.
4.14) Consider the preferential network growth model and let $n(k,t)$ be the number of nodes with connectivity k at time t. Find the analytical expression for the steady state distribution of $n(k)$ for different probabilities of adding new nodes p (with $1-p$ being the addition of new links).
Qlesson: In this case, each time one adds a link the denominator in eq 4.28 grows with 2 but the nominator with less. Thereby $\gamma = 1 + 2/(2-p)$ that approaches 2 when $1 - p \sim 1$.
4.15) Let's, for the time being, ignore the network aspect and simply simulate the merging/creation model in terms of a set of integer numbers k_i, $i = 1, 2, \ldots, n$, (with $n = 1,000$) which are updated according to

$$k_i, k_j \rightarrow k_i = k_i + k_j \;\; and \;\; k_j = 1 \; . \tag{4.39}$$

Show numerically that this generates a steady state distribution of the sizes of the numbers in the set $n(k) \;\propto\; \frac{1}{k^{\tau}}$ with $\tau \sim 1.5$.
Qlesson: Merging with constant influx can indeed generate power laws.
4.16) Simulate the merging/creation model in terms of a set of integer numbers k_i,

$i = 1, 2, ...n = 1000$, which are updated according to

$$k_i, k_j \rightarrow k_i = k_i + k_j - 1, \quad and \quad k_j = \delta \, , \tag{4.40}$$

where δ is 0, 1 or 2 with equal probability and we only allow updates where all $k_i \geq 0$. Thus, some of the $1,000$ numbers can be zero, for later to be merged and replaced with nonzero numbers from elsewhere in the system. Show that this procedure generates a steady-state distribution of the sizes of the numbers $p(k) \propto \frac{1}{k^\tau}$ with $\tau \sim 2.5$.
Qlesson: In merging the resulting distributions are very dependent on a finite but persistent loss term.

Lessons:

- Networks are both about connecting large systems, but also about keeping individual nodes isolated from most other nodes.
- Networks often have broad degree distributions, $N(k) \propto \frac{1}{k^\gamma}$ with $\gamma \in [2, 2.5]$. Such scale free networks can be efficient in transmitting diseases because of the large value of its amplification factor

$$\mathcal{A} = \frac{\int k^2 n(k) dk}{\int k n(k) dk} - 1 \, ,$$

  a quantity that can used to estimate the critical point for signal transmission across the network (percolation threshold).
- Real Networks are formed by some dynamic process, taking place on much longer time scales than the dynamics associated to signaling or transport across the network. This chapter suggested two possible dynamics that generate power laws: "*Rich gets richer*" and perpetual *merging* with a noisy evaporation or injection.

Supplementary reading:

Newman, Mark. Networks. Oxford university press, 2018.

Cohen, Reuven, and Shlomo Havlin. Complex networks: structure, robustness and function. Cambridge university press, 2010.

Barabási, Albert-László. "Linked: The new science of networks." (2003): 409-410.

4.4 Appendix: Formal solution to merging

More formal arguments of the scaling for the simple merging and creation model presented above can be found in (Takayashu *et. al.* 1988) and in (Minnhagen *et al.* Physica A 340, 725 (2004)). Here we present a direct solution using Generating functions (proof by Ruijie Wu). A generating function is a way of encoding an infinite sequence of numbers (here $p_k = p(k)$) by treating them as the coefficients of a formal power series.

This appendix provides an example of how to use this powerful method for discrete models. Define

$$G(x) = \sum_{i=1}^{\infty} p_i x^i \tag{4.41}$$

where the variable x is not in itself interesting, but rather here to allow us to calculate the p_i's from differentiating with respect to x. We see directly that $G(0) = 0$ and $G(1) = \sum_i p_i = 1$. Now express

$$\begin{aligned} G^2(x) &= \sum_{i=1}^{\infty} p_i x^i \sum_{i=j}^{\infty} p_j x^j \\ &= \sum_{k=2}^{\infty} \sum_{i=1}^{k-1} p_i p_{k-i} x^k \\ &= \sum_{k=2}^{\infty} 2 p_k x^k \,, \end{aligned} \tag{4.42}$$

where we in the last equation use the basic equation defining p_k, i.e. the steady state equation for the merging-creation process,

$$\sum p_i p_{k-i} = 2 p_k \tag{4.43}$$

Now the sum from $k = 2$ to ∞ can be written in terms of whole sum from $k = 1$ to ∞ minus the contribution from $k = 1$, i.e. minus p_1:

$$\begin{aligned} G^2(x) &= 2 \cdot (G(x) - p_1 \cdot x) \Rightarrow \\ G(x) &= 1 \pm \sqrt{1 - 2 \cdot p_1 \cdot x} \,. \end{aligned} \tag{4.44}$$

Thus with the help of the basic recursion equation for p_k we got an expression for the generating function G. The sign choice $\pm$ is fixed by the constraint that $G(0) = 0$. Thus,

$$G(x) = 1 - \sqrt{1 - 2 \cdot p_1 \cdot x}$$

and the size of p_1 is set by $G(1) = 1 \Rightarrow p_1 = 1/2$:

$$G(x) = 1 - \sqrt{1 - x} \,. \tag{4.45}$$

We can now Taylor expand this expression for the generating function, yielding

$$G(x) = G(0) + G'(0)x + \cdots + \frac{1}{i!}G^{(i)}(0)x^i + \dots \,, \tag{4.46}$$

with the i'th derivative equal to

$$G^{(i)}(x) = \frac{1 \cdot 3 \cdot 5 \cdot 7 ... (2i-1)}{2^i} \cdot \frac{1}{(1-x)^{(2i+1)/2}} = \frac{(2i)!}{2^i \cdot i!} \cdot \frac{1}{(1-x)^{(2i+1)/2}} \,,$$

where we express the i'th order derivative in terms of factorials, thus allowing later use of approximate equations for these. The generating function is:

$$G(x) = \sum_{i=1}^{\infty} \frac{1}{4^i} \cdot \frac{(2i)!}{i! \cdot i!} \cdot \frac{1}{2i-1} \ \cdot \ x^i \tag{4.47}$$

Now we identify each order in x between this expression and the definition equation for the generating function $G(x)$:

$$p_n = \frac{1}{4^n} \frac{(2n)!}{n! \cdot n!} \frac{1}{2n-1} \tag{4.48}$$

Using Stirling's approximation $n! \sim \sqrt{2\pi n} \cdot (n/e)^n$:

$$p_n \propto \frac{1}{4^n} \frac{\sqrt{2n}}{n} \frac{(2n/e)^{2n}}{(n/e)^{2n}} \frac{1}{2n-1} \sim \frac{1}{\sqrt{n}(n-1)} \sim 1/n^{3/2} \,, \tag{4.49}$$

which is indeed the scaling guessed from assuming that the second return scale as the first returns of a random walker.

Chapter 5

Agent-based models

"The human brain is capable of a full range of behaviors and predisposed to none."

Stephen Jay Gould

5.1 Introduction

In this chapter we attempt to understand aspects of our social/biological world while assuming that it is built of many entities or agents that - by repeated actions - allow larger scale organizations to form.

Agent-based models are entering into the mainstream of computational approaches to economic and social systems. By simulating complicated economic relationships in terms of many different types of agents, explored recently in the economic literature by, e.g., Le Baron, it is hoped that one can obtain realistic models of societal dynamics. Here, we instead advocate another type of agent-based model which can describe "bottom-up" self-organization of complex systems. The models we describe have only few rules for simple agents, which are then iterated billions of times. Heterogeneity, diversity and complex behavior should be an emergent property, not an input. We term this class of models **"bottom-down"** approach, with the value of a model being greatest, when it is built on only few assumptions and parameters. A clearly "wrong" model is often more useful than an unclear model.

5.1.1 Schelling model of racial segregation

We will introduce agent-based models (ABMs) using several simple examples. ABMs were originally introduced by Von Neumann [71], in an effort to deal with system properties of many relatively simple agents that repeatedly use specified rules of mutual engagements [71, 72, 73, 74, 75]. Agent-based models can be defined both in terms of identical agents or in terms of a few archetypes of agents that together define a system, see Fig. 5.1.

Figure 5.1: **Agent-based models can involve intrinsically different agents.** Each agent is defined in terms of its particular caricature of a strategy. Examples include the Schelling model or the Rock-Paper-Scissor Game, both discussed in the subsequent chapter. Agents may also have identical basic characteristics, but then develop different properties as a function of "life experiences" during many updates of mutual interactions. The photographs were taken around a coffee bar at the Eastern shore of Taiwan.

ABMs are suited for addressing self-organization and emergent phenomena. ABMs have been used to study properties of living and, in particular, social systems, including segregation [72], traffic jams, evacuation behaviour [74], social insect organization [76], stock market dynamics, as well as dynamical pattern formation — with ABMs taking the form of a cellular automaton [71, 77, 78, 73]. We already encountered examples of agent-based models in for the spreading by contact (directed percolation, chapter 2), and by the "rich-get richer" behavior (preferential attachment in chapter 4). Here we will teach ABMs by examples, first using dynamics on simple geometry.

Boundary formation and segregation may most easily be modelled by using the famous ABM model proposed by T.C. Schelling [72]. This ABM was suggested as a description of spontaneous segregation of white and black neighborhoods in American cities. In a simple formulation the Schelling model consists of two types of agents, one "black" and one "yellow". Initially, one assumes an equal number of either color, and positions the agents randomly in a two-dimensional lattice with periodic boundary conditions in both directions.

At each step:

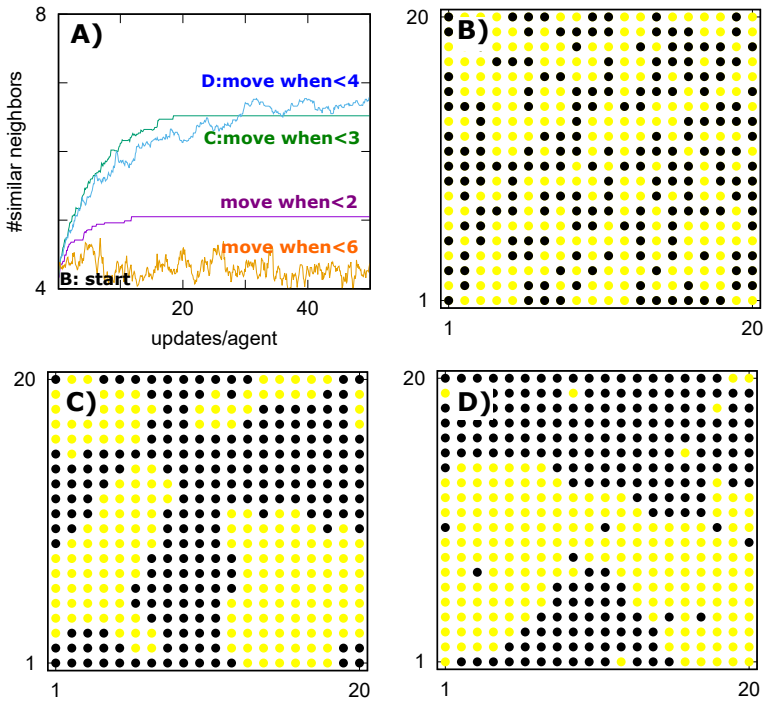


Figure 5.2: **Dynamics of the Schelling model.** Over time, segregation of agents is obtained. The two colors, "black" or "yellow" show different agents, with and equal numbers of each. We use an 8-neighborhood rule on a square lattice with periodic boundary conditions. A) Time development of system under different rules for when an agent moves, considering his the race of his 8 nearest neighbors. B) Initial random configuration of agents before any movements. C) An agent moves to a random other position if she has fewer than 3 neighbors of same color, i.e. 6 or more neighbors of opposing color than himself. This correspond to green curve in panel A. D) An agent moves if he has less than 4 neighbors of same color than herself (blue curve in panel A).

- Select one agent and let this agent move, if her number neighbors of the same color is smaller than a certain threshold. When moving, we just replace the agent's position with another random agent. This swapping takes place irrespective of, whether any of the agents gains or looses with this move, the move is only driven by current stress.

The central parameter is the threshold of equally-colored neighbors. A simulation of the model is shown in Fig. 5.2. From panel A) one observes that the system segregates. This segregation occurs to an extent that depends on the threshold used: if an agent is required to move already when she has fewer than three neighbors of same color as yourself (green curve in Fig. 5.2A)), the system evolves to a state where most agents on average have 6.5 neighbors of same color, that is, a strongly spatially-segregated system. This outcome is illustrated in panel C). In fact, if one also moves when there are three neighbors of same color, nearly the same degree of separation is obtained (Fig. 5.2D).

Thus, segregation can be obtained with a relatively moderate racial prefer-

ence, which is the main result of Schelling. Even if one only moved when one has 0 or 1 neighbors of same color, there is noticeably segregation. Noticeably, if the requirement is such that you are even pushed out with relatively few of opposing color, segregation is weakened, as people are pushed to move all the time, and boundaries tend to break up.

Noticeably, the above model is slightly simpler than the original model, where there were also empty spaces and people moved to these empty spaces. Our moves simply made two agents swap position, and not just one agent that moves into an empty space. Further, we do not ask whether the agent replaced gains with the move. If you move when you are frustrated because you only have 3 neighbors of the same color as yourself, then an opposing color will automatically be satisfied, as it then will have 8-3=5 neighbors with the color threshold satisfied. However, for a threshold that forces you to move also when you have 5 neighbours of your own color, our version of the model will force a differently colored neighbor into a frustrated situation. However, the result for threshold of six in Fig. 5.2A remains qualitatively similar.

Segregation models have been suggested to be important for a range of biological problems, ranging from evolution of toxin producing bacteria [79] to spatial sorting of different cell types into tissues, using the differential adhesion hypothesis [80, 81]. The differential adhesion hypothesis assumes that similar types of cells attract each other more strongly, than cells from different tissues.

5.1.2 Globalization in a nutshell

Globalization is the prominent characteristics of today's manufacturing and trade on intercontinental scales. This has not always been the case. Global trade emerged on a large scale when transportation costs decreased for more than a century, fueled by technology and inventions such as railroads [82] and containers [83]. This development opened for a competition between different geographical regions, with further optimization of manufacturing of simple products for global markets [**?**], allowing manufacturers to take advantage of *economies of scale* up to the global market.

Within the last two decades the invention and broad adoption of the internet made prices transparent to consumers beyond their regional scope and on a truly global scale. Local retailers often lost the resulting competition with other retailers far away, with few particularly competitive retailers suddenly finding themselves turning into quickly growing internet retailers — Resulting in a "winner-takes-all" type of dynamics. This dynamics adds the new facet of the pros and cons of agglomeration [84] and globalization [85, 86, 87, 88].

Following [89] we discuss a simple agent-based approach to illustrate the interplay between transportation costs and information barriers in a model describing manufacturing and trade. Thus each agent can produce products at a cost per product that is decreasing with his overall production, but on the other hand is limited in distribution by a transportation cost. This is implemented by allowing each agent to order a product where it is offered at

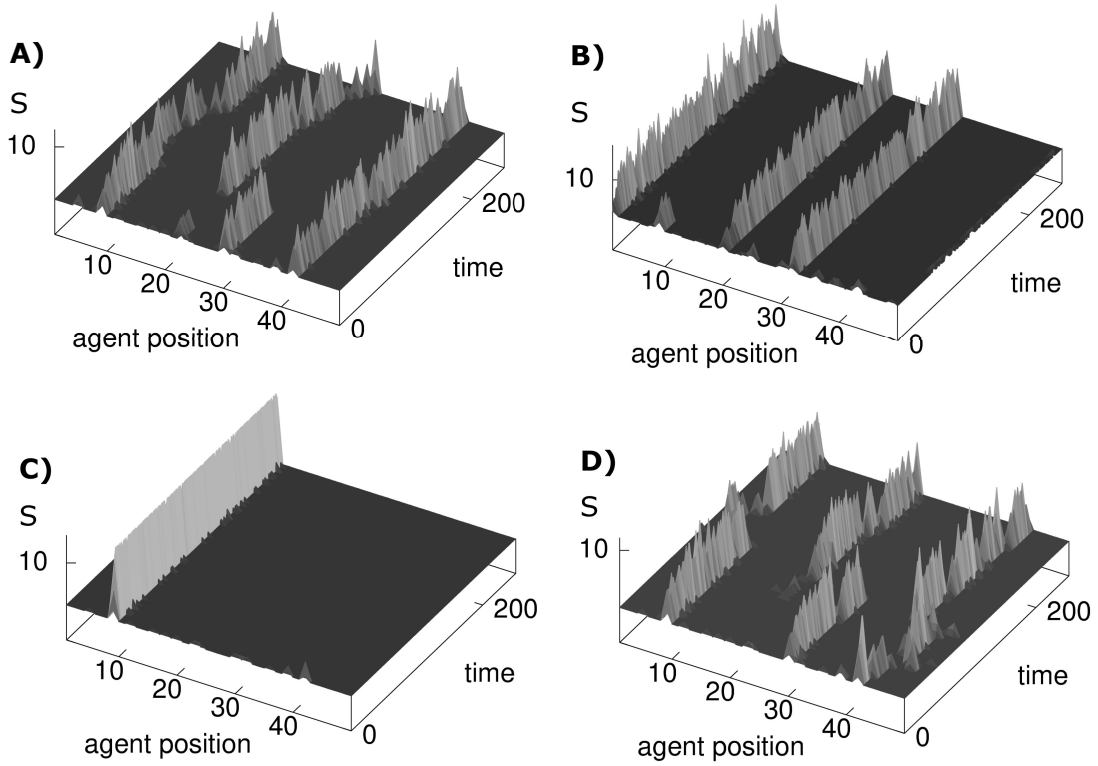


Figure 5.3: **Economies of scale.** Company location and size x as a function of time for $L = 50$ agents organized on a line and a production cost per unit that scales as $c \propto s^{-0.5}$. A) High noise ($\tau = 15$ transactions per unit time), high transport cost $\sigma = 0.05$ and information horizon $h = \infty$. B) Low noise $\tau = 25$, $\sigma = 0.05$ and $h = \infty$. C) $\tau = 15$, low transport cost ($\sigma = 0.01$) and $h = \infty$. D) As in C) but with a finite information horizon $h = 10$.

the lowest price, including the transportation cost to the producer.

Define a one-dimensional lattice with L agents at positions $x = 1, 2, 3, \ldots, L$. Each agent act as both producer and consumer, and we do not count the actual exchange as a reduction in the buyers capacity. Thus each exchange strengthen the seller, but have no direct consequence for the buyer.

In our model, agents for simplicity only act by giving points to each other, not loosing by doing so. At each update one random agent at position x rewards (makes one unit reward to) the agent at position y that can supply him with the good at lowest cost. This selected reward will determine a position y that as a consequence will accumulate a gain. The cost comes from production, transport, and possibly associated tariffs, which all are simply added together. Thus, for an agent at position x, the cost c of a product from position y is set to be

$$c = s(y)^{\gamma} + |x - y| \cdot \sigma. \tag{5.1}$$

In equation (5.1), s is the current company size at position y and s^γ with $\gamma \leq 0$ is the production cost per unit for that company — incorporating the notion, that larger companies are able to lower their production costs. A reward of one unit is subsequently assigned to that position y which provides the lowest cost for x:

$$reward(y) \to reward(y) + 1 \quad \text{for the position } y \text{ with minimal } c\,.$$

A value $\gamma \sim 0$ means that production cost does not diminish much with company size, whereas a lower γ would reflect an increasing effect for an *economies of scale*. $\gamma < -1$ is not realistic, as it would mean that the total production cost of many product units, namely $s \cdot s^\gamma = s^{1+\gamma}$, becomes cheaper than for just one unit. However, $\gamma \sim -1$ may be realistic for the software or movie industries, whereas traditional factory production may have moderately negative γ closer to 0.

The parameter σ quantifies a transportation cost that, for simplicity, is assumed to be linear in the distance between the position x of the consumer and the position y of the producer. Notice that this proportional dependence is markedly different from the exponential "iceberg cost" assumed in the economic literature [90]. In fact, one may even expect modern shipping costs to increase more slowly that proportional to distance, however, for simplicity, we keep the linear dependence here. Finally one may supplement the model with a tariff parameter β, quantifying that a customs barrier between position x and position y might be added to equation (5.1).

The model is executed in time steps, where each time unit consists of τ trading updates as defined above. During these τ steps the value of s does not change. After these τ updates, new company sizes $s(y)$ are assigned to be equal to previous ones plus the accumulated orders at site y during these updates:

$$s(y) = 1 + reward(y), \quad \text{and then reset } reward(y) = 0\,. \tag{5.2}$$

There is no direct memory to the earlier size of the company but, nevertheless, they tend to remain localized because of the sensitivity of accumulated rewards to production capacity at the previous production period.

Altogether, the model has three parameters γ, σ, and τ. In addition, tariffs may be added for externally imposed customs, and the model should in principle be extended to include pre-factors in front of s^γ, in order to take into account the variation in labor costs considered in model descendants of [86]. That is, products with small γ would presumably have a large cost of producing already the first product (as all subsequent products are nearly free). The system size L is irrelevant, as long as it is much larger than the domain scale set by the other parameters. γ and σ quantify incremental production cost and transportation cost for a unit of product, whereas τ is proportional to the time it takes to rebuild the production apparatus for the considered product type.

Fig. 5.3 explores the dynamics of the one-dimensional model with periodic boundary conditions (a ring of sites) using an intermediate level of *economies of scale* exponent $\gamma = -0.5$. The first three panels illustrate emergence of production centres (denoted "companies" in the following), reflecting the positive feedback between consumers and the *economies of scale*.

Fig. 5.3A illustrates that a given manufacturer may collapse while others emerge. Notice further that the emergence of new companies often occurs close to the positions of the previously collapsed ones. This inheritance reflects the memory associated to the geography of surrounding companies that survive the collapse of the one in question. In other words, when a company disappears, it leaves vacant a wide business niche because of the cost associated to distance for local customers to deal with companies farther away in the larger neighborhood.

Comparing Fig. 5.3A and B, one notices that lowering τ destabilizes companies. Remember, that low values of τ (as in Fig. 5.3A) correspond to the case where it only costs a few product units to build a new production facility. Therefore, a higher start-up cost will tend to stabilize existing production centers. Comparing Fig. 5.3C with Fig. 5.3A one sees that lower distribution costs may stabilize even a product with low τ.

Fig. 5.3D introduces another limit on avalability of products in terms of an "information horizon". With this the agents are only allowed to explore prices of the nearest say $h = 10$ neighbor agents in search for the lowest price (modeling a traditional "offline" economy). Fig. 5.3D uses the same other parameters as Fig. 5.3C and illustrates that a low information horizon has an effect comparable to a larger transportation cost (compare with Fig. 5.3A).

For relatively small transportation cost, the productions centers become large, whereas the noise in allocating customers in the time interval τ becomes small. In this limit an "equilibrium" production center should supply customers up to a distance x where the gain by *economies of scale* $\frac{d}{dx}s^\gamma \propto \frac{d}{dx}x^{D\gamma}$ balances the increase in transport cost σ by increasing x further. By differentiating the cost of products originating from a region with radius x in D dimensions, $cost(x) = x^{D\gamma} + \sigma x$:

$$\begin{aligned} \frac{d}{dx}cost(x) = 0 \quad &\Rightarrow \\ x^{D\gamma-1} \quad &\propto \quad \sigma \Rightarrow \\ x \quad &\propto \quad (1/\sigma)^{1/(1-\gamma D)} . \end{aligned} \tag{5.3}$$

Here, we only include the overall scaling, and not the tendency that small γ products tend to have larger costs for the first product (i.e. σ should be interpreted as the transportation cost per production cost for first product).

Overall, the size of a production center or its associated customer base is governed by the balance between the positive feedback of an *economies of scale* and the negative feedback set by transport. Within this analogy, the patterns in Fig. 5.3 are reminiscent of the ones found in reaction diffusion systems [91] where a local positive feedback is combined with a spatially extended

inhibition. Turing suggested these reactions as a way to understand pattern formation in the development of an embryo.

5.1.3 Information spreading on social scales

Information may spread like infection waves: Information gathering is also important on larger scales, allowing individual organisms to optimize their survival and proliferation. For social animals, information is collected by communication, involving some sort of language. Interestingly, language and communication within our own species also seems to follow rules that result in wave-like behavior.

It has long been observed that linguistic features, like some diseases, spread outward from an originating center. A beautiful example is the geographical distribution of the word 'snail' in Japan. This was investigated by *Yanagita* [92], where it was found that ancient forms of the word still existed in the southern and northern parts of the country but not in the middle. He concluded, using his wave theory, that this reflected the strong influence of Kyoto, Japan's former capital.

Following [93], consider the dynamics of culture spreading around a strong pulsating culture center. As a proxy for the spreading of cultural traits one may use the spreading of words where the key feature is that new words are more prone to be adopted than older ones:

- Information is sorted, using that **New** is better than **Old**.

As a good model system we consider word spreading in Japan, which presents a long history where a single strong center teeming with ideas that subsequently spread over the country.

Figure 5.4 shows the geographical distribution of swear words across Japan. There are about 20 words present in the map, and the overall trend is that old words are found far away from Kyoto whereas new words are found close to it. The circles drawn indicate the swear words' centre of mass distributions with respect to the absolute distance from Kyoto. The data also shows that the gap between to adjacent circles are not uniform but tend to grow with increasing distance away from Kyoto. From old records of when each word first appeared in Kyoto, the speed of swear word propagation has been estimated to be $v_{\text{word}} = 1$ km/year (0.5-2 km/year). Accordingly, the words in the northern and southern parts of Japan are found to be about 500 years old.

The ABM for "word" spreading is defined on a two-dimensional square lattice on which words, after being coined in the culture center, spread. N agents are then placed on the lattice. The model aims to capture the ongoing spreading of new words that originate stochastically at the center with a frequency f_{word} from the centre:

- At each update a new word is initiated at the domain center with probability f_{word}/N, and assigned a birth time according to a time counter.

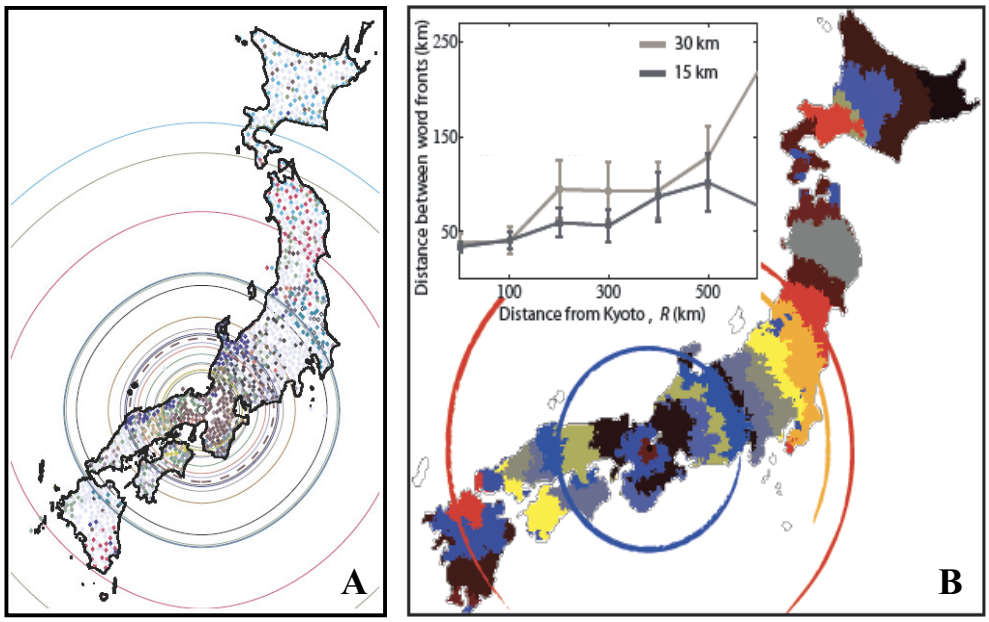


Figure 5.4: **Swear word dynamics in Japan.** The left panel shows the distribution of swear words as measured in the 1980's. The geographical distribution of concentric circles around Kyoto is the result of 600 years of history. The right panel shows a snapshot of a simulation of the spatial dynamics of word spreading over the Japanese mainland. Blue and red circles show two examples where the same word form is found symmetrically on either side of Kyoto. The graph in the upper left corner shows the mean distance between two adjacent fronts (averaged over many runs) as a function of distance from Kyoto. The orange broken circle belongs to a word which only is present at Kyoto's east side. The probability that a word coexists on both sides decays with distance away from Kyoto. In the insert of Panel B on see this in terms of the width of the respective word regions: As distance from Tokyo increases, the region of surviving words tends to increase. Figure reproduced from ref. [93].

- At each update a lattice site is chosen, and its word is communicated at equal probability to one of the four neighboring sites.

- If the word is younger than what is already present on the site chosen, the word overwrites the order word at this site. If the word is transmitted to a site where an even newer version exists, the older word is ignored. If a word is transmitted to a site where it already is, then in effect the system is not changed.

As words spread, they always retain their original birth time, assigned at origination at the center. In Fig. 5.4 the 2-d lattice is constrained within the land boarders of Japan, thus allowing us to include the simplest geographic features.

The simulations depends on the frequency of new words f_{word} originating

from Kyoto and also on the size of the squares into which we coarse grain space. With larger but fewer patches, fluctuations increase and the likelihood that a word dies out becomes larger. The figure 5.4 shows the case where each "agent" represents a lattice withe square size of $\Delta = 30$ km, and where frequency of new words was calibrated such that about 20 words remain simultaneously on main island, as can be counted from the data in panel A).

The size of Δ was adjusted to fit the increasing distances between words as one moves out from the center, see Fig. 5.4. Importantly, the words are quite different from center to periphery of Japan.

The language spreading model with the basic assumption that "'new" overrules "old" resembles a minimal disease spreading model, where people get infected and subsequently immune to each disease. In this process, subsequent waves of emerging new diseases become possible [94, 95].

Questions:

5.1) Simulate a Schelling-like model in two-dimensions for a 40×40 site system where each site has 8 neighbors, and only one color is allowed to move away from present location. Set threshold for moving when having less than 3 neighbors of same color as yourself. Simulate this system for a 3 color system, only allowing one color to make active moves. Simulate the system for long times to verify coarsening. Use periodic boundary conditions.
Qlesson: One could get segregation driven by only one race. But it should be possible to distinguish in a 3 race system

5.2) Simulate the globalization model in one dimension akin Fig. 5.3A. Plot the average size of companies as a function of the transportation cost and time-scale τ for a fixed value of economy of scale $\gamma = -0.5$. Simulate an $N = 100$ system placed on a line, for at least 2000 updates per agent in the system.
Qlesson: The distribution of company sizes is quite independent of τ.

5.3) Simulate and visualize spreading of signals along a one-dimensional line, with new words appearing at position $x = n/2$ with high frequency (for example each time each agent has been involved in one word exchange). At each step, select two neighbors, and let the youngest word spread to replace the oldest word. Also simulate the model when new words are inserted more rarely. Qlesson: With fast word innovation one will see words on the right and left sides of system rarely being the same because the survival on the two sides are exposed to big fluctuations around the insertion point.

5.2 Information Battles

5.2.1 Hub dominance or Social Fragmentation

Hierarchy formation is important for stability and function of colonies of many social animals. For ants, hierarchical organization allows for prioritization of resources and work. The hierarchy in human society may in addition be coupled to the flow of information between its members.

Mini Tutorial: Suggest positive feedback mechanisms that could favor maintenance of high connectivity/central hubs in social networks.

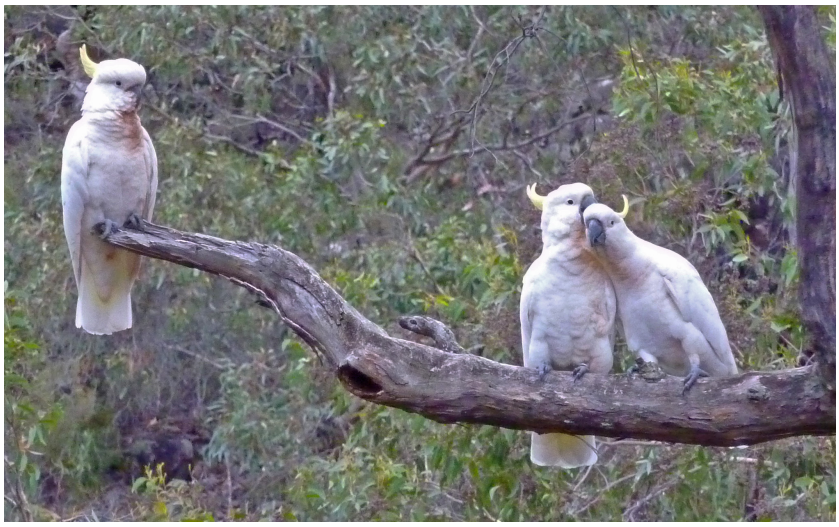


Figure 5.5: **Preferences in communication.** Segregation and social structure may emerge as consequence of preferences in communication, here visualized by some that communicate and some that do not.

Simple agent based modeling of spontaneous hierarchy formation typically emulates a positive feedback between winning and future chance of winning [96, 97]. Interestingly, social hierarchies are also predicted by an ABM acting through information gathering and networking [98, 99]. We will see that a hierarchy around a central node is a natural consequence of information gathering of individual agents, because of a positive feedback between being centrally placed and having access to newer information.

The ABM of [99] considers dynamics of a social network consisting of N agents that is connected with a fixed number of L links, see Fig. 5.6. Each agent is assigned an individual memory that includes three vectors:
(1a) N pointers to each of the $N-1$ other agents in the system that show which friend [1)] provided information about the other agent,
(1b) the age of information pinpointing the direction to each of the $N-1$ other agents [2)]
1c) a priority vector that allocates space proportional to the interest in each of the other agents in the system.

[1)] A friend is an agent which the agent in question has communicated with. An agent uses friends in its contemporary map of pointers. Each of these pointers is directed toward the friend that provided him with the newest information about any particular person in the system. The number of pointers equal the number N-1 of persons in the network, and many pointers may go to the same friend.

[2)] The information age that is used to compare quality and subsequently update the information when communicating with neighboring agents.

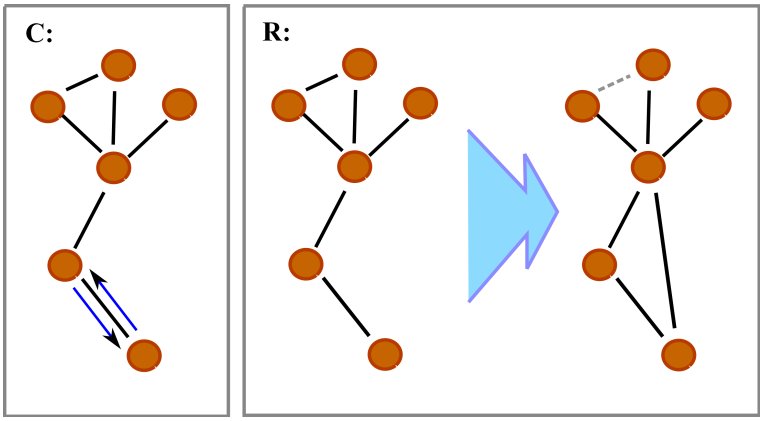


Figure 5.6: **An agent-based model for social networking.** The model consists of a communication step (C) and a rewiring step (R). During communication agents share information with nearest neighbors on the network, allowing new information about the current position of each agent to spread in a process mimicking the swear word spreading across Japan. During the rewiring a new link is formed to a friend of a friend, mimicking social climbing towards the source of new information [98, 99].

The memory in (1a) and (1b) is a map that is updated based on the principle of new is better than old information, whereas the priority in (1c) is only used in the more complicated version that give segregation.

The 3 vectors above are updated with each encounter. Thus people just talk about people. In the simplest model people only update their knowledge about overall direction towards other people, and how new their knowledge are.

The network model is executed in time steps, each consisting of one of the two events (see Fig. 5.6):

- **Communication (C):** Choose a random link and let the two agents connected by the link communicate about a third agent selected by one of them. The two agents also update their information about each other. [3)]

- **Rewiring (R):** In a process termed "social climbing," let a random agent use the local information to form a link to a friend's friend to shorten its distance to a selected third agent. Subsequently a random agent loses one of its links.

[3)] When two agents communicate about a third agent, they first decide which of them has the newer information about network location of the third agent in question. That newest information is considered the most reliable. The agent with the older information then updates its information using the agent with the newer information. Thereby, each agent builds a map of the contemporary social network. This information will be reliable if communication is much more frequent than rewiring [98].

The communication step C is executed much more frequently than the rewiring step R, thereby allowing each agent to build a reliable contemporary map of where other agents are. R defines a network dynamics where individuals change their neighborhood by gradual social climbing to friends of friends [4)].

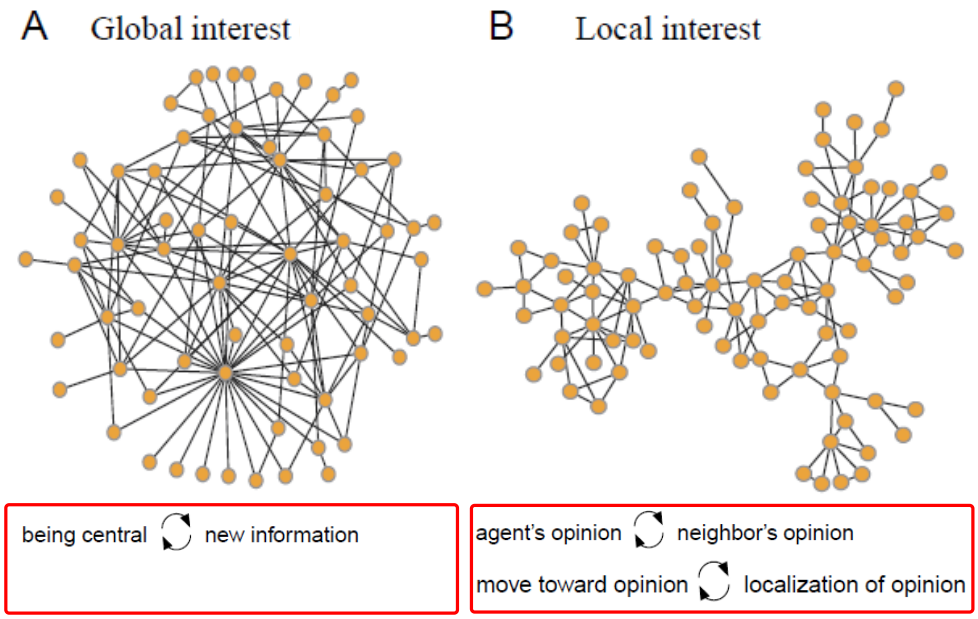


Figure 5.7: **Network evolution through information hunt.** ABM where agents create links to neighbors of neighbors, and remove random links. A) A network snapshot in the steady state of an evolving social network where everybody attempts to climb towards each other in an ongoing hunt for information about each other. For details see ref. [98]. B) Snapshot of a segregated network, where the social climbing of agents is limited by the agents prioritization of recent experiences, leading to self-organization of a limited information horizon [99]. Lower panels emphasize the main positive feedback that drives the topology of the communication networks shown.

The ABM is first simulated such that agents prioritize each other equally all the time. This mimics the case where "broadly minded" agents simply hunt for new information about everybody. The network then develops the hierarchical structure shown in Fig. 5.7 A). This hierarchical structure reflects the positive feedback between being central and having access to new information about other agents. That is, agents with new information are attractive in the R-move of other agents. As a consequence, central agents tend to gain links and networking reinforce a hierarchy based on information access.

The behavior of the ABM changes dramatically, when agents prioritize their "information hunt" based on what they are used to hear, see Fig. 5.7B).

[4)] In case an agent is completely disconnected from the system, it reconnects on the basis of a weighted choice from its own prioritized list.

Following Spencer's conjecture of proportionality between interest and previous experience [100], this version of the ABM lets an agent's priority memory be filled with other agents' names proportional to their occurrence in recent gossiping. In practice, the priority memory is updated during the communication step, where one of the communicating agents uses her prioritized memory to select the agent of interest to "talk" about. Subsequently both agents increase the memory slot about this discussed agent at the cost of reducing some other memory.

The two versions of the ABM for dynamical social networks illustrates: 1) that agent based models indeed may reflect common experiences from social networking; 2) that information spreading or containment may act in a close positive feedback loop with contemporary social organization; 3) that social segregation may emerge also among identical agents due to a positive feedback involving re-enforced interests in local neighborhoods.

5.2.2 Emergence and Decline of Wrong Paradigms

Human history contains a number of epocs, each dominated by certain themes. Themes are often centered around scientific ideas, which each become so dominating on large scales that nearly everybody is affected and influenced by their prevalence in the ongoing process of human communication/thinking (Kuhn, Thomas (1962). The Structure of Scientific Revolutions). In science these themes are often centered around single words or concepts, with recent examples including climate change, chaos, or string theory. These phenomena have a real basis — but also include a large social factor associated to people communicating and reinforcing each other.

Typical features of paradigm shifts are the relatively sharp initiation, rapid growth up to a near-global awareness level, which is followed by a slow decline where the ability to sustain interest is weakened by new ideas. Sometimes, scientific concepts even escape the scientific community to the global public, and become common themes that influence the frame for future cultural development.

In social systems it could be believed that opinion formation is governed by cooperative effects in the sense that two persons together have a much stronger convincing ability than one person alone.

Instead of opinions, we here consider the spread of ideas, or concepts, or more generally: orientations, that are open-ended in the sense that there is an infinity of varieties. Furthermore, we consider these ideas to have a small probability α for being initiated. In fact, we assume that each new idea appears spontaneously only once. Finally, and most importantly, we assume that each agent can only hold any particular idea during *maximally one* continuous time period. When changing to a new idea, the agent never returns to any of the ideas that she had at earlier times. **This amounts to assuming that all the ideas we consider are false, that is, that when an agent is convinced**

that the idea is false, she remains convinced of this.

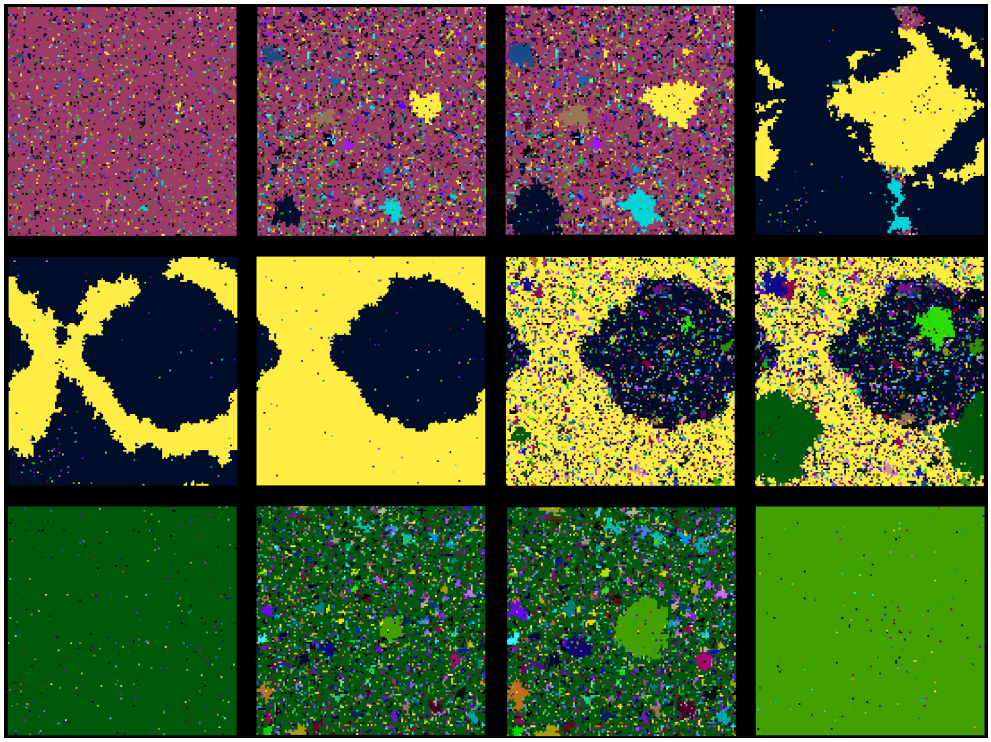


Figure 5.8: **Model for idea spreading.** 12 consecutive snapshots of a $N = 128 \times 128$ system with $\alpha = 25 \times 10^{-6}$. The time intervals between the pictures are not equidistant, as can be seen from the times given for each panel (times correspond to Fig. 2b). Time is measured in units of full sweeps, that is, one update for each agent.

The model is defined on a 2-d square lattice of $L \times L$ sites, each occupied by an agent. Each agent i can be assigned a number r_i which can take any integer value. This number plays the role of a particular idea, concept, or opinion. At any time-step one random agent i is selected, and the following two actions are attempted:

- One of the nearest neighbors j to the agent i is selected. Denoting by n_j the total number of agents with integer value equal to that of j, we with probability n_j/N let the agent i change its integer value to that of its neighbor j, provided that i never assumed that particular integer value before. In case it had, no update is made.

- With probability α another random agent k is assigned a new random integer which does not appear anywhere else in the system. Thus α represents the "innovation" rate.

A key difference to previous models of opinion spreading is the rule that old opinions are never repeated. Practically, one could of course repeat a particular integer in the simulation, provided that it does not exist anywhere in the

system. This is because an integer that is not on the lattice would not be distinguished from a new number by the model. Another feature of the above model is the factor n_j/N, which implies that a minority concept has more difficulty in spreading than a more widespread idea. This particular feature represents cooperative effects in social systems, and is nearly the same as just selecting two agents to then influence another. This feature is included in a way that we 1) allow cooperativity to act on long distances but at the same time restrict propagation to spreading on a 2-d plane, 2) avoid discussion of detailed neighborhood updates related to where the two agents are located, and 3) allow a single idea to nucleate from one person (with probability $1/N$). Mini Tutorial: What would happen if there was no memory in the above model, that is, that every idea could spread proportional to the number of current followers, irrespective of history?

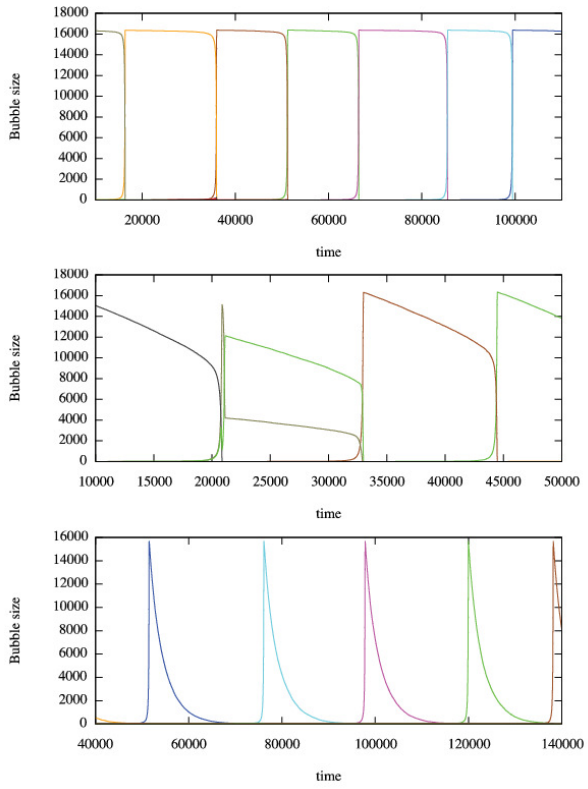


Figure 5.9: **Three time series of the sizes of the dominant states of the system.** At $\alpha = 0.4 \times 10^{-6}$, $\alpha = 25 \times 10^{-6}$, and $\alpha = 400 \times 10^{-6}$, respectively. Time is measured in units of sweeps (updates per agent). Notice that the length of each period does not change substantially with α, and in fact becomes more regular with larger α.

The collective effects associated to the cooperative coupling lead to globally coherent states, that sometimes are replaced by new coherent states through a

system sweeping "avalanche dynamics" with a deterministic part governed by

$$\frac{dn}{dt} = \left(\frac{n}{N}\right)^2 \propto n^2 \tag{5.4}$$

$$n(t) = \frac{1}{t_c - t} \tag{5.5}$$

that is, it is divergent (reaching system size) at some finite time t_c. Thus, the start of new paradigms is slow, but the final rise is fast.

Fig. 5.8 shows 12 subsequent states of a system driven by the model. The snapshots reflect states of the simulation shown in Fig. 5.9b, starting at time $t = 62000$. The first panel shows the system shortly after a new idea swept the system, leaving the system in a coherent state dominated by this particular idea. A few agents have different colors (i.e., ideas), representing the effect of a finite innovation rate α over the short time interval after the dominating idea took over. The second and third panels show the system closer to the next transition (at $t \sim 68,000$), where several ideas have nucleated some sizeable clusters of coherent colors. Panels 4 and 5 correspond to the spike at $t \sim 68000$ which subsequently leaves the system with two mutually coexisting coherent states that persist until they are erased by a new "avalanche" (panels 8 and 9). Finally panels 9-12 describe the evolution of the system from $t = 82,000$ to $t = 93,000$. This period is characterized by the dominance, erosion and subsequent replacement of one state with another.

Figure 5.9 shows three time series for the rise and fall of different leading communities, illustrating the behavior at low, intermediate, and high values of the "innovation" rate α. In all cases one sees a sharp growth of the dominating community, followed by a slower decline. Remarkably, the lengths of domination periods are quite insensitive to α. However, as seen from Fig. 5.9 a-c, the nature of the decline of the dominating state depends on α:

- For low α, the dominating state nearly remains intact until it is replaced by a rare single nucleation event that suddenly replaces the old state with a new one. As a consequence, low noise only rarely leads to situations where more than one state nucleate at the same time.

- At intermediate α the decline is substantial and many nucleating states are competing. Sometimes, two nucleating states grow and interfere which subsequently results in a period where there are two frozen states. This reflects events where one of the major communities was defeated in some part of the system by another, and therefore cannot re-invade that region again. Thereby a substantial minority community can remain protected by its immunity to the prevailing majority.

- Large α results in a complete erosion of the dominating state, before a new nucleating state can grow. This growth will be in an environment where it also has to compete with ongoing erosion from other nucleating states. Because the winner is the result of many events, the distribution

of time intervals between global state changes becomes more regular than for lower noise. At even higher α, the on-going activity prevents nucleation, thereby leaving the system in a permanently noisy state with multiple small domains that are constantly generated and replaced.

The model presented here is in a class of opinion formation models studied in statistical physics and complex systems. Common the these models is exchange of opinion and alignment of opinions. The peculiarity of the model described is the infinite opinion space and most importantly the repression of previously rejected opinions. In case one removes this constraint and allows old ideas to re-invade the same site again, the nucleation process will be rare and the winner will persistently dominate. Without immunization, new ideas are typically removed shortly after introduction by the cooperative re-invasion of the dominating idea. The system has no memory of all these small "noise events", in contrast to the memory that is inherent in human inventions.

Mini Tutorial: Remove the requirement of two persons agreeing before anyone can be convinced. What do you think would happen with the behavior then?

The model provides a new frame looking at the interplay between dominance of prevailing concepts supported by a large number of followers, and the striking inability of these concepts to defend themselves against new ideas when the situation is prone to takeover. The increased vulnerability of a dominating idea or paradigm with age is in our model seen in the steady increase in the number of competing ideas, and a parallel decrease in its support. For intermediate or large innovation rates, the takeover is a chaotic process with multiple new states competing on short time scale. The final takeover is on a much shorter time scale than the decline. Existing paradigms are eroded in a pre-paradigm phase for the next paradigm much as envisioned by Kuhn (T. S. Kuhn, "The Structure of Scientific Revolutions", 1st. ed., Chicago: Univ. of Chicago Pr., 1962.) New paradigms are born fast, ideally aggregating in a real scientific competition between the many random ideas that emerged in the pre-paradigm phase.

Questions:

5.4) Consider the paradigm model, with cooperative idea spreading. When you are the first to get an idea in a system on $N = L \times L$, what is the average time until the idea is spread to two persons? And given n persons have the idea, what is the time it spreads to one more person?. Simulate a 10×10 system where each agent can get an idea with probability 0.01 at each time one agent try to transmit a message to a neighbor. Plot popularity (number of followers) of some ideas in the simulation. Notice that each agent needs a memory of the last 100 ideas it was exposed to, and is not allowed to take any ideas in this list.

Qlesson: Its by far hardest to spread the idea the first time.

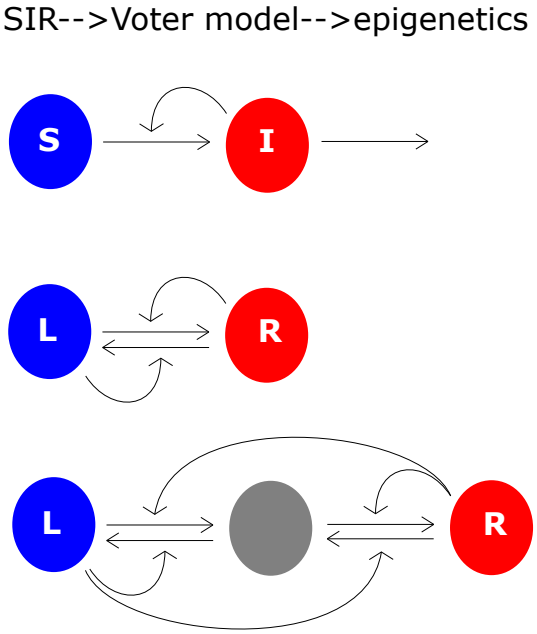


Figure 5.10: **Schematics showing the SIR, Voter, and epigenetics models.** The three models in this chapter, building on interacting populations. Straight arrows show possible transitions, curves arrows show facilitating interactions. The upper panel shows the SIR model, the middle panel the Voter model and the lower panel the two step Voter model with bi-stability.

5.3 Mass Action Kinetics and Epidemics

Interactions between individuals are often thought to be proportional to their densities:

$$Frequency\ for\ an\ encounter\ between\ A\ and\ B \propto \rho_A \times \rho_B \tag{5.6}$$

where ρ_A and ρ_B are the densities of A, respectively B. This simple approach of course assumes that there is no real spatial effect, or that events A and B are not confined to, for example, moving around a network that eventually could be depleted by such encounters.

There are two classical real-world examples which use this "well-mixed" approximation: Epidemics and Ecology. In epidemics, the approximation is found in the SIR/SIRS model from the 1930s and in ecology in the Lotka-Voltera equation that was developed a few decades earlier. We will here go through one of those examples, and use this example to introduce stochastic models in well mixed populations.

According to the World Health Organization (WHO), infectious diseases causes about 25% of human death worldwide, associated to 1,415 known species of infectious organisms [101]. This includes a wide variety of pathogens, of which a majority can also be transmitted between one or more animal species and us. Here we will briefly mention the simplest model for describing the spreading of one idealized disease in a homogeneous population of "agents". More complicated models should take into account limited immunity, or immunity of limited duration, and also the fact that about 60% of known human diseases also have non-human hosts [102, 103].

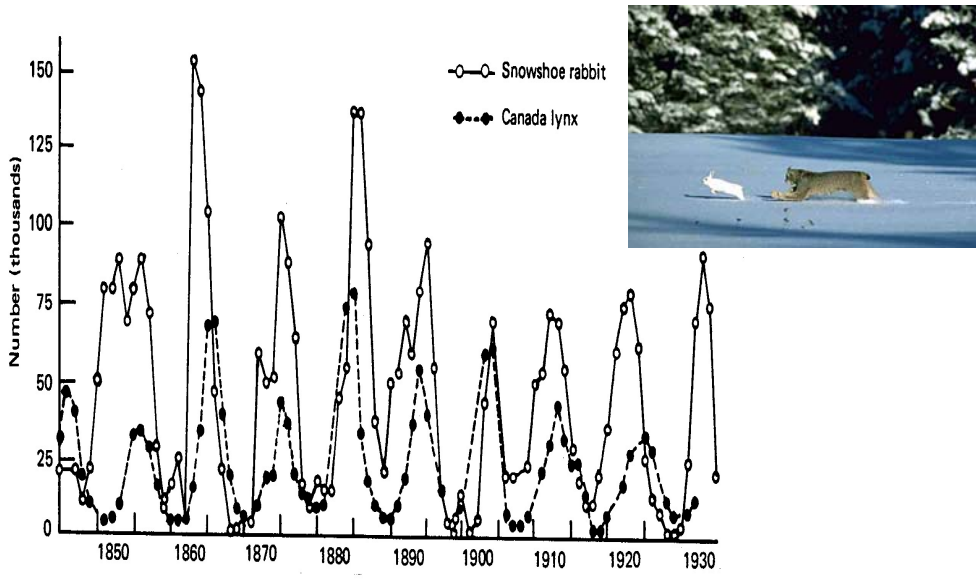


Figure 5.11: **Interacting species in real life.** It often appears that a lynx population grows when a rabbit population is large, suggesting the famous Lotka-Volterra coupling between their populations: $dx/dt = a \cdot x - \eta \cdot x \cdot y$ and $dy/dt = \beta \cdot x \cdot y - \delta \cdot y$ where x is the prey and y the predator population (and the parameter β/η the amount of predator that is produced for each prey that is eaten).

Epidemic models are very similar to models for spreading information and ideas. The classical epidemic model is the SIR (Susceptible-Infectious-Recovered) model, that divides the population into fraction of susceptible individuals (S), infectious individuals (I) and recovered individuals (R). The latter individuals may be dead or immunized and thereby are removed from further spreading of the disease.

The earliest mathematical treatment of disease infection for malaria was done by Ross, [**?**], who also introduced mosquito nets to reduce malaria. The simplified mass-action kinetics for a well-mixed population that include the recovery/removed state can be found in (Kermack, W. O. and McKendrick, A. G. "A Contribution to the Mathematical Theory of Epidemics." Proc. Roy. Soc. Lond. A 115, 700-721, 1927):

$$\frac{dS}{dt} = -\lambda \cdot S \cdot I \tag{5.7}$$

$$\frac{dI}{dt} = \lambda \cdot S \cdot I - \gamma I \tag{5.8}$$

$$\frac{dR}{dt} = \gamma \cdot I \tag{5.9}$$

where the parameter γ decides for how long individuals are infectious (time $\sim 1/\gamma$) and the parameter λ/γ subsequently determines how many individuals each infected person can infect. A central parameter is the so called R_0-factor:

$$R_0 = \frac{\lambda}{\gamma}\,, \tag{5.10}$$

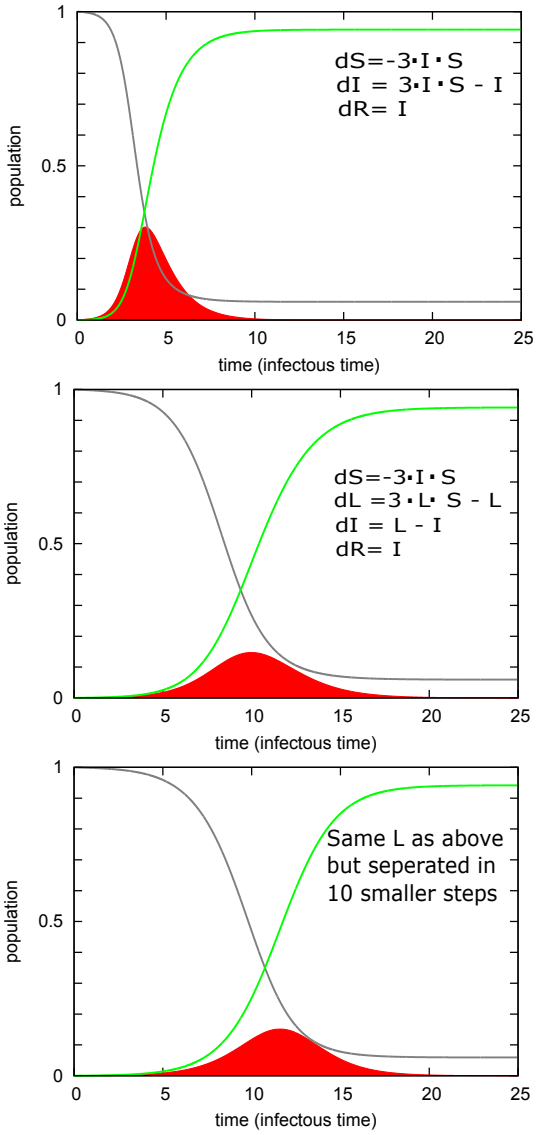


Figure 5.12: **Simulating epidemics.** The classical SIR model and a variant where one includes a latency time where people are infected but cannot infect others.

which is central in thinking about how widespread the disease becomes before herd immunity sets in.. R_0 is the number of infections that each infected individual cause at the beginning of the disease (when few individuals have been infected).

As the disease spreads, the amplification number tends to decrease because people get immunized (or die), leaving fewer susceptible individuals. R_0 is very large for measles ($\sim 10 \to 15$), it is about two for ebola, about 1.3 for common influenza. For the Covid-19 it is estimated to be between 2 and 3.

Dividing the first with the last equation one obtain:

$$\frac{dln(S)}{dR} = -\lambda/\gamma = -R_0 \Rightarrow S(\infty) = e^{-R_0(1-S(\infty))} \tag{5.11}$$

where we use that $S + I + R = 1$, $S(0) = 1$ and that $R(\infty) = 1 - S(\infty)$ since there are no infected individuals after the epidemics have died out. For an illustration of the solution see Fig. **??**. Thus, if an epidemic with $R_0 >> 1$ really would follow the SIR model with, $S(\infty) \ll 1$ and the number of "survivors", that is, those never infected, declines with R_0 as

$$S(\infty) \sim e^{-R_0} .$$

This is a very small number, and much smaller than the so called herd immunity limit.

Herd immunity is instead calculated from the size of S when the disease stop growing exponentially, i.e. when $dI/dt = 0 \rightarrow S = 1/R_0$ which indeed is much larger than e^{-R_0}. The herd immunity is the level at which the epidemic stop if one avoid "overshooting", i.e. avoid that the many infected continues infect after S have decreased to $1/R_0$. Herd immunity is also what one wants to obtain with vaccination strategies, since vaccinating a fraction $> 1 - 1/R_0$ of the population would secure that $dI/dt < 0$, and thus that an epidemic could not propagate.

If $R_0 < 1$ then the disease cannot spread, corresponding to a percolation that is limited to a finite cluster. In that case $S(\infty) = 1$ is the only solution to 5.11. When $R_0 > 1$, on the other hand, the disease indeed spreads. If a fraction q is immune to the disease, then the effective $S \rightarrow S \cdot (1 - q)$ and the real spreading will occur with an effective $R = R_0 \cdot (1 - q)$ which becomes smaller than 1 when

$$R_0 \cdot (1 - q) < 1 \rightarrow q > 1 - \frac{1}{R_0} . \tag{5.12}$$

Thus, for an R factor of 10 one needs to vaccinate more than 90% of the population.

Recovered individuals can not be re-infected in the SIR model. Further, the model assume that one become infectious imidiately after infection. Figure 5.13 also explore the effect of relaxing these conditions, thus having having a latency period between infection and being infectious and becomming susceptible again after some longer time interval. The latency period effectively delays the progress of the disease, but does not change the long term fraction of infected people needed to obtain herd immunity. In Fig. 5.14 we show data for recurrent epidemics of some disease (influenza).

5.4 Agent perspective on Covid-19

Considerable evidence indicates that superspreaders are important in the spread of COVID-19 (See Sneppen, Taylor & Simonsen, medRxiv 2020, `https://doi.org/10.1101/2020.05.17.20104745`). Examples of superspreading events include an outbreak in South Korea in which a single infected person who attended five night clubs in one night caused at least 50 new infections, and a 2.5 hour choir rehearsal in Skagit, Washington where 52 out of 61 attendees were infected. Moreover, multiple studies of COVID-19 have quantitatively assessed the heterogeneity of infectivity among infected individuals, finding that 1% to 20% of infected people cause about 80% of new infections.

Given the observed evidence that superspreaders/superspreading events are important in COVID-19 transmission, models should not rely on a single parameter such as the basic reproductive number (R_0), because doing so obscures the considerable impact of individual variation in infectivity on an epidemic's trajectory. See

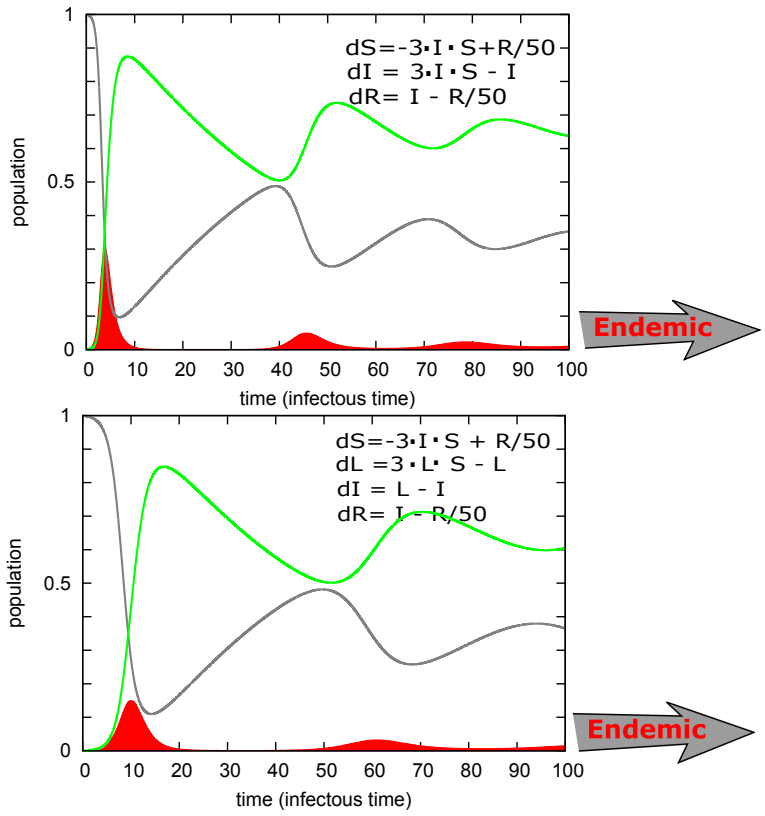


Figure 5.13: **Model including reinfections.** The SIRS model, where recovered people can get reinfected after a longer time interval. This model is relevant, if the immunity obtained has a time limit. Alternately, steady state can be obtained if new individuals are born without immunity. The simulations are started with a fraction of 0.001 individuals infected and the remaining $S = 0.999$ susceptible.

Fig. 5.15. Agent based models, however, are very well-suited to investigate the role of superspreaders. Like standard compartmental SEIR models, they can easily reproduce the epidemic curves observed in a population. Unlike purely compartmental models, however, agent-based models can adjust individual infectivity and mimic repeated social interactions within defined groups. In an agent-based model an agent goes to the same workplace in the morning and home to the same household at night. In contrast, inhabitants of standard compartmental models go to a new workplace and home to a new family in every time step.

Model: *We developed agent-based model with three simulated sectors of social contact through which the disease can be transmitted (Fig. 5.16). Each agent was assigned to one "home" and one "work" unit and participated in random "other" contacts. Each home had an average of 2.1 members. Agents 20-70 were assigned a "workplace," a Poisson distributed cluster of average size 6 agents; to simulate interactions between workplaces, each agent's connections were assigned to two random persons outside this cluster. "Other" contacts were chosen at random from the*

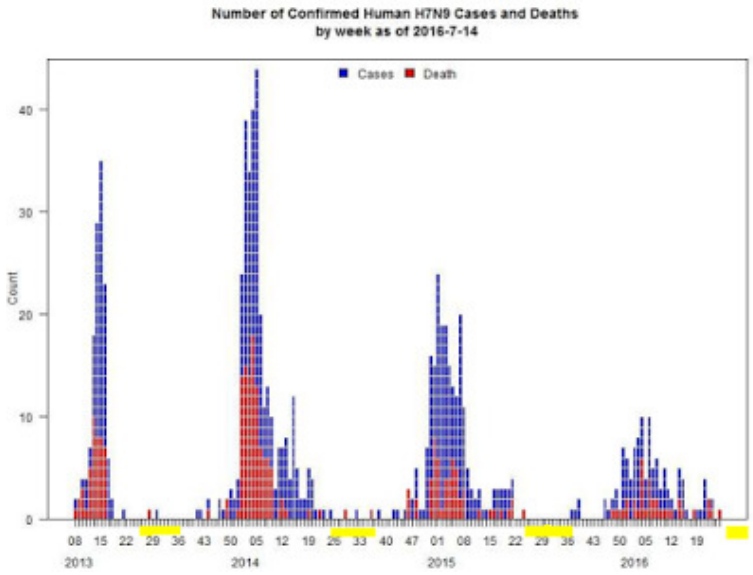


Figure 5.14: **Data showing recurrent epidemics.**

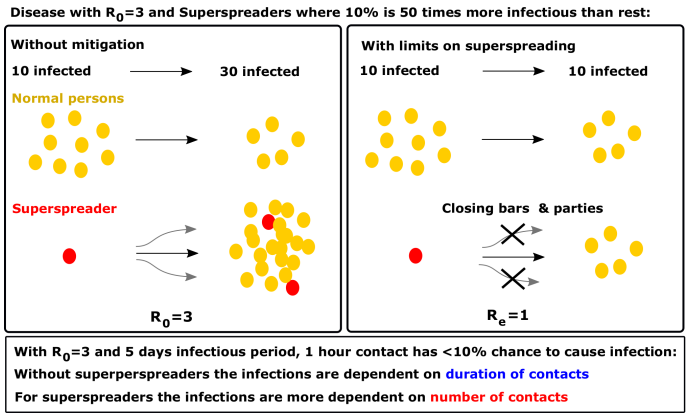


Figure 5.15: **Superspreaders as persons with larger virus shedding:** Left panel show normal epidemic, right panel show how distribution of infections would be if 10% infected 50 times more than rest. Notice bottom philosophy: Superspreaders can infect people during brief encounters, whereas a normal person only have substantial risk for spreading the disease using long exposure.

entire population. Progression of disease was modeled in a Susceptible, Exposed, Infected, Recovered (SEIR) framework, with agents passing through each stage according to preset rules (Figure). The exposed period was set to 5 days, extending from infection to symptom onset. Agents became infectious 2.5 days after infection and remain so through day 3 after symptom onset. All transitions between stages were implemented as a corresponding probability per time to pass to next stage. Age-dependent conditional probabilities governed progression from symptomatic illness to hospitalization and intensive care that was calibrated to a death rate of 0.3%.

Most agents were assigned an infection activity parameter (s_i) of 1, indicating

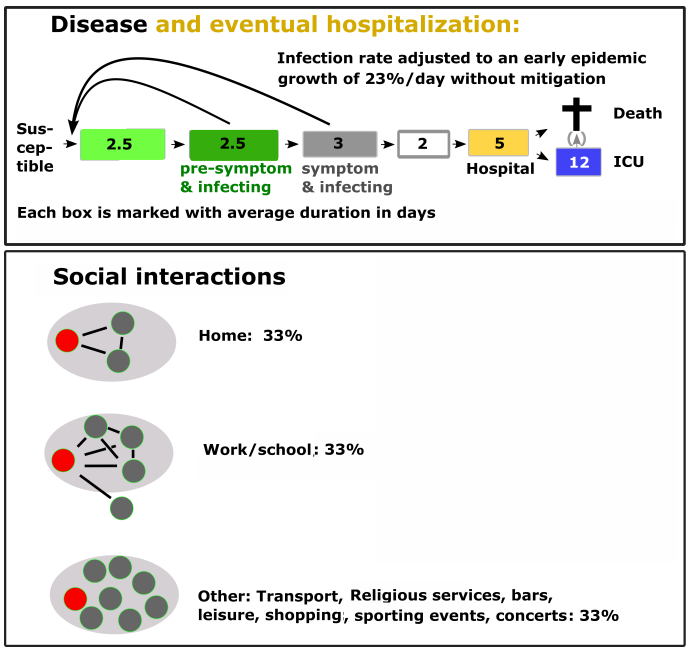


Figure 5.16: **Agent based model for Covid-19 epidemic:** Top panel show progress of disease in each infected person. Bottom panel show the social structure with each persons contacts divided into tree different social circles.

that the agent has one chance of transmitting the virus at a given contact. A chosen proportion of agents were designated as superspreaders, with $s_i = 50$*. Simulations were run in a population of 1 million seeded with 100 infected agents. In defined time steps* Δt *within an agent's infectious period, each infected agent was chosen for a contact with an age-dependent probability. For each chosen agent we assigned a contact in one of the three sectors: home, work/school, other. The were selected with probabilities such that they occur in a ratio of 1:1:1 across the population. Contacts were selected so that 1/3 occurred in each of our three social sectors, resembling social science data from Mossong et al. (2008).*

We first simulated epidemic trajectories in our socially structured model both without and with superspreaders. We initially modelled superspreaders were by designating 10% of the population as having 50 times greater infectivity than the other 90%. We adjusted the infectivity to produce a growth rate of 23%/day, the rate observed in multiple settings in the early stages before mitigation was implemented(17). We then applied two simulated mitigation strategies in turn, removing first "work" and then "other" contacts.

When we allowed contacts of all types (i.e., no mitigation), the epidemic trajectories were virtually identical whether superspreaders were present or not (Fig. 5.18 a, d). When we eliminated "work/school" contacts, the epidemic curves with and without superspreaders were still similar: in both cases the epidemics had been broadened and flattened somewhat, with somewhat lower peaks in cases and in ICU demand (Fig. 5.18 b, e).

When we included superspreaders in the model, however, the benefits of prevent-

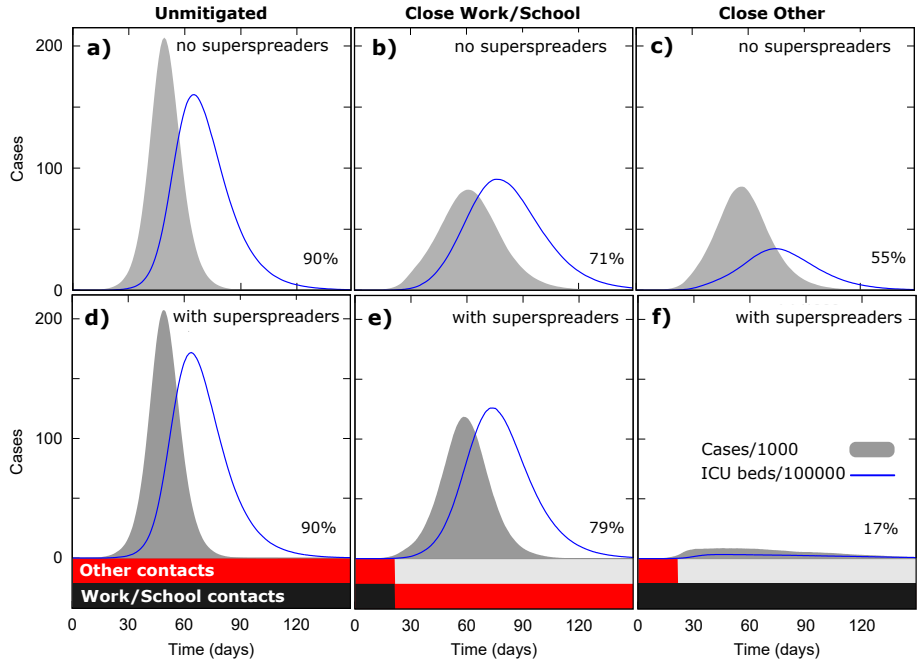


Figure 5.17: **Agent based model for Covid-19 epidemic:** Top panel show simulation without superspreaders, lower panel show simulation with superspreaders.

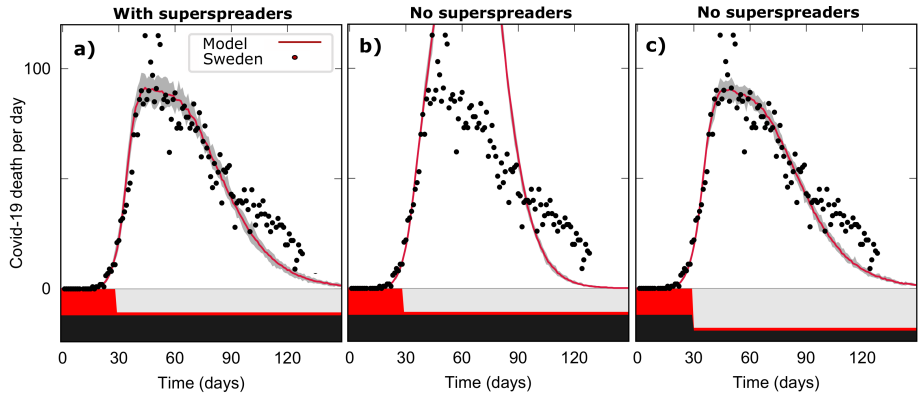


Figure 5.18: **Agent based model fit to Sweden Covid-19:** The figure show that the model with superspreaders provide a fit to Swedish mortality data for Covid-19 that is more plausible than a model without superspreaders (rightmost panel). The shown data was the record until 1 july 2020.

ing "other" contacts randomly chosen from the population became much greater than the benefits when we did not include superspreaders (Fig. 5.18 c, f). The projected number of case were both substantially smaller when superspreaders were included in the model.

Questions:

5.5) Simulate the SIR model with $\gamma = 1$ and $\lambda = 10$, starting with $I = 0.000001$

and $S = 1$. Assume that all individuals R are dead, but that there is a birth process by adding a term $+0.1 \cdot S \cdot (1 - S)$ to the equation for S. Simulate the long time dynamics of this disease.
5.6) Formulate an extended SIR model, where there are two populations and infection from one to the other occurs (but not the reverse). Assume equal population sizes and same parameters γ and λ for all allowed infections. Compare the population collapse in the two populations for $\gamma = 1$, and $\lambda = 5$.
5.7) Simulate the SIRS model with $\lambda = 5$, $\gamma = 1$ and starting with $I = 0.0001$ and $S = 1$ (corresponding to a population of 10,000). Assume, in addition to the standard SIR model, that R is converted to S with rate 0.01. Simulate the long time dynamics of this disease. Simulate the long time dynamics of a disease in a Gillespie algorithm with a population size 10,000. Assume that there is always one infecting individual (of the total population of 10,000. This prevents extinction of the disease).
5.8) Construct a agent based model for an epidemic where 10% do all infections, but all are equally susceptible. Assume a normal SIR framework, with an infectious period of 10 days and that the infection rate is such that each of the superspreaders can infect 30 other persons in the beginning of the epidemic. Consider a society with 10000 persons.
a) Assume first that persons contact each other randomly across the population, and follow epidemic trajectory starting with 1% infected.
b) Assume instead that each person is embedded in an Erdos-Reynei network with average connectivity $k = 5$. Use same infection parameters as before and calculate epidemic trajectory starting with 1% of population infected.

5.5 Persistently competing states

Every major religion today is a winner in the Darwinian struggle waged among cultures, and none ever flourished by tolerating its rivals. - E. O. Wilson

Mini tutorial: Can you mention any example of meta-stable systems in physics/your surroundings?

Mini tutorial: What is the probability to stay in a potential well of depth E at temperature T?

Competing states are part of society, where opinions spread through social contacts [104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 98, 99, 93, 114]. Heavily studied systems are the "voter models" [107, 108], where agents take one of two opinions $+1$ or -1, and update these by repeatedly setting the states of pairs of agents to be equal. Fig. [**?**] shows coarsening in a voter model that starts with five different states assigned randomly to a number of agents on a one-dimensional line. In fact, the Voter model will always coarsen to a state where eventually only one opinion survives, and all agree on everything. However, if one adds some external noise to the model then the dynamics will stop this coarsening, and the system instead stabilizes at a finite level of coarsening given by the level of the noise. Other interesting approaches include the Axelrod model [104], where opinions are multidimensional

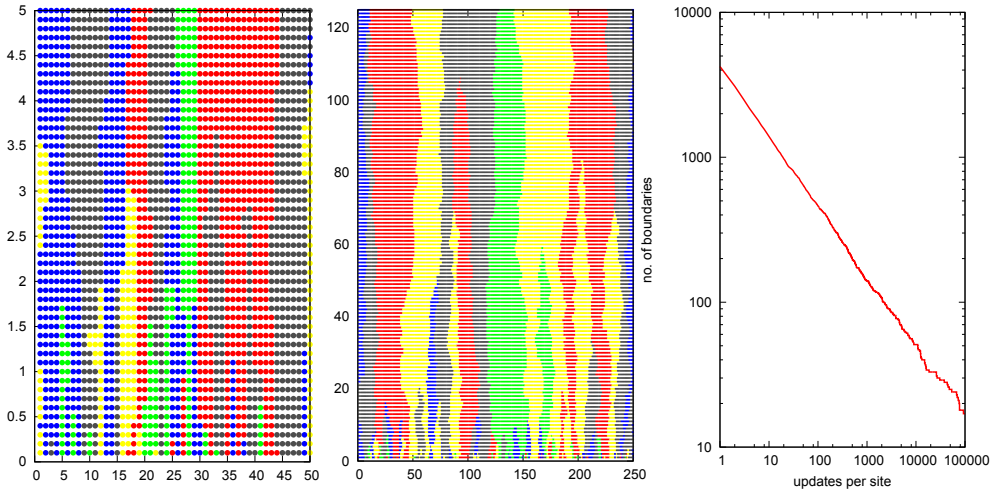


Figure 5.19: **Dynamics of the Voter model**. At each step one selects one site and then sets its state to the same as its neighbor. In the simulation above we assume five different states. The system always coarsens and the dynamics of boundaries perform random walks. Coarsening happens when two random walkers meet, and annihilate the opinion between them. Notice the self-similarity of the two coarsening pictures, where a five times larger system coarsens to a similar number of patches after a 25-fold longer time. Because all boundaries perform random walks the domain sizes grow as $\sqrt{time}$ and the number of boundaries coarsens as $1/\sqrt{time}$, see rightmost panel for simulations of an L=100,000 system.

(and agents only communicate to the extent that they at least share some opinions with each other). Also this model coarsens, but without noise it will freeze in a state of non-communicating clusters. Allowing for noise in the communication rule, where also agents without anything in common sometimes communicate, ultimately leads to a uniform state.

5.5.1 Voter model with cooperativity

We now want to introduce cooperativity, and first do this in terms of a simple two-state model. These two states are denoted by L and R, and the basic idea is that two R's are needed to convert one L→R.

The two-state model is described in Fig. 5.20. The system can be mathematically characterized by the fraction of R-states $r = R/N$, where N is total number of agents in the system. Given a certain value of r, the opposing fraction is $1 - r$.
In case recruitment is attempted, there is a probability r to pick an M-agent, and a probability $(1-r)^2$ that two subsequent random sites are opposing this state. Thus, the probability that $r \to r - 1/S$ due to recruitment is $m(1-m)^2$.
In case a direct (non-recruited) event is attempted, $r \to r - 1/N$ with probability r. Thus, the probability that $r \to r - 1/N$ due to direct transition is βr.

Adding up all recruitment and direct transitions per unit time, the change in r is given by a Langevin equation (with a noise term denoted ξ that takes into account

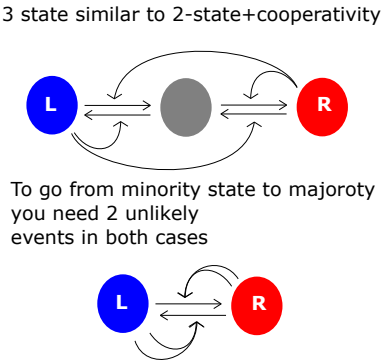


Figure 5.20: **Models for bi-stability.** The three-state model (upper panel) does not require cooperativity. The schematic of the two-state model (lower panel) indicates, that two representatives of R are needed to convert one L into an R. This process represents cooperativity. Each site represents an agent that can be either in the R or in the L state. Transitions between these two states are in part random, and in part recruited: At each update of an agent i, with probability β the agent i is set to the opposite state. Subsequently, two other agents are chosen, and if these two are in an equal state then the state of another random agent conforms to this state.

that events are discrete and in random order):

$$\frac{dr}{dt} = (r^2 \cdot (1-r) - r \cdot (1-r)^2) - \beta \cdot r + \beta \cdot (1-r) + \xi(t)\,. \tag{5.13}$$

Here, the average noise $\langle \xi \rangle = 0$ and variance of the noise $\langle \xi\xi \rangle_t \propto 1/N$. The above equation can be rewritten as

$$\begin{aligned} \frac{dr}{dt} &= r(1-r)(2r-1) + \beta \cdot (1-2r) + \xi \\ &= (r(1-r) - \beta)\,(2r-1) \;+\; \xi \end{aligned} \tag{5.14}$$

The above equation have one steady state solution ($dr/dt = 0$) when $\beta > 1/4$ and 3 solutions for $\beta < \beta_c = 1/4$. For small β there are therefore two stable solutions: One at low, another at high r, separated by a barrier at the unstable state with $r = 1/2$.

Effect of Cooperativity: In case one only required one methylated site to make the recruitment, eq. 5.14 would be replaced by

$$\begin{aligned} \frac{dr}{dt} = (r(1-r) - r(1-r)) \;\; &+ \;\; \beta(1-2r) + \xi \\ &= \quad\; \beta \cdot (1-2r) + \xi\,, \end{aligned}$$

which naturally self-organizes to one solution, namely $r \sim 1/2$. Therefore, cooperativity is inherently coupled to driving the system away from the intermediate state.

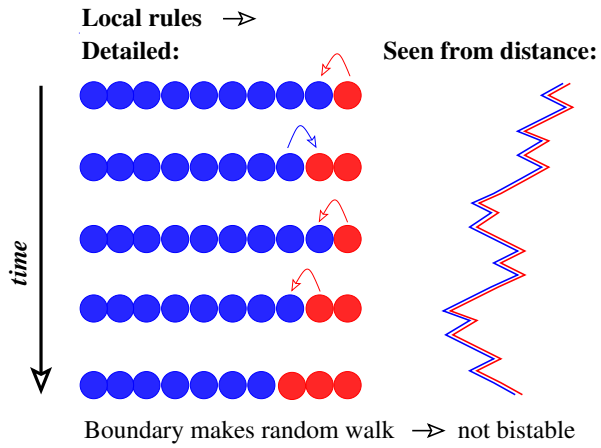


Figure 5.21: **Considerations for bistability in one dimension.** Local model in 1-d cannot give bistability, even when two neighbors in the same state are needed to convert a site. To see this, assume again a balanced model, where R and L are equally "strong". First, the system will coarsen into several domains. Subsequently, the interface between the part of the system with R and the part with L is insensitive to the majority state and therefore simply performs a random walk. The movement does not depend on who is part of the majority, that is, there is no cooperativity. Note the analogy to the argument for lack of phase transition in the one-dimensional Ising model (Sec. 1.6).

one-dimensional version does not give bistability: In the one-neighbor-only system there is no clear threshold between silenced and active states. The boundary between R and L regions wanders along the DNA as a random walk, see Fig. 5.21. This contrasts the non-local model system which is strongly pushed away from intermediate states and spends the vast majority of its time in a low R or high R configuration.

The difficulty of obtaining clear two state behavior in the neighbor-only model reflects transition dynamics which are similar to those found in the one-dimensional Ising model (Ch. 1) or the helix-coil transition in polymer physics. In fact the fact that the 1 dimensional cooperative Voter-like-model cannot give bistability resembles the classical arguments for the absence of phase transitions in one dimensional systems.

5.5.2 Bi-stable Environments

The climate in the Sahara has been different, even within the current warm period on the planet. About 6000 years ago large areas in the Western and Southern Sahara were covered by vegetation. The subsequent collapse is presumably due to some climate change, perhaps associated with slight changes in the Earth's orbit around the sun. But the land vegetation is also exposed to local feedback mechanisms, where rain favors vegetation, which in turn again favors more rain.

One model of the above is (*Brovkin et al. J. Geophys. Res. 103 (1998)* and *Liu et al, Geophysical Research Letters 33 (2006)*) which suggest a positive feedback

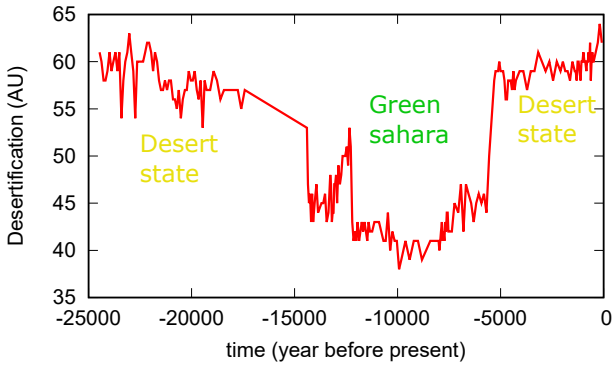


Figure 5.22: **Bistable vegetation model.** Switch to and from desert stage in Western Sahara, measured by dust in sediments off the coast. 6000 years ago the "green state" terminated.

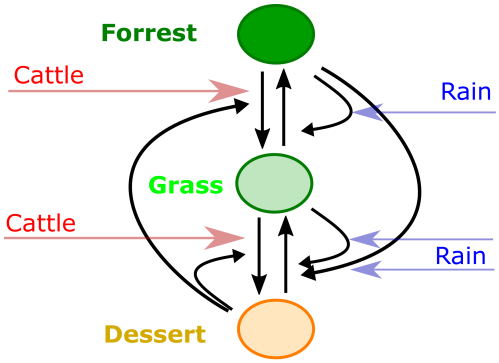


Figure 5.23: **Model for a micro climate.** Imagine a lattice, where each site can be either tree (T), grass (G) or dessert (D). All states tend to spread with some rate to the neighbor states, with cattle being responsible for the removal of grass and forest whereas random seeding (β) accomplishes the opposite. The extreme states (tree and dessert) influence conversion of each other through their influence on water drainage and wind erosion. The overall state of the system can be perturbed by external drivers, like rain (favoring growth) or cattle (destroying vegetation).

between vegetation V and rain R.

Consider now a model, where spatial sites are treated as agents that influence neighbor sites. The model is outlined in Fig.5.23, and implemented on a lattice. The interactions reflect the impact that the presence of one state has on growth or decay of other nearby environments. For example, desert (D) may drain water away, and thus would tend to convert forest (F) to grassland (G), and grassland to desert. Reversely, if forest maintains water it might thus facilitate the growth of more grass and forest by maintaining a more humid micro-climate with more water in the soil. For simulation results by G. Halvorsen, see Fig. 5.24.

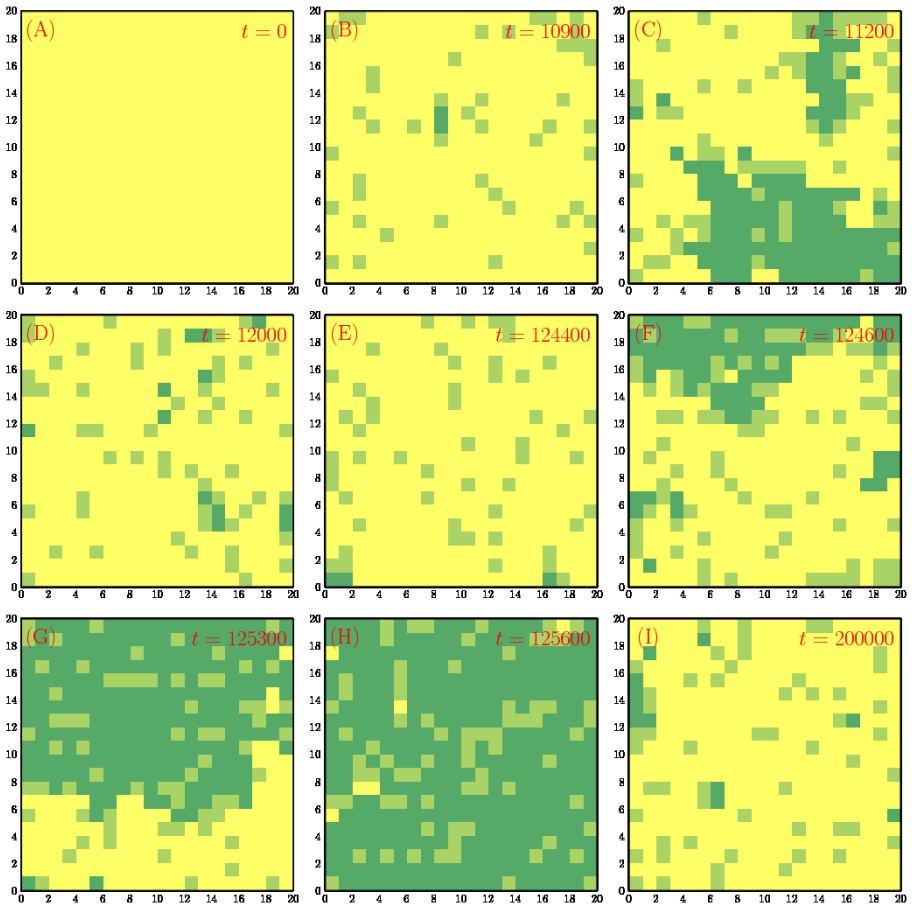


Figure 5.24: **Simulation of the spatial implementation of the model in Fig. 5.23.** Sites interact with neighbor sites and convert these dependent on their own state. Notice that the subsequent panels correspond to very different time intervals (as labeled in orange).

5.6 The Gillespie Simulation Method

Sometimes an agent based model can be simplified to a differential equation that corresponds to some kind of well mixed approximation. However, one may still want to maintain some stochasticity associated to the underlying discreteness of the processes. This can often be done by an event based simulation, with the sole additional input being the size of changes that is associated to each of the independent processes.

Gillespie (1977) developed a stochastic approach to kinetics, using an *event-based algorithm* to deal with many molecules that react with each other as they randomly collide with each other. The main idea can be transferred to stochastic simulation of differential equations, that anyway are supposed to represent random encounters between individuals.

The main assumption is that all events are random Poisson processes, and thus that the probability for the next event of any particular event is decaying exponentially with some rate set by this particular event type. Thus the chance that this

event happens after time Δt is

$$P_{next-event}(after\ \Delta t) = exp(-r\Delta t) \tag{5.15}$$

To select when an event actually occurs, the exponential decay specifies the cumulative probability that the event occurs at times larger than Δt. One should select uniformly a random number $ran \in [0,1]$ and find the δt which solves $exp(-r\Delta t) = ran$. The Δt selected in this way will be exponentially distributed. Accordingly, if the current time is t, then the next event should be assigned to occur at $t + \Delta t$ with

$$\Delta t \ = \ -\frac{1}{r} \cdot \ln(ran)\,. \tag{5.16}$$

Here, $1/r$ is the average time to the next event.

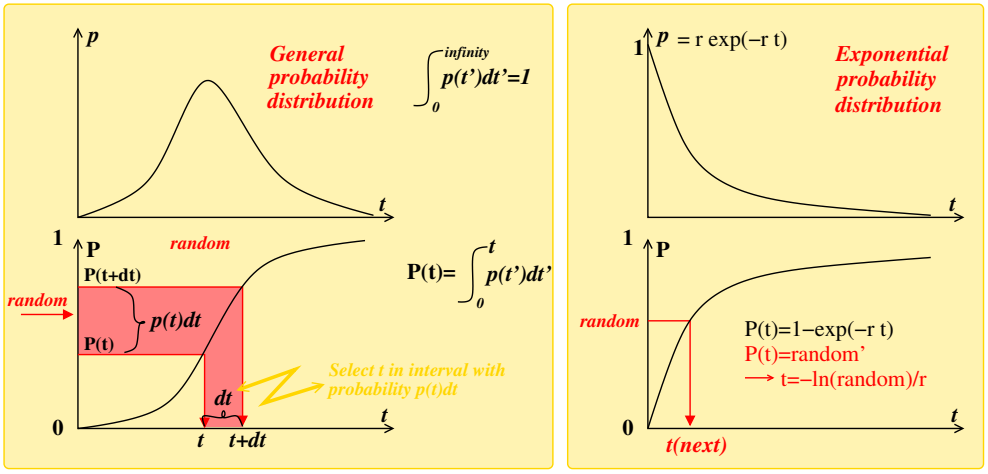


Figure 5.25: **Time selection in the Gillespie Algorithm.** Selection of t according to some predefined distribution function $p(t)$. One constructs the cumulative distribution $P(t)$, selects a number *random* uniformly in $[0,1]$, and finds the t that achieves $P(t) = random$. Thereby, the part of $p(t)$ that has high values will be selected at higher likelihood, because there are more *random* values that correspond to values where $P(t)$ has a large slope. That is, if p is doubled within an interval $[t, t+dt]$, the slope of P is doubled and the probability to select t is also doubled. If $p(t)$ is an exponential, then $P(t)$ is also exponential and the solution is analytic (as shown in the right panels).

In the typical case where several competing events can take place, one at any time needs to make a list of times with one time for when each of these events should occur next. This could, for example, be events where a variable increases, and competing events where the same variable decreases:

- $x \to x+1$ with rate r_1 (that might be a function of x)
- $x \to x-1$ with rate r_2 (that also might be a function of x)\,.

Other variables might also be included and inter dependent on one-another.

One update consists of selecting the first (earliest) events in this list, and then making the change specified by the reaction chosen. Subsequently, the rates for many of the other events may then change, which will then serve as new input for the next event. Also, one should keep track of the total time during the simulation, always updating this time with a step size given by the time used for the selected events.

Mini tutorial: If doubling the event sizes, how should one scale the rates to maintain the same average behavior?

After initialization with a start number for all variables an update step in the event driven algorithm reads:

1 Monte Carlo step: Generate random numbers ran_i to determine the time-step $\Delta t_i = -\frac{1}{r} \cdot \ln(ran_i)$ for all potential events i, and select the first event for updating.

2 Update: Increase the time by the timestep generated (from Step 1) and update the variables with the change associated to the selected event.

The key assumption is that we consider systems without memory. That implies that any event only depends on the quantified state of the system, and that this occurrence is independent on how much time has passed since there last was a change in the system.

Mini tutorial: How could the event sizes be larger than one unit in a system?

Mini tutorial: If event size is doubled for one of the processes, what would that mean for the resulting noise?

Now equations like

$$\frac{dx}{dt} = Q - \frac{x}{\tau}$$

can be rephrased in terms of a stochastic dynamics of the variable x. That is, imagine that x is the number of molecules, and changes happens in units of one molecule $x \to x + 1$, respectively $x \to x - 1$:
Q would then be the rate for the change $x \to x + 1$.
x/τ would be the rate for change of $x \to x - 1$.

The above changes can also be different for each of the terms in the differential equation. For example, the change in increase could occur by making 2 units at each event, while the decay term still could be in 1 unit at a time. This all depends on the underlying "physics" of the problem. If each term is changed in steps of Δ_i the simulation of a dynamics with $i = 1, 2, \ldots$ different processes should proceed as:

1 Monte Carlo step: Generate random numbers ran_i to determine the time-step $\Delta t_i = -\frac{\Delta_i}{r} \cdot ln(ran_i)$ for all potential events i, and select the first event for updating.

2 Update: Increase the time by the generated time-step (from Step 1) and update the variable x with the change Δ_i associated to the selected event.

Questions:
5.9) Draw 100,000 random numbers $ran(i)$ uniformly between 0 and 1, and for each number set $x_i = -\ln(ran(i))$. Plot the histogram of x_i. Fit an exponential function to this histogram.
Qlesson: It is simple to simulate exponential distributions.
5.10) Use the Gillespie algorithm to simulate the dynamics of $dx/dt = 12 - x$ with x changing in steps of 1. Plot over 1,000 events. Then redo simulation when production of x changes in steps of 4, while removal still changes in steps of one. Compare the variation (standard deviation) in the timeseries of x in the two cases.
5.11) Draw 100,000 pairs of random $t_1(i)$, $t_2(i)$ each from a distribution $exp(-t^2/2)$, with $t > 0$, that is, as the right side of Gaussian with center in 0 and standard deviation one, and compare the distribution of $t_2 - t_1$ for all cases where $t_2 > t_1$ with the distribution of t_2.
Hint: if one draws 12 numbers uniformly between 0 and 1, their sum is Gaussian with standard deviation 1 and mean 6.
Plot the histogram of $dx = x_i(2) - x_i(1)$ for all pairs where $x_i(2) > x_i(1)$.
Qlesson: The distribution of second events, given the first have occurred, is different when using Gaussian distributions (or any other distribution than an exponential).
5.12) Make a Gillespie simulation of the 2-state recruitment model with cooperativity as formulated in terms of the different processes in eq. 5.13. Let m vary between 0 and 1 in steps of 0.04 and set $\beta = 0.1$. Change step size to 0.03 and check how stability of one of the states increases.
Qlesson: Should correspond to a agent based model with $N = 25$.

Lessons:

- Emergence and aggregate properties can emerge as a consequence of many iterations of simple update rules between pairs of agents.
- Often simple estimates of behaviour can be obtained in the well-mixed limit, where the probability for each reaction is simply the product of densities:

$$Frequency\ for\ an\ encounter\ between\ A\ and\ B \propto \rho_A \times \rho_B\,, \tag{5.17}$$

  where ρ_A and ρ_B are the densities of agent type A, respectively type B. However such approaches have their limitations in systems where individuals have repeat their interactions to certain other specific agents.
- Models which are formulated in terms of well mixed populations can be extended to include noise by using event based simulations (Gillespie simulation). In this case, each reaction i is assigned the time when it happens the next time:

$$t(next) = t(now) - ln(random)/rate_i\,, \tag{5.18}$$

  where $rate_i$ is the rate of the corresponding reaction given the state of system. The first of all reactions is then set to occur at its designated time, and time updated accordingly.

Supplementary reading:

Van Kampen, Nicolaas Godfried. Stochastic processes in physics and chemistry. Vol. 1. Elsevier, 1992.

Samanidou, Egle, et al. "Agent-based models of financial markets." Reports on Progress in Physics 70.3 (2007): 409.

Farmer, J. Doyne, and Duncan Foley. "The economy needs agent-based modelling." Nature 460.7256 (2009): 685-686.

5.7 Appendix

It may sometimes be useful to perform stochastic dynamic using a Langevin equation, where one follow the development of a variables in small fixed time steps, or perhaps even in the Fokker Plank formalism where one propagate a whole ensemble of systems.

5.7.1 Langevin versus Fokker Planck equation

In physics the mobility associated to the how fast a particle can be dragged through a viscous medium. There is a fundamental relation between mobility μ and molecular noise quantified by the diffusion constant D. If x fulfil the Langevin equation

$$\frac{dx}{dt} \;=\; -\mu \;\cdot\; \frac{dV}{dx} \;+\; \sqrt{2D}\cdot\eta(t) \tag{5.19}$$

with $\langle \eta(t'')\eta(t')\rangle = \delta(t''-t')$ then one can construct a probability distribution for x at a given time t:

$$P(x,t)dx = probability \;\; for \;\; x\in[x,x+dx] \;\; at \;\; time \;\; t \tag{5.20}$$

By using a physical insights the two terms in the Langevin equation can be recast in terms of loss and gain of particles within $[x,x+dx]$, $P(x,t)\cdot dx$. $P(x,t)$ will evolve according to the Fokker Planck equation:

$$\frac{dP(x,t)}{dt} \;=\; -\frac{d}{dx}J \tag{5.21}$$

where the current J is given by

$$J \;=\; -\mu\cdot P\cdot\frac{dV}{dx} \;-\; D\cdot\frac{dP(x,t)}{dx} \tag{5.22}$$

At equilibrium $P = constant$ implying that $J=0$ and thus

$$P(x,t) = P(x) \propto e^{-\mu V(x)/D} \tag{5.23}$$

which can only be $\propto\; e^{-V(x)/k_BT}$ if $\mu = D/k_BT$. This famous equation by Einstein states that the mobility (μ) is proportional to diffusion constant D. The diffusion constant D has dimension of a mean free path times a typical (thermal) velocity. The diffusive part of the Fokker-Planck equation describes how an initial localized particle spreads out, in flat potentials to a a Gaussian with spread $\sigma\propto\sqrt{Dt}$ that simply follows from the central limit theorem. The convective part of the Fokker-Planck equation states that the particle moves downhill, with a speed proportional to both dV/dx and the mobility μ.

5.7.2 Kramers equation

Escape from a potential well is an old and important problem in physics, chemistry as well as for a number of larger scale bological problems related to stability of genetic switches and punctuated evolution in abstract fitness landscapes. We here present the derivation proposed by Kramers [115] for escape from a potential well.

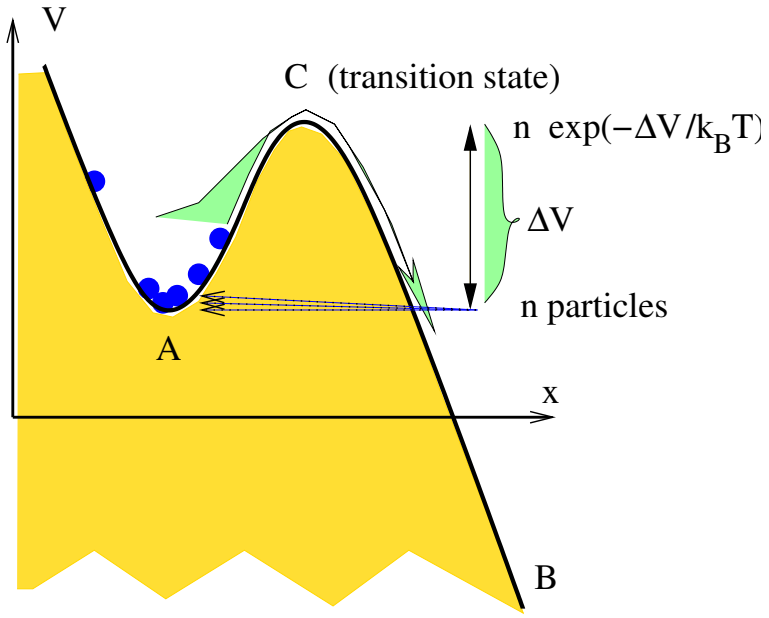


Figure 5.26: **Escape over a 1-d potential.** A particle is confined at A, but allowed to escape over barrier C to the outside of the well. In the figure we show a number of particles, to illustrate the approach by Kramer.

Consider a 1-dimensional potential well as in figure 5.26 where point A is at the bottom of the well and point B is somewhere outside the well. Following Kramer we consider the stationary situation where the current leaking out from the well is insignificant. Thus current J in the corresponding Fokker-Planck equations is a constant number that is independent of position x and the corresponding P is constant in time. The position independent current can be rewritten

$$J \;=\; -\frac{D}{T}P\frac{d}{dx}V \;-\; D\frac{dP}{dx} \;=\; -D\cdot e^{-V/k_BT}\cdot\frac{d}{dx}\left(P\cdot e^{V/k_BT}\right) \tag{5.24}$$

When rewritten and integrated from point A to point B:

$$J\cdot e^{V/k_BT} \;=\; ;-D\cdot\;\frac{d}{dx}\left(P\cdot e^{V/k_BT}\right) \tag{5.25}$$

or

$$J \;=\; -D\;\frac{\left[Pe^{V/k_BT}\right]_A^B}{\int_A^B e^{V/k_BT}dx} \tag{5.26}$$

where the quasi stationary condition states that $P_B \sim 0$ and $P_A \approx$ local equilibrium value:

$$J \;=\; D\;\frac{P_A e^{V_A/k_BT}}{\int_A^B e^{V/k_BT}dx} \tag{5.27}$$

The value of P_A can be estimated from using a harmonic approximation around the minimum A, and setting P_A equal to the corresponding max of the corresponding Gaussian density profile. Thus around point A:

$$V(x) \approx V_A + \frac{1}{2}\frac{d^2V}{dx^2}(x - x_A)^2 = V_A + \frac{1}{2}k\,(\delta x)^2 \tag{5.28}$$

For a particle with mass m this is a harmonic oscillator with frequency $\omega = \sqrt{k/m}$. The peak density P_A is given by normalization of $\exp(-k\delta x^2/(2k_BT))$ (everything has to be counted as if there is one particle in the potential well that can escape). The escape rate (J per particle):

$$r \;=\; \frac{\omega_A \tau}{m}\sqrt{\frac{mk_BT}{2\pi}}\,\frac{e^{V_A/k_BT}}{\int_A^B e^{V/k_BT}dx} \tag{5.29}$$

The remaining integral is calculated by a saddle point around its maximum, i.e. around the barrier top at point C:

$$\int_A^B dx\; e^{V/k_BT} = e^{V_c/k_BT}\int_{-\infty}^{\infty} dx\; e^{-k_c^2\delta x^2/(2k_BT)} \;=\; \frac{\sqrt{2\pi k_BT}}{\sqrt{k_c}}e^{V_c/k_BT} \tag{5.30}$$

which with $k_c = m\omega_c^2$ gives the final escape rate for overdamped motion:

$$r \;=\; \frac{\omega_a}{2\pi}\cdot\omega_c\tau\cdot\exp(-\frac{V_c - V_A}{k_BT}) \tag{5.31}$$

This equation can be interpreted in terms of a product between a number of attempted climbs:

$$number\;\; of\;\; attempts \;=\; \frac{\omega_A}{2\pi} \tag{5.32}$$

multiplied by the fraction of these climbs that can reach c, simply given by the Boltzmann weight $e^{-(V_c-V_A)/k_BT}$. Finally just because a climb reaches the saddle point, it is not given that it will pass. The chance that it will pass is $\omega_c \cdot \tau$ which is equal to one divided with the width of the saddle, in units of the steps defined by the random kicking frequency $1/\tau$. I.e. imagine that the saddle is replaced by a plateau of $w = 1/(\omega_c\tau)$ steps, and we enter the first (leftmost) of these steps. We then perform a random walk over the plateau, with absorbing boundaries on both sides. As this is equivalent to a fair game, the chance to escape on the right hand side is $1/w$. Thus one may interpret the overdamped escape as:

$$\begin{aligned} r \;=\;& (\;\; Attempts\;\; to\;\; climb\;\;) \\ &\cdot\; (\;\; chance\;\; to\;\; reach\;\; top\;\; given\;\; it\;\; attempts\;\;) \\ &\cdot\; (\;\; chance\;\; to\;\; pass\;\; top\;\; given\;\; it\;\; reached\;\; top\;\;) \end{aligned} \tag{5.33}$$

and one immediately notice that a higher viscosity, meaning a lower τ implies that escape rate diminishes. This is not surprising, as a higher viscosity means that everything goes accordingly slower, and therefore also the escape.

Lesson: The escape from a potential well is exponentially difficult in the inverse temperature, where $T \propto D$ from Einsteins equation. In our Agent based models, the effective temperature would be the variance of the noise term. When noise in

our Gilespie simulation is double as big then the escape is as if the "temperature" is four times bigger.

Chapter 6

Econophysics

For then, since gold was soft and blunted easily, man would deem
it useless, but bronze was a metal held in high esteem.
Now the opposite: bronze is held cheap, while gold is prime.
And so the seasons of all things roll with the round of time:
What once was valuable, at length is held of no account,
while yet the worth of which was despised begin to mount.

Lucretius, De Renum Natura, Book 5 (60 years before crist)

Figure 6.1: *Unlimited growth of a computer currency, at present (autumn 2017) with a total capitalization value of about 50 billion US dollars. And this in spite of having no backing from country or bank.*

6.1 Analysis of a Time Series

Mini tutorial: Why is money valuable for a society

J.K. Galbreith statement *"The only function of economic forecasting is to make astrology look respectable"* is also a implicit reflection on the fact that if one could predict the future then it would be easy to make profit in for example the stock

market. Time series analysis of stock prices in part reflect the ancient dream of predicting the future from the past in order to make profit. Much effort is put into the analysis of time series of especially stocks, and anyway, as we will see, then they are inherently un-predictable. We will here outline some of the simplest measures.

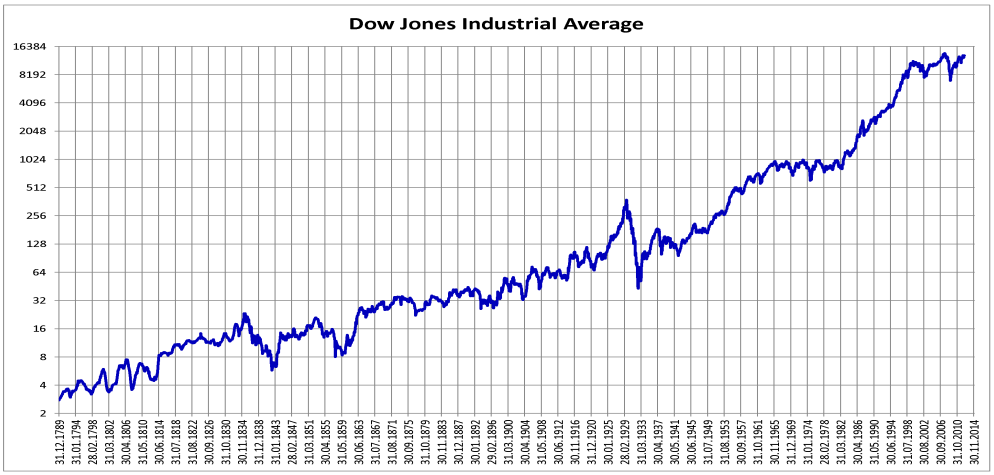


Figure 6.2: **Dow Jones:** *An index following the average of the major shares in USA. The index increases with about a factor 4,000. For comparison, the US public debt changed from* $\sim 10^8$*\$ in the period 1800-1850 to* $\sim 5 \times 10^{12}$*\$ in year 2000.*

Fig. 6.2 shows a stock market index during a 200 year period. The index is calculated as the average of many shares, and should thus in principle be much less variable than individual shares. In spite of this, there are indeed wild fluctuations, with occasional collapses where the overall value of all stocks drops by a factor 10 over a relatively short period. In fact, when one inspects stock markets across the world, then nearly all of them have had about one reduction by a factor 10 during the last century. Value is dynamic.

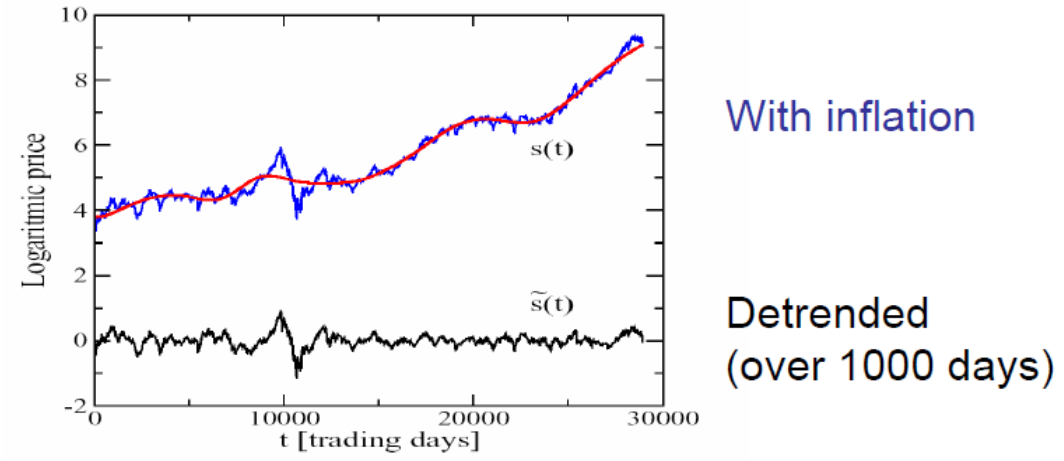


Figure 6.3: **De-trended Dow Jones index.**

To first approximation the market exhibits a biased random walk. More precisely, de-trending for the overall increase due to general growth of the economy/inflation, log(*price*) follow a random walk. In Fig. 6.3 we show the de-trended Dow-Jones index, removing trends that are more than about 5 years long. The random walk hypothesis was first put forward more than a century ago by Bachelier [116], and has been recently supported by analyzing price fluctuations $W(t)$ as function of time:

$$W^2(T) = \langle (\log v(t+T) - \log v(t))^2 \rangle_t = \langle (\Delta s(T))^2 \rangle \,, \tag{6.1}$$

where the average is taken over all starting times t of intervals of duration T in the available time series.

For a random walk $W(T) \propto T^{0.5}$, whereas most stock markets show $W(T) \propto T^{0.55 \to 0.65}$ with the lowest values of the Hurst exponent for the oldest markets. Notice that one can define the Hurst exponent in terms of both the variance of prices over a time interval with length T, or instead just define it in terms of the variation after a time interval T. In both cases it involves sampling a lot of different starting points!

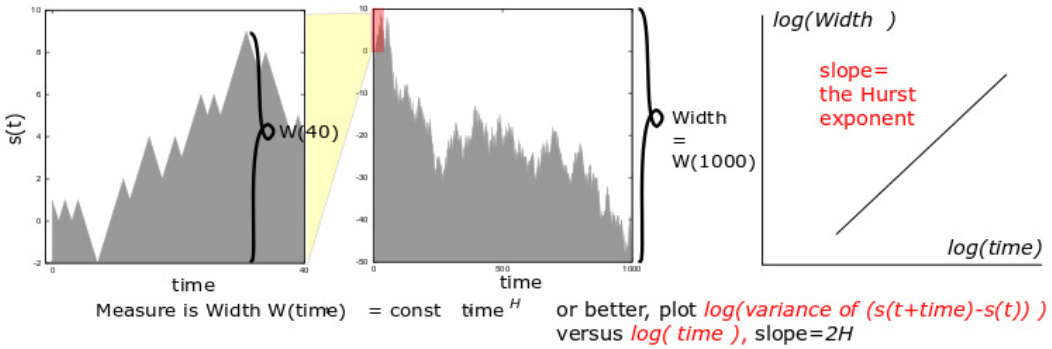


Figure 6.4: **Hurst exponent simplified.** The scaling between the spread in $s = \ln(v)$ when measured over different times T. Thus the spread in $\Delta s = s(t+T) - s(t)$ is a function of T.

To characterize the stochastic dynamics of a time series one uses the Hurst exponent. The Hurst exponent is defined by the scaling of the typical change in price over a time interval of length T

$$\langle (\Delta s(T))^2 \rangle = \langle (s(t+T) - s(t))^2 \rangle_t \propto T^{2H} \tag{6.2}$$

where one normally follows the logarithm of the price $s(t) = \log(v(t))$. This measurement is performed by averaging over all starting points t in a given time series, using the prescription shown in Fig. 6.4.

In economic time-series one follows the logarithm of the price because it is the relative change in price that actually matters. That is, this determines how much your investment gives in return. Thus, if a share changes value from 10 to 11, or from 100 to 110, it is the same relative change, and the same change in $\Delta \log$. The scaling assumption in the above equation reflects the near-random walk behaviour of the market, where deviations grow with time with some exponent, that in fact is close to that of a random walk.

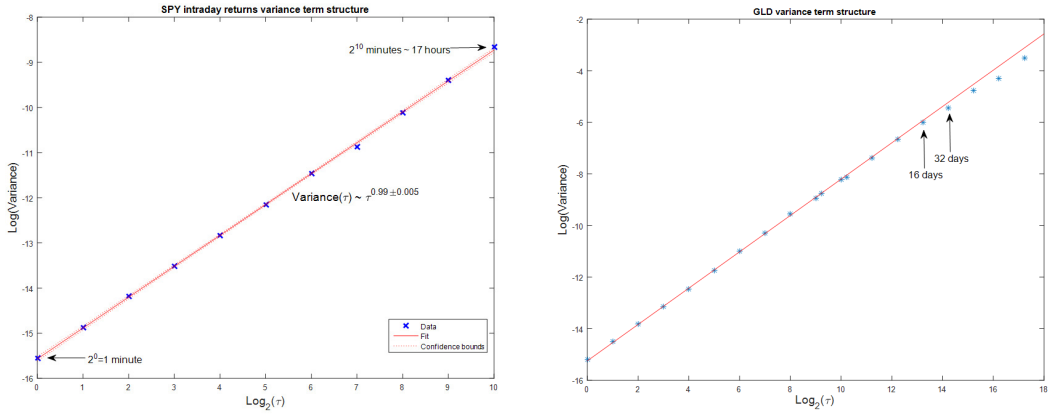


Figure 6.5: **Example of scaling of variance with time.** Shown on different timescales: The left panel focuses on the trading within one day; The right panel includes trading up to one year. In both cases the slope is close to one, corresponding to $H = 1/2$. Figure from a blog post by Ernie Chan in QUANTITATIVE INVESTMENT AND TRADING IDEAS, RESEARCH, AND ANALYSIS (2017).

The correlation between the past and future is related to to the Hurst exponent H. Consider the variation around the present time, $t_0 = x$, with forecast at a time T in the future $\Delta s(T) = s(T+x) - s(x)$ whereas the historical counterpart is given by $\Delta s(-T) = s(x-T) - s(x)$. Thus, we want to calculate the correlation between past and future

$$C \equiv \frac{\langle (s(t) - s(t-T)) \cdot (s(t+T) - s(t)) \rangle_t}{\langle (\Delta s(T))^2 \rangle_t} = \frac{\langle -\Delta s(-T) \cdot \Delta s(T) \rangle_t}{\langle (\Delta s(T))^2 \rangle_t}, \tag{6.3}$$

a calculation that demands a few intermediate steps.

The denominator in eq. 6.3 is the variance over the time interval T, which can be re-expressed as:

$$\begin{aligned} f(T) &= \langle (\Delta s(T))^2 \rangle_x = \langle s^2(x+T) + s^2(x) - 2s(x)s(x+T) \rangle_x \\ &= 2(\langle s^2(x) \rangle - \langle s(x)s(x+T) \rangle_x) \propto T^{2H}, \end{aligned}$$

where we use the assumption that an average over all starting time points x makes $\langle s(x+T)^2 \rangle_x$ and $\langle s^2(x) \rangle_x$ equal.

The numerator in eq. 6.3 can be re-expressed

$$\begin{aligned} \langle -\Delta s(-T) \Delta s(T) \rangle_x &= \langle -(s(x-T) - s(x))(s(x+T) - s(x)) \rangle_x \\ &= -\langle s(x-T) \cdot s(x+T) \rangle_x + \langle s(x-T) \cdot s(x) \rangle_x \\ &\quad + \langle s(x) \cdot s(x+T) \rangle_x - \langle s^2(x) \rangle_x \\ &= -\langle s(x) \cdot s(x+2T) \rangle_x + \langle s^2(x) \rangle_x \\ &\quad + 2\langle s(x) \cdot s(x+T) \rangle_x - 2\langle s^2(x) \rangle_x \\ &= \frac{1}{2} f(2T) - f(T). \end{aligned} \tag{6.4}$$

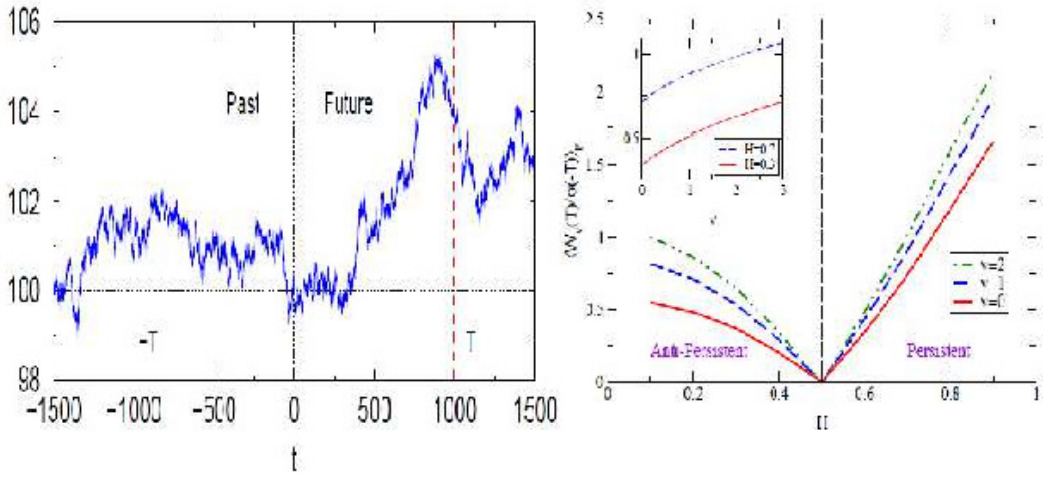


Figure 6.6: **Past→ future.** Left shows an example of a time series with Hurst exponent $H = 0.40$, generated by a wavelet method (not pensum). Right panel examines the average return of investment as a function of H, where one buys according to trend [117]. The red curves shows the profit when one buys on the way up, and sells on way down in $H > 0.5$ markets, and oppositely in $H < 0.5$ markets. The two other curves invest proportional to size of the past price change $\nu = 1$, respectively to this change squared $\nu = 2$. Thus, weighting the trend pays off even more. All returns are measured in units of the spread in volatility during the time interval considered, and the curves in fact scale proportionately to this as the horizon T for investment increases.

Using that $f(T) = const \cdot T^{2H}$ one obtain [118, 117]

$$C = \frac{\langle -\Delta s(-T) \cdot \Delta s(T) \rangle_t}{\langle (\Delta s(T))^2 \rangle_t} = 2^{2H-1} - 1 \tag{6.5}$$

Thus, an ordinary random walk with $H = 1/2$ has $C = 0$, whereas an $H > 1/2$ walk implies that the past price difference $\Delta s(-T) = s(0) - s(-T)$ is most likely maintained for $\Delta s(T) = s(T) - s(0)$. That is, if the price increased during the past month, then it on average will increase also during the next month. In contrast, in an $H < 0.4$ market the price fluctuations will tend to revert.

To get an interpretation of the above correlation, consider a stock that on a time scale T follow the trend with probability p and reverses it with probability $1 - p$. The variance for one step of this walk is $\langle (\Delta s(T))^2 \rangle_t = 1$. The numerator in eq. 6.5 is given by the sum of two contributions, one for following the trend,and one for reversing the trend

$$\langle -\Delta s(-T) \cdot \Delta s(T) \rangle_t = 1 \cdot p \cdot 1 + 1 \cdot (1-p) \cdot (-1) = 2p - 1 \,.$$

Accordingly, using eq. 6.5 one find that p is associated to the Hurst exponent by $2p - 1 = 2^{2H-1} - 1$ or $p = 4^{H-1}$. This is the probability to follow the trend:

$$\text{Propability (follow the trend)} \sim 4^{H-1} \,, \tag{6.6}$$

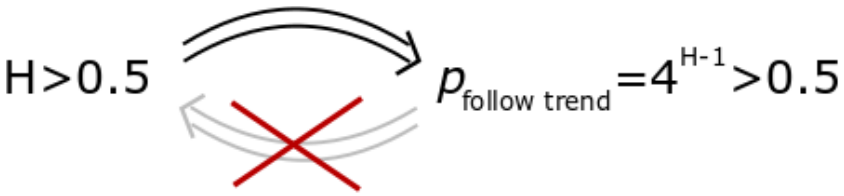


Figure 6.7: **Following the trend.** When $H > 0.5$ then there is more than a 50% chance that the next move is in same direction as the previous move. However, the reverse is not true! A tendency to follow the trend typically implies a random walk with a longer "persistence length," i.e. a longer time before the walk changes direction. On this longer timescale the walk will still be a random walk.

a statement that qualitatively should be true for all time intervals where the walk can be characterized by H (*see* also Fig. 6.7). In particular for $H = 1/2$ then the above probability is equal to $1/2$, reflecting a true unbiased event. In the questions we will try to use this to gain profit in correlated markets.

In the $H > 1/2$ case, a winning strategy is to "bet" on the trend: Buy when it is bull market, and sell when it becomes bear market [117]. Thus for $H > 1/2$ one should:

$$\text{Buy at } t \text{ if } s(t-T) < s(t) \tag{6.7}$$
$$\text{Sell at } t \text{ if } s(t-T) > s(t) \tag{6.8}$$

whereas this strategy should be reversed in a $H < 0.5$ market, see Fig. 6.6. Noticeably, electricity markets have $H = 0.40$ [119, 120]. Again we emphasize that this buy-sell strategy would work on with trading intervals anywhere inside the timescale where the walk is characterized by the Hurst exponent H.

Finally, as a small notice, then an walk with hust exponent H has a fractal dimension $2 - H$ when one consider the position as function of time (the walk embedded in 2-dimensional space-time). From this one can calculate the first return for various H-walks. Do this! (for help see chapter 2)

- **Summary:** In spite of all the people thinking, talking and dealing, the resulting market nearly behaves as a random particle exposed to Brownian noise. Second order correlations presumably reflect crowd panic.

Questions:
6.1) Simulate a walk where the logarithm of a price (s) moves one step up or one step down at each time-step. Let the probability to continue in the same direction as in the previous step be $p = 0.75$. Investigate the Hurst exponent for this walk numerically. Redo the simulation for $p = 0.99$. *Hint*: Just calculate the variance for one hundred simulated time-series of length 100, hundred time-series of length 1,000, and hundred timeseries of length 10,000. Plot the variance of end points on a log-log scale. (You can equivalently use one very long time-series and extract various segments from it).

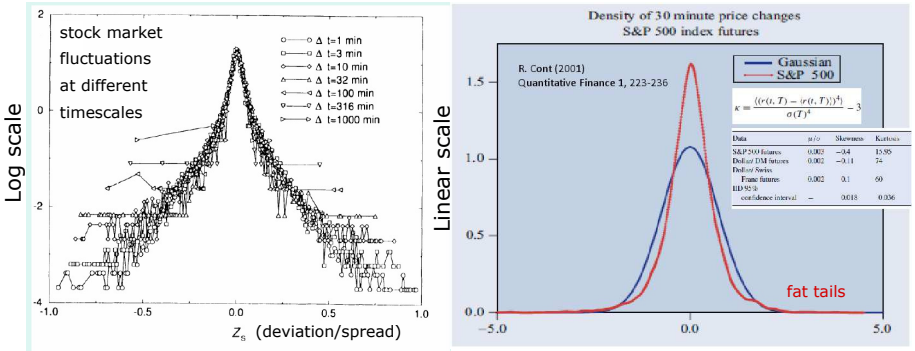


Figure 6.8: **Fat tails:** *Distribution of short time-scale fluctuations exhibit fat tails. The left panel show short timescale fluctuations of an index, re-scaled with the timescale over which one examine the fluctuations. In right panel, the red and blue curve have same variance, but different Kurtosis. Kurtosis quantifies 4'th moment, normalized by second moment squared. It is more sensitive to tails in distribution than second moment, and would thus be divergent when* $p(tail) \propto 1/\Delta s^{\tau}$, *with* $\tau \leq 5$.

Qlesson: Any finite $p > 1/2$ still leads to a random walk, just with a correlation time that is proportional to $ln(1-p)$ (what is the pre-factor?, what would happen for $p < 1/2$?).

6.2) Simulate a random walk of uncorrelated up and down movements of s, where step sizes δ are chosen from the fat-tailed distribution $P(\delta) \propto 1/\delta^3$. Visualize the walk. Calculate the Hurst exponent by simulation.

Qlesson: Notice that the mean squared displacement diverges.

6.3) Plot eq. 6.5 as a function of the Hurst exponent H, and interpret this in terms of profit of a sensible strategy. Devise an investment strategy and calculate the maximum average profit per investment step for an $H = 0.4$ market.

Qlesson: Act as if tomorrow would be opposite to today.

6.4) Generate a market profile by the upper envelope of directed percolation, using a critical value of p (and restarting a new seed at last present seed when all live sites in the DP dies out). That is, when the upper branch dies out, one experiences a sudden collapse. Analyze the Hurst exponent of this market. Try to devise an investment strategy to make money in this market, and simulate the investment strategy assuming that it is the logarithm of the price that follows this trajectory.

Qlesson: This is a persistent walk (exponent 0.63) with occasional collapses that can be very very large. Follow the trend but bet hedge (see later).

6.2 Fear-Factor model

Mini tutorial: How can the time variation of the sum be asymmetric when parts are symmetric.

"In economics, the majority is always wrong." by John Kenneth Galbraith. This classic quote can in fact be quantified by considering the coordinated movement of many stocks. To explore economic time series we now consider inverse statistics

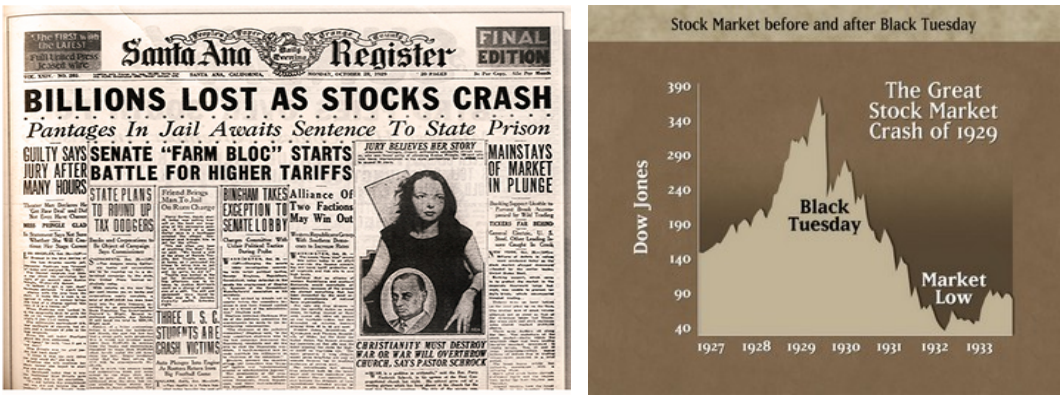


Figure 6.9: **Fear and Panic in 1929.**

[**?**, 121]. In turbulence one often measures velocity differences as a function of distance, and obtains the famous Kolmogorov scaling. However, one could also consider the inverse statistics that measure the time or the distance until the next large fluctuation in relative velocity occurs. Thus, the inverse statistics focus attention on the laminar/calm regions of the fluid [**?**], with large distances corresponding to large laminar regions. In economics the corresponding measure is associated to the time it takes before one obtains a given return on an investment. This will take a long time when stocks are calm, or when fluctuations are in the opposing direction as the one that is aimed at.

Let $v(t)$ denote the asset price at time t. The logarithmic return at time t, calculated over a time interval Δt, is defined as $\Delta s(T) = s(t+T) - s(t)$, where $s(t') = \log v(t')$. We consider a situation in which an investor aims at a given return level, ρ, that may be positive (being "long" on the market) or negative (being "short" on the market). If the investment is made at time t, then the inverse statistics, also known as the "investment horizon," is defined as the shortest time interval $\tau(t) = T$ fulfilling the inequality $\Delta s(T) \geq \rho$, given that $\rho \geq 0$. For losses $\rho < 0$ one similarly defines the first time T where $\Delta s(T) \leq \rho$. The inverse statistics histogram, or in economics, the "investment horizon distribution", $p(\tau_p)$, is the distribution of waiting times T for obtaining the strike price. It is obtained by averaging over all initiation times t in the available time series.

The data set used is the daily close of the DJIA covering its entire history from 1896 until today. Fig. 6.10 depicts the empirical inverse statistics histograms for the investment horizon distributions. The distributions are shown for a return of 0.05 with open blue circles and a return of -0.05 with open red squares. The histograms possess well-defined and pronounced maxima, the optimal investment horizons, followed by long $1/t^{3/2}$ power-law tails.

Remarkably, the optimal investment horizons with equivalent magnitude of return level, but opposite signs, are different. Thus, the market as a whole, monitored by the DJIA, exhibits a fundamental gain-loss asymmetry. As mentioned above, other stock indexes including SP500 and NASDAQ, also show this asymmetry, while, for instance, foreign exchange data on currencies do not.

It is even more surprising that a similar well-pronounced asymmetry is not found for any of the individual stocks constituting the DJIA. This can be observed from the insert of the figure, which shows the results of applying the same procedure

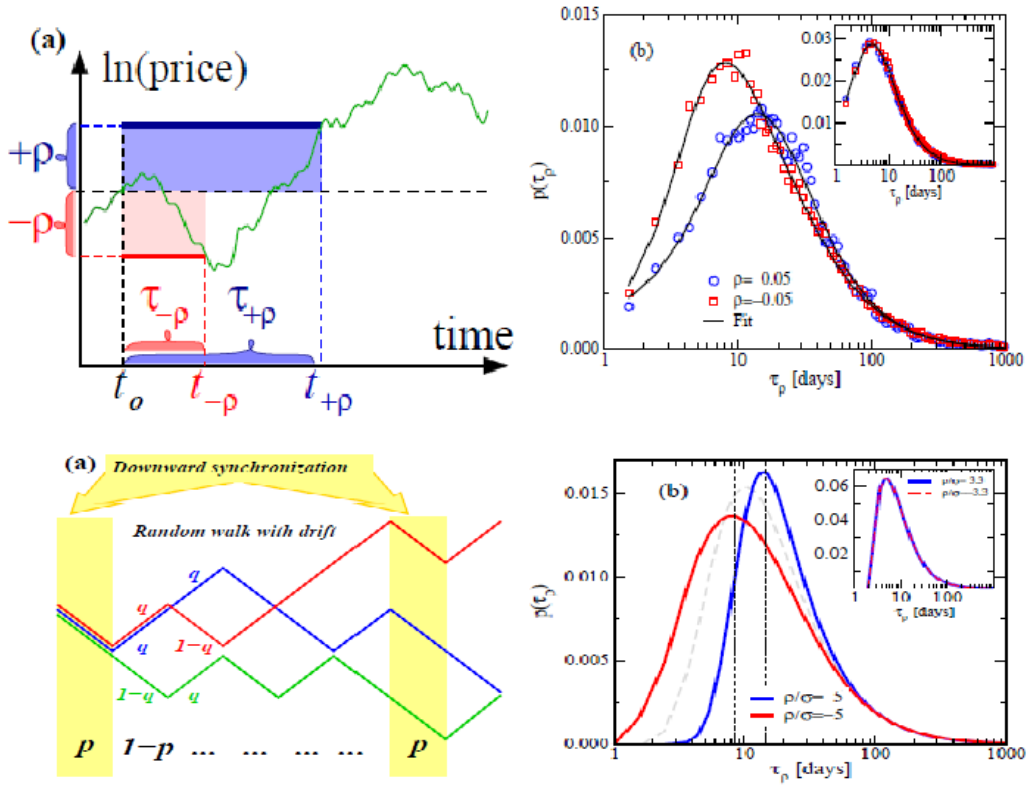


Figure 6.10: **Inverse statistics and Fear-factor model.** The upper two panels show the definition of "strike price", and the distribution as measured from the de-trended Dow-Jones index. The blue curves show the number of days when the price first exceeds the current price by 5%, the red when it first lies 5%below its current price (inset shows corresponding distributions for individual companies). To read the curves, the x-axis labels the day following the investment and the y-axis labels the probability that the price reaches the 5% deviation at that day. Lower panels define the model and show predicted strike-price distributions.

individually to these stocks, and subsequently averaging to improve the statistics. Hence, the question is: why does the index exhibit a pronounced asymmetry, whereas the individual stocks do not? This question is addressed by the fear-factor model introduced below [122].

The main idea is the presence of occasional short periods of dropping stock prices synchronized between all N stocks contained in the stock index, e.g., during crises, see Fig. 6.9. In essence, these collective drops are the cause (in the model) of the asymmetry in the index.

It is assumed that the stochastic processes of the stocks are all equivalent and consistent with geometrical Brownian motion. This implies that the logarithms of the stock prices, $s_i(t) = \log v_i(t)$, follow standard, unbiased, random walks

$$s_i(t+1) = s_i(t) + \epsilon_i(t)\delta, \quad i = 1, \ldots, N\,, \tag{6.9}$$

where $\delta > 0$ denotes the common fixed log-price increment (by assumption), and

Figure 6.11: **Fear is a strong driver at stock markets.**

$\epsilon_i(t) = \pm 1$ is a random time-dependent direction variable.

At certain time steps, chosen randomly with fear factor probability p, all stocks synchronize a collective draw down ($\epsilon_i = -1$). For the remaining time steps, the different stocks move independently. Thus the shares behave independently with probability $1-p$. To assure that the overall dynamics of every stock behaves equivalent to a geometric Brownian motion the typical movement needs a slight bias.

Let q be the chance to move up ($\epsilon = +1$) in calm periods, and $1-q$ the probability to move down ($\epsilon = -1$). If the probability to have collective fear and synchronous downward move is p the probability to move up for one company is $(1-p)\cdot q$ whereas the probability to move down is $p + (1-p)\cdot(1-q)$. A neutral walk demands that these probabilities have to be equal:

$$(1-p)\cdot q = p + (1-p)\cdot(1-q)\,, \tag{6.10}$$

fixing q in terms of the probability for overall fear:

$$q = \frac{1}{2\cdot(1-p)}\,. \tag{6.11}$$

$q > 1/2$ is a "compensating" drift that governs the non-synchronized periods [1]. From the price realizations of the N single stocks, one may construct the corresponding price-weighted index, like in the DJIA, according to

$$I(t) = \frac{1}{N}\sum_{i=1}^{N} v_i(t) \tag{6.12}$$

and investigate inverse statistics for this (Fig. 6.10). Overall result: DJIA is reproduced with one collective fear that occurs with probability $p = 0.05$ per day, corresponding to one panic event per month or so. The other parameter is $\rho = 5\cdot\sigma$,

[1] Note that there are only solutions when $p < 0.5$. For larger p the market is doomed, as it is not possible to compensate the overall disasters

where σ is the standard deviation of the volatility of the index (average stock movement) and we use an index of $N = 30$ shares. For DJIA the typical daily fluctuations have $\sigma = 1\%$.

We conclude that the asymmetric synchronous market model captures basic characteristic properties of the day-to-day variations in stock markets. The agreement between the empirically observed data, here exemplified by the DJIA index, and the parallel results obtained for the model give credibility to the point that the presence of a "fear-factor" is a fundamental social ingredient in the dynamics of the overall market (*see* also the cartoon in Fig.6.11).

- **Summary:** Crowd behavior and panic on even the relatively small scale of a once in a month event can be seen by using of inverse statistics.

Questions:
6.5) Consider the fear factor model with 10 stocks that move one step up or down, all starting at 1,000. With probability $p = 0.05$ all stocks move down simultaneously. What should the probability for other up, respective down, movements be in order to let individual stocks perform a unbiased random walk? Simulate the system and plot the time series for the average stock price.
Ups and downs of the average are asymmetric, but the average change is zero with Hurst exponent 1/2.

6.3 Models of economic time-series

6.3.1 A model of Economic Bubbles

There have been multiple examples of economic bubbles [124, 125], including the Dutch Tulips (1637), the South sea company (1711-1720), and the Japanese stocks (in the 1980'ties) and perhaps the contemporary phenomenon of bitcoins. Cryptocurrencies are particularly interesting because of their purity: Their "fundamental" value is presumably zero [125, 126]. While there exist some considerations on positive feedback [127, 128], bubbles [129] and depressions [130], the economic literature is not settled on the non-equilibrium nature of value [131, 132, 133, 134]. Following Cecilie Toftdahl Olesen & Kim Sneppen (2019) we here present an agent based model of value, assuming that it is given solely by attention. In a nutshell, a topic is proposed to be "interesting" in proportion to how often one has heard about it [135].

Fig. 6.14 illustrates our model in terms of the amount of memory M that the agents allocate to the good, and the number of exchanges S in the society at large. These two aspects of value may not be in equilibrium with each other. In particular, the dynamics is out of equilibrium when the positive feedback of a current fashion acts on a faster timescale than the dominating negative feedback.

Consider N agents that trade in D different goods. The goods do not exist physically, but are recorded in terms of two types of memory.

First, each agent has a memory that consists of μ slots which each can be assigned a number corresponding to one of the D product types. If an agent has several memory slots assigned to the same good, she is more likely to talk about this good.

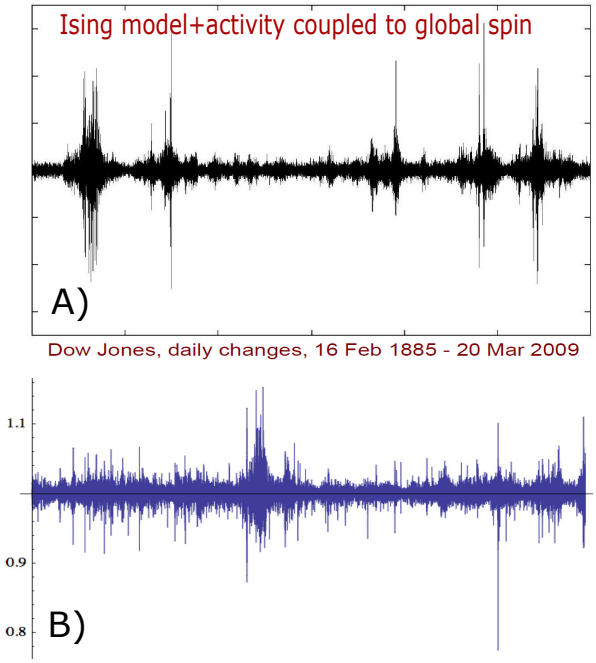


Figure 6.12: **Market model for volatility.** A) $M(t+1)-M(t)$. The volatility it is supposed to reproduce the data shown in B) illustrating daily returns, $(S(t)-S(t-1))/S(t-1)$, for the Dow Jones stock market index. Fluctuations are correlated: When variations on one day are large, then they most likely are large again the next day [123]. The directions of these fluctuations are uncorrelated! Volatility clustering is sometimes also discussed in terms of the GARCH model (Tim Bollerslev 1986).

Figure 6.13: **Herding.** A market driven by herding can give huge volatility (Kalton, in the Economist)

Second, there is a globally shared memory of how much each product was traded during the last $\tau \cdot N$ timesteps.

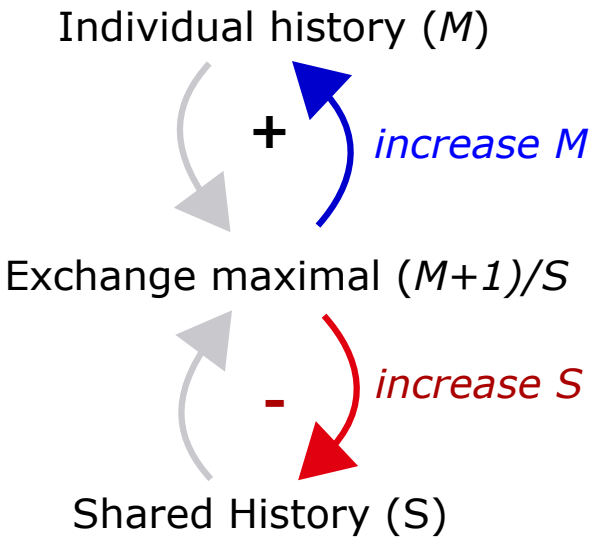


Figure 6.14: **Bitcoin model.** Model with positive (blue) and negative feedbacks (red) on the trading activity of products. The positive feedback amounts to peer-to-peer communication, whereas negative feedbacks come from common history of all past transactions.

The memories are implemented in lists:

- Local Perception: Agent i has memory "slots" m_{ij}, $j = 1, \ldots, \mu$ that each can be assigned one of the D product types.
- Global Perception: All agents have access to the information of the common global trading activity S_j associated to $j = N \times \tau$ earlier trading events. Each of these global trading positions is assigned one of the D product types.

The model is executed in steps. At each step, two random agents i and j are picked. Agent i now selects the good k that he/she considers to have the highest relative value:

$$p_{ik} = \frac{1 + \sum_{l=1}^{\mu} \delta(M_{il} - k)}{\sum_{l=1}^{N \cdot \tau} \delta(S_l - k)} = \frac{1 + M_i(k)}{S(k)}\,, \tag{6.13}$$

and decides to share this to agent j. Here the δ function is $= 1$ if the corresponding memory slot referring to product type k. The sum in the numerator thereby counts the number $M_i(k)$ that the product k occurs in the memory of agent i. The added number 1 in the numerator avoids absorbing states where a product is absent in the memory of all agents. When several products has same maximal p_{ik} one randomly chose one of these to be the active one.

The exchange causes the following changes in our memory lists:

- First, one adjust the local memory by inserting the chosen product k in one randomly chose place x, $M_{jx} = k$, in the receiving agent j.
- Second, one adjusts the shared memory S by inserting the chosen product k in one randomly chosen place y, $S_y = k$, in the global memory.

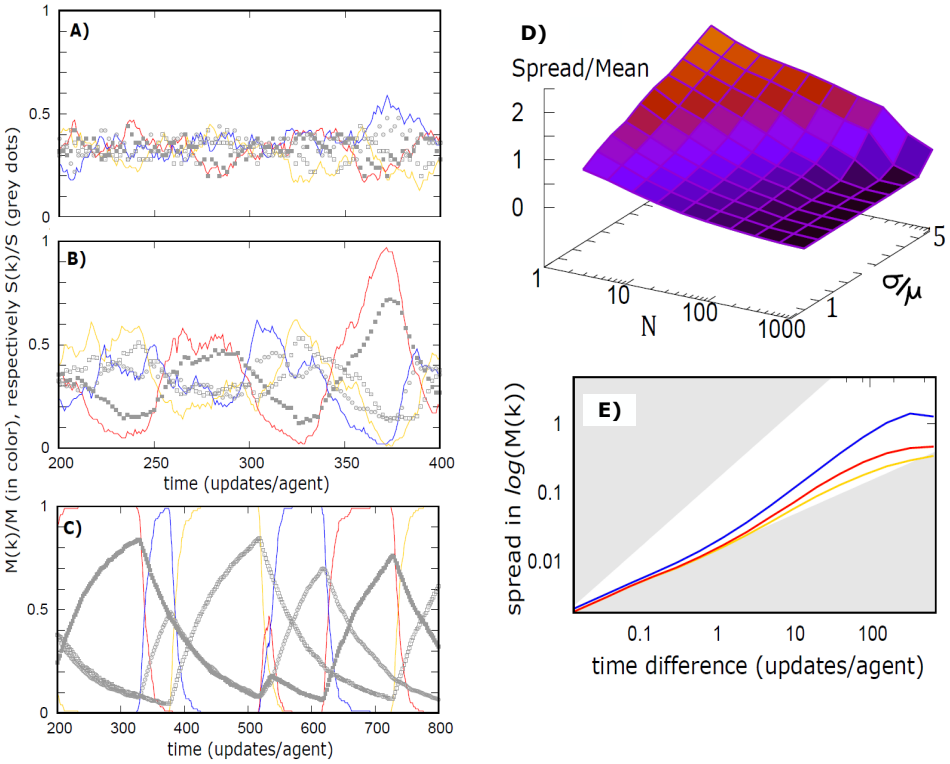


Figure 6.15: A-C) Attention $M(k)$ (colors) for each of $D = 3$ different products in an economy with $N = 10$ agents each having $M = 10$ memory slots. Grey dots show the global market share G of each product. A) $\tau = 0.5 \cdot \mu$, B) $\tau = \mu$, C) $\tau = 2 \cdot \mu$. D) Coefficient of variation for the total memory $M(k)$ of a product. The simulation uses an individual memory $\mu = 20$ and an economy with $D = 20$ products. E) Spread (square root of variance) in $M(k)$ as function of time separation Δt between measurement points. Parameters: $\mu = 200$, $D = 100$ with $N = 100$ agents. τ is 100 (yellow), 200 (red) and 400 (blue) respectively. Grey shaded areas are bounded by slope $H = 1$, respectively $H = 0.5$.

When the new memory is inserted, an old memory "bit" is discarded. Thereby, τ defines the characteristic time for adjustment of the global market, whereas μ is the lifetime of the individual memory.

The global trade activity of a product k, $S(k) = \sum_{u=1}^{\tau} \delta(S_y - k)$, reflects the common history/processes shared among all agents. When $S(k)$ of a product k is large it means that it has been traded a lot in the past. Increases in $S(k)$ could for example reflect production of the "traded" product, with an increase in this number making each copy less valuable.

Fig. 6.14 highlights the two feedback mechanisms in the model: A positive feedback of fashions/viral marketing on a peer-to-peer scale, and a negative feedback that acts through the dynamics of a common marketplace. The lower panel in Fig. 6.14 show the total local memory, $M(k) = \sum_i^N M_i(k)$ that is allocated to a particular good k. The simulation was done for a small economy where the common history have a relatively long lifetime $\tau > \mu$.

Fig. 6.15 explore the impact of length of global memory, with the common market G adjusting faster (A), equal (B) respectively slower that the memory of the

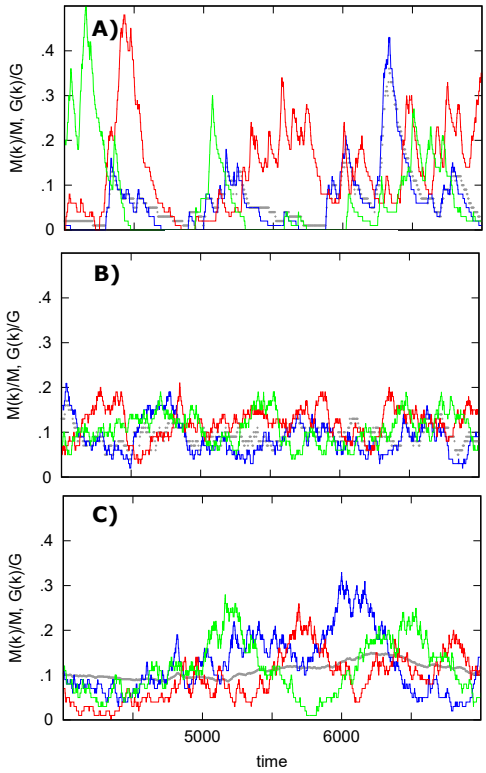


Figure 6.16: Simulation with $N = 1$ (memory just copied from one position to another). All panels show 3 products out of $D = 10$. Grey dots show $S(blue)/S$. A) Standard model with $\mu = \tau = 100$. B) Linear model where copied memory is selected proportional to $(1 + M(k))/S(k)$. Again $\mu = \tau = 100$. C) Linear model as in B) but with $\mu = 100$ and $\tau = 1000$.

agents (C). There is a shift from random fluctuations around an equilibrium value ($= 1/D$), persistent fluctuations, to oscillation like behaviour. The latter pattern of alternate dominance (D) has resemblance to the oscillations obtained with frustrated bi-stability in some biological circuits [136, 137].

The stochastic oscillations in Fig. 6.15C) are characterized by shifting dominance of products in an order that is partly random. That is, when diversity is high then products which are not dominating tend to adjust toward each other, and the next winner is selected by chance/contingency events.

Fig. 6.15D) explore the variation in the attention allocated to a given product $M(k)$. The top panel show that large variations occur when the length of the global memory τ is larger than a threshold. This threshold increases with the local memory μ and increase proportional to the logarithm of the number of agents N. Thus the requirement (on τ/μ) for emergent fashions only grows slowly with N.

Fig. 6.15E) explore the variation in total memory M as function of the time interval Δt over which these variations are measured. For guidance we also mark behaviour corresponding to Hurst exponent $H = 0.5$, respectively $H = 1.0$. There is a limited scaling regime, with an apparent exponent that varies between a random walk ($H = 0.5$), and a fully persistent walk $H = 1$ (as in Fig. 2C).

Fig. 6.16 investigate $N = 1$ case corresponding to a direct copy of memory from one position to another in both the M-list and the S -list. Panel A show that even

similar length of the two memories, $\tau = \mu$, leads to large bubbles. Comparing the blue curve $M(k = 1)$ with the corresponding grey dots $S(k = 1)$ one see that the two memory lists are nearly in equilibrium. Noticeably, their difference is enough to drive quite large fashions.

Fig. 6.16B,C) relax the assumption of copying the product with maximal $(M(k)+1)/S(k)$. Instead we at each step select a random product to be copied with probability $P(k) \propto (M(k) + 1)/S(k)$. One see that variations is smaller (panel B same parameters as panel A). However increasing τ to $\tau >> \mu$ again leads to large variations in value. Thus the soft proportional selection can cause "bubbles" if the two timescales are widely separated.

Mini tutorial: Consider a model where one select with probability $\propto ((M(k) + 1)/S(k))^2$. How would that behave compared to the two models explored in above text?.

Other models of fashions from mimicking behaviour of agents was introduced in [138, 139], where in particular the model of Galam et al. explored a cooperative threshold for establishing dominance. Thresholds for dominance also emerged in the cooperative voter model from chapter (see also [140]). These competing scenarios aims to describe objects/topics that are not exposed to negative feedback. Religions is an example where such negative feedback seems to be abolished, at least for very long timescales. In particular the monotoistic religions obtained dominance in their respective regions on the timescale of thousands of years. In contrast, our model predict fashions that dynamically replace each other.

In our model the amplitude of the fashions depend on a relative slow negative feedback. Thus they depend on a "reality" that do not always keep up with the demand. This speaks to engineering of fashions, emphasizing that they require to *NOT* produce enough of a product to fulfill demand.

In our model then a high production (possibly introduced by sudden increase in S) would lead to a declining M/S and a subsequent collapse in attention. In the real world such behaviour could be caused by accumulated action of many companies each producing maximally to optimize their own profit. This has analogy to the game theoretical "tragedy of the commons" by [141]. In this light "Brand names" allow tightly controlled production, and thereby a slower negative feedback.

The presented model only considered two feedback loops, while real products in their complicated reality would be exposed to history on a variety of timescales: Production, production facilities, education, common sense, fundamental user value, saleability, and social euphoria. In addition Fisher [130] suggested feedback associated to accumulation and spreading of debt as a driver for depressions. Some of the above feedback can fuel a cycle of reinforcements, and bubbles may emerge when these are relatively fast. In any case, we here focus on cases where these "bubbles" ultimately become unstable, reflected in the assumption of a negative feedback on the longest timescales.

Our model suggest to view collective human minds as elements of an excitable media with a dynamics of transient "bubbles" of attention. Excitability and dominance emerge from a relatively fast positive feedback, whereas its finite duration is

associated to some sort of slower acting reality. A perspective that is partly compatible with the phenomenological description of the waves of war presented in "War and Peace" by Tolstoy [142].

Mini tutorial: How could the presented model be implemented as a model of an excitable medium (in a 2 dimensional space)?.

Questions:
6.6) In the above model we use the heat bath method, therefore repeat a simulation of the ising model for a 10 times 10 system as function of beta (1/temperature) and plot the energy and average magnetization as function of of temperature.
Qlesson: It works
6.7) Simulate the above Ising inspired model for volatility in a market model using a $N = 10 \times 10$ system with $\beta = 0.7$ and $\alpha = 1$, respectively $\alpha = 2$ and 5.
Qlesson: By coupling agents to their overall average in some sort of frustrated ways, an extended system can exhibit irregular dynamics. Perhaps this can be done more elegantly than with the Ising model using the recruitment models from chapter 5...
6.8) Simulate the N=1 version of the fashion model above, using $D = 5$ products and $\mu = 20$ and $\tau = 10$, respectively $\tau = 20$ and $\tau = 50$. Estimate Hurst exponent.
Qlesson: Notice the sensitivity of results with τ
6.9) For the $D = 2$ case the "bubble model" only contain the variables $m_1 = M(1)/\mu$ and $s_1 = S(1)/\sigma$ and can be studied through the eqs:

$$\begin{aligned} \frac{dm_1}{dt} &= \left(\frac{m_1+\epsilon}{s_1}\right)^\gamma \cdot (1-m_1) - \left(\frac{m_2+\epsilon}{s_2}\right)^\gamma \cdot (1-m_2) \\ \theta \cdot \frac{ds_1}{dt} &= \left(\frac{m_1+\epsilon}{s_1}\right)^\gamma \cdot (1-s_1) - \left(\frac{m_2+\epsilon}{s_2}\right)^\gamma \cdot (1-s_2) \end{aligned} \tag{6.14}$$

with $\theta = \sigma/\mu$ in terms of earlier parameters. Thus $\theta > 1$ correspond to a dynamics of s that is relatively slower. The above equations can be simulated in an event based/ Gillespie simulation with update size $\epsilon = 1/\mu$ in favour of product 1 happening with rate $rate_1 = \left(\frac{m_1+\epsilon}{s_1}\right)^\gamma$. Such an event will induce a change $m_1 \to m_1 + \epsilon$ with probability $1 - m_1$ and a change $s_1 \to s_1 + \epsilon/\theta$ with probability $1 - s_1$. When m and/or s of product 1 is updated, the corresponding record of product number 2 is similarly reduced ($m_2 = 1 - m_1$, $s_2 = 1 - s_1$). Simulate the equations using a Gillespie update with $\theta = 2$. $\gamma = 2$ and step-size $\epsilon = 0.04$.

6.4 Bet hedging

6.4.1 Bet hedging in random walk markets

Following Namiko Mitarai's notes from the Stochastic Dynamics course we consider the stochastic differential equation with changes that are proportional to your investment (including multiplicative noise)

$$dv = \mu \cdot v \cdot dt + \sigma \cdot v \cdot dw \tag{6.15}$$

where dw is calculated as the change that happens in the future:

$$dw = w(t+dt) - w(t) \tag{6.16}$$

where w is a random walk process that is normalized such that

$$\langle (w(t) - w(t'))^2 \rangle = |t - t'| \tag{6.17}$$

(equivalent with $dw/dt = \eta(t)$, where $\langle \eta(t)\eta(t') \rangle = \delta(t-t')$) Thus the typical total increment squared over time interval dt is accordingly of size $(dw)^2 \sim dt$. Now in economy we are interested in relative changes, and again define $s = \log\, v(t)$. Using Taylor expansion and $(dw)^2 = dt$ we obtain

$$\begin{aligned} ds &= \log v(t+dt) - \log v(t) \\ &= \log(v+dv) - \log(v) \\ &= \frac{1}{v}\cdot dv - \frac{1}{2v^2}(dv)^2 \\ &= \frac{1}{v}(\mu \cdot v \cdot dt + \sigma v \cdot dw) - \frac{1}{2v^2}(\sigma \cdot v \cdot dw)^2 \end{aligned} \tag{6.18}$$

where we expanded to first order in dt (equal second order in dw). This the gives

$$\begin{aligned} ds &= \left(\frac{1}{v}\cdot\mu\cdot v - \frac{1}{2v^2}\cdot\sigma^2\cdot v^2\right)\cdot dt + \frac{1}{v}\cdot\sigma\cdot v\cdot dw \\ &= \left(\mu - \frac{\sigma^2}{2}\right)\cdot dt + \sigma\cdot dw \end{aligned} \tag{6.19}$$

Notice that this last equation only have additive noise, and not multiplicative noise! By integration this equation one obtain

$$s(t) = s(0) + \left(\mu - \sigma^2/2\right)\cdot t + \sigma\cdot(w(t) - w(0)) \tag{6.20}$$

where w was a simple random walk and s was the log of the value of your assets.

Notice that your average capital

$$\langle s - s(0) \rangle / t = \langle \log v(t)/v(0) \rangle / t = \mu - \frac{\sigma^2}{2} \tag{6.21}$$

easily may decrease in value, even if the average return looked positive ($\mu > 0$). [Or in other terms $\langle exp(\sigma w t) \rangle = exp(\sigma^2 t/t)$.] This is reminiscent of the earlier introduced games, where the average looked good, but the actual long term performance will be bad. The risk cost on your average performance, because a 50% down movement, followed by a 50% up movement does not cancel out. This reflect the fact that the down-movement is taken from a large absolute value than the up-movement. In the above derivation, the corrections comes about because of the Taylor expansion of the $ln(v)$ to more than first order (see eq. 6.18). The above equation are one cornerstone in evaluating the value of risk, used for setting price on future expectations, see Fig. 6.17.

As long as μ is positive one can however ALWAYS gain money on the market, by gambling with a fraction of your money $(1-x)$ (and putting x under your pillow or in some other safe deposit). The typical growth rate will then be

$$growth\ rate = \mu\cdot(1-x) - \frac{\sigma^2}{2}\cdot(1-x)^2 \tag{6.22}$$

which for small $1-x$ will be positive if $\mu > 0$. Optimal investment fraction can then be found by differentiation.

$$1 - x = \frac{\mu}{\sigma^2} \tag{6.23}$$

which can in principle exceed 1, reflecting a situation where you can borrow (here assumed at interest 0). Notice that this bet-hedging require that you always keep a fraction x away, so f market goes down, then you need to take from your safe money to keep the fraction constant. Reversely if market goes up, you will sell the asset to keep your fraction of safe money equal x. A more extensive discussion of the bet-hedging associated to this equation is done in Namiko Mitarais course in block 3.

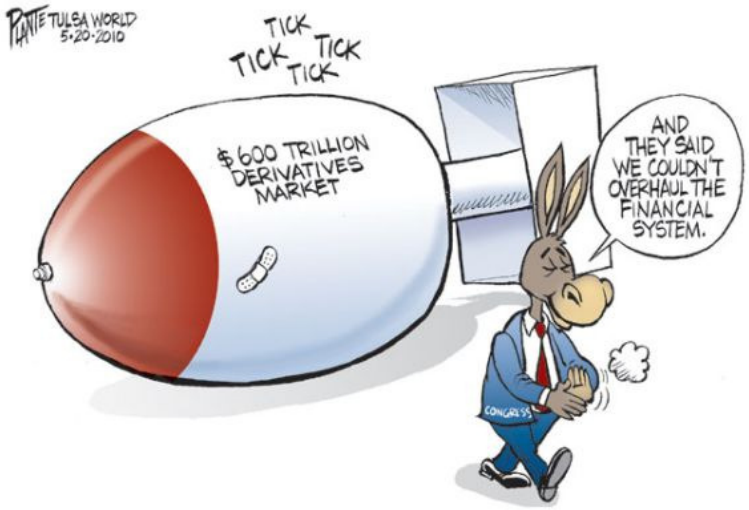


Figure 6.17: **Bet hedging.** The equation that relates risk (volatility, $= \sigma^2$) and average rent μ is a cornerstone in valuating the price of stocks in future, where one for example can buy the right to sell a given share at a given price half a year from now (without buying the share). This has become a huge market, with potential instabilities.

Noticeably, the above formalism is sensitive to the assumption about Gaussian uncorrelated random walks, where steps is always small, i.e the assumption that

$$\langle (w(t) - w(t'))^2 \rangle = |t - t'| \tag{6.24}$$

which in particular is not true if these sometimes are very big changes. (then the variation on short time intervals can be large, and not approaching zero as the above equation predict).

6.4.2 Bet-hedging with occasional catastrophes

Mini tutorial: What would a million dollars (or equivalent in their time currency) placed in a main roman bank around 0 bc have turned into today?

The size of this "investment" is in fact an information theoretical problem. It reflect optimization of long term growth rate in much the same way as gamblers

on horse races may optimize their portfolio [143], and have lately been applied to biology by for example Bergstrøm and Lachman [144] and Kussel and Leibler [145, 146]. Here we use the simplified formulation of [147], directly applicable to simple win-loose games.

Consider a game with two outcomes, one good event where everything invested get amplified by Ω, and alternatively a bad outcome where all invested capital is lost. The probability to loose is set to p, and the probability that the "bet" is successful is $1-p$. Assume for example the quit and double game, $\Omega = 2$ and $p = 1/2$, which may be modified if one for example plays with a false coin, or if one have additional information.

Given that you have a capital K, one may ask two questions:

- What is the optimal investment fraction when one play the game one time
- What is the optimal investment fraction when one can play the game many times, but only using whatever is left of the original capital.

For the one round of the game the average outcome of an invested capital of unit of money is

$$(1-p)\cdot\Omega + p\cdot 0 = (1-p)\cdot\Omega \tag{6.25}$$

When this product exceeds 1, one apparently would have the max average gain if one invest everything.

When playing the game many times, one accordingly also maximize the average return when one invest everything at each round. However the chance that you as a single player have any money left after t bets require t wins in a row, and thus becomes exponentially small

$$propability(solvent|t) = (1-p)^t \to 0 \tag{6.26}$$

as the number of bets progresses. I.e. after many time-steps, the chance to have anything left is near zero, but if you are lucky, then your capital is near infinite. Therefore it is wrong to try to optimize the average outcome. In repeated games, one should instead try to optimize the **typical** outcome. That is, in a $p = 1/2$ game, one will on average win half the games, and looses the other half. Therefore the typical gain after two games will be the product of returns for a win game and a loose game:

$$Capital \propto Win \cdot Loose \tag{6.27}$$

and if $Loose = 0$ you will typically have zero money after an equal number of wins and looses.

To be more quantitative, we now allow the player to maintain a fraction x of his capital in a safe, and only playing with the remaining fraction $1-x$ at each round. The faction x is a constant throughout all repeated rounds, and thus specifies a strategy.

In the above scenario with $p = 1/2$ and say $\Omega = 3$ corresponding to a 50% chance of tripling your fortune, the typical fortune after two games will be

$$Capital \propto (\Omega(1-x)+x)\cdot x \tag{6.28}$$

a function which have a maximum at $x = (1/2)\cdot\Omega/(\Omega-1)$. Thus for quit or doublet, $\Omega = 2$ and $x = 1$, i.e. one should not play at all. However for $\Omega = 3$, then $x = 3/4$

and one should thus play with 1/4 of capital at each round.

In general, the optimization of growth over all possible sequences of events of duration t would be obtained by maximizing

$$N(t) = N(0) \cdot \sum_{b=0}^{t} C(t,b) \cdot p^b \cdot (1-p)^{t-b}((1-x)\Omega + x)^{t-b} x^b \tag{6.29}$$

where $C(t,b)$ is the number of ways b bad events can be distributed among t total events. Optimizing the above $N(t)$ would be optimizing the average. Instead we will look at the typical contribution to N, that is where the red part of the above sum contribute most. The average of the binomial part of the above equation (shown with color red) have an expected number of bad events

$$b = p \cdot t \tag{6.30}$$

This lead to optimization of

$$\begin{aligned} N(t) &\propto ((1-x)\Omega + x)^{t-pt} x^{pt} \\ &= \left(win^{1-p} \cdot loose^{p}\right)^t \\ &= \left(((1-x)\Omega + x)^{1-p} x^p\right)^t = e^{t\Lambda(x)} \end{aligned} \tag{6.31}$$

where the average long term growth rate

$$\Lambda(x) = (1-p) \cdot \log(\Omega(1-x) + x) + p \cdot \log(x)$$

which then should be optimized with respect to the fraction x kept in the safe "bank". The first term in the growth rate is the logarithmic growth rate under good conditions where where the invested fraction $1-x$ is multiplied by Ω, while the reserves x remains unchanged. The second term is the logarithmic growth rate when the bet is lost. The expected value of the (logarithmic) growth rate Λ is then given by the mean trajectory with an average of $1-p$ good events and p losses per time unit

We emphasize, that the growth rate weights the *logarithms* of multiplicative growth factors of the entire capital under two conditions with their respective probabilities of occurrence. Maximization of Λ with respect to x secures the long-term optimal growth rate [143].

In contrast to its short-term counterpart, the long-term logarithmic growth rate $\Lambda(x)$ usually reaches its maximum at some x^* between 0 and 1. In the economics literature this is denoted the Kelly-optimal investment ratio [143]. It describes the optimal fraction of capital that a prudent long-term investor should keep in relatively safe financial assets such as bonds while investing the rest in more risky assets such as stocks [148]. At the Kelly-optimum the derivative should be zero:

$$\begin{aligned} \frac{d\Lambda(x)}{dx}|_{x^*} &= -(1-p) \cdot \frac{\Omega - 1}{\Omega(1-x^*) + x^*} + p \cdot \frac{1}{x^*} = 0 \Rightarrow \\ x^* &= p \cdot \frac{\Omega}{\Omega - 1} \end{aligned}$$

Hence for very large potential profit ($\Omega >> 1$), the optimal strategy is to maintain a safe fraction which is equal to the probability that you loose.

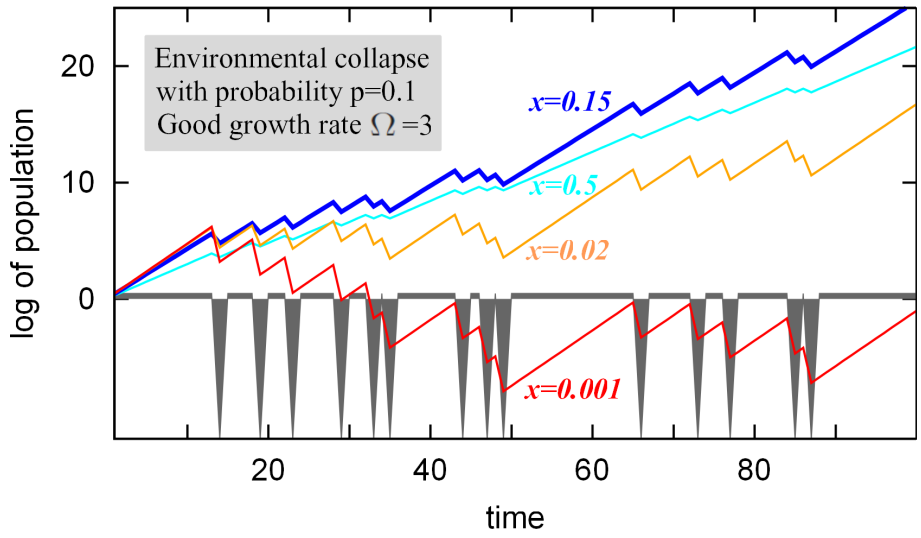


Figure 6.18: **Dynamics of capital.** One can play a game where one wins by factor $\Omega = 3$ by probability $1 - p = 0.9$, and looses all investment with probability $p = 0.1$. The blue curve is the growth of the Kelly-optimal strategy with a fraction $x^* = p \cdot \Omega/(\Omega - 1) = 0.15$ in the bank. The orange and red curves show sub-optimal strategies with $x = 0.01$ and $x = 0.001$. Conversely, the cyan trajectory simulates an over-cautious strategy with $x = 0.5$.

Put in practical use: Imagine that an investment agent suggest you a 12% per anuum investment for a 20 years investment. Your return after 20 years is the $\Omega \sim 10$ fold. You have to decide how big fraction of your capital you are going to invest. In practice $\Omega/(\Omega - 1) \sim 1$. Thus you should keep a bigger fraction of your capital in your bank account than the probability that the investment agent is a crook. If this probability is 50%, keep at least 50% of your money in a safe deposit.

- **Summary:** Hedge your bets

Questions

6.10 Consider a game where your invested money is multiplied by $\omega < 1$ when you lose and a factor $\Omega > 1$ when you win. Reconsider the above equations, and derive the optimal bet hedging strategy. Discuss the derived equation in the limit where $\Omega \gg 1$, $p \ll 1$ and $\omega \ll 1$.

Qlesson: In that limit then you should bet hedge with a fraction of the money equal the difference between the probability that things go bad, minus the loss when it goes bad. I.e. it is equal to the difference between a probability and a fraction.

6.11 Simulate the long time (500 updates) development of a capital that grows with rate $\Omega = 2$ during good times, but is exposed to catastrophic events with probability $p = 0.1$. In case of these events they lose everything invested. Simulate the development of initial capital of 1 when using the Kelly optimum value of x. Also simulate the development with other values of x, e.g. $x = 0.01$ and $x = 0.9$ and compare outcomes. Repeat simulation with finite disasters, say that bad events leads to reduction of invested fortune with a factor $\omega = 10^{-2}$, respectively $\omega = 0.5$.

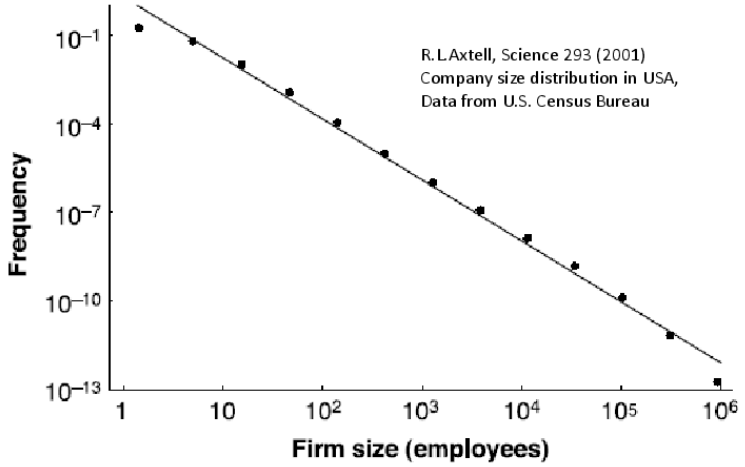


Figure 6.19: **Company size distribution in the USA.** This distribution exhibits a scale free behavior with exponent -2 (data from R.L. Axtel, 2001).

Hint: simulate the development of the log of the capital (where each event amount to addition or subtraction of the log of the change).
Qlesson: There is an optimum, but the gain with varying around that optimum is quite soft.

6.12) Consider the "Trimurti model" (Maslov and Sneppen, PLoS computational biology (2015)) based on:

- Exponential growth, $(dC_i/dt \propto C_i)$
- Finite world, $(\sum C_i < 1)$
- Bad things happen, $(C_i \to 0)$

That is consider N=100 companies with sizes C_1, C_2,...C_{100} where $\sum C_i = 1$. At each step set one random company to size a very small value $C_i = \gamma = 0.00001$, and then rescale all companies such that their sum becomes equal to 1. The last step correspond to collapse of a company, and start-up of a new (small) company. Investigate development of company sizes with time. Simulate average company size distribution. Repeat model with the modification that smaller companies have slightly larger collapse rate $(p(C \to \gamma) \propto C^{-0.2})$
Qlesson: catastophes on single company scale transcends into system wide collapses/revolutions.

6.13) Explore investment schemes to try making a profit in the above markets. Try it on a market with 20 companies, and $\gamma = 0.001$. Try to optimize the scheme, eventually including bet-hedging.

6.14) Daniel Bernoulli (1738) proposed a simple model for speculation of real markets, based on a quit-or-double game, where agents are eliminated once he reaches zero wealth and new agents enter the system with fortune $s = 1$ (keep one agent in the system at all times). The time average wealth distribution of this quit-or-double

game is obtained by iterating fortunes $s \to 2 \times s$, respectively $s \to 1$ with equal probability. Calculate analytically this distribution and compare with Fig. 6.19. Hint: The probability to reach fortune $s = 2^j$ or more is the probability to win at least j subsequent games, i.e. $(1/2)^j$.
Qlesson: One gain obtain a unfair wealth distribution from a fair game.
6.14) Simulate the quit-and-double game for a society with 1000 agents. What is the survival time distribution? Hint: One could equivalently simulate one agent for many time-steps where each collapse set agent to size 1. Then a sample of his fortune over a long sequence of situations with be identical to the 1000 agents, because they are anyway non-interacting.
Qlesson: 2^{-t}. Very short lifetime.

Lessons:

- Persistent walks is captured by Hurst exponent$H > 1/2$.
- Stocks are more correlated when they fall, than when they increase in value.
- Economic bubbles is a challenge to the assumption of optimally acting agents, rather indicating positive feedback in social systems.
- Bet hedging is a way to deal with unknown future, and is associated to the fact that 50% downturn and a 50% upturn does not balance, i.e. $0.5 \cdot 1.5 < 1$.

Supplementary reading:

Farmer, J. Doyne, Eric Smith, and Martin Shubik. "Economics: the next physical science?." arXiv preprint physics/0506086 (2005).

Mantegna, Rosario N., and H. Eugene Stanley. Introduction to econophysics: correlations and complexity in finance. Cambridge university press, 1999.

Peters, Ole. "The ergodicity problem in economics." Nature Physics 15.12 (2019): 1216-1221.

Bibliography

[1] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[2] Ulli Wolff. Collective monte carlo updating for spin systems. *Physical Review Letters*, 62(4):361, 1989.

[3] Helmut G Katzgraber. Introduction to monte carlo methods. *arXiv preprint arXiv:0905.1629*, 2009.

[4] Stephen G Brush. History of the lenz-ising model. *Reviews of modern physics*, 39(4):883, 1967.

[5] WP Wolf. The ising model and real magnetic materials. *Brazilian Journal of Physics*, 30(4):794–810, 2000.

[6] E Ao Guggenheim. The principle of corresponding states. *The Journal of Chemical Physics*, 13(7):253–261, 1945.

[7] Benoit B Mandelbrot. The variation of certain speculative prices. In *Fractals and scaling in finance*, pages 371–418. Springer, 1997.

[8] Taisei Kaizoji, Stefan Bornholdt, and Yoshi Fujiwara. Dynamics of price and trading volume in a spin model of stock markets with heterogeneous agents. *Physica A: Statistical Mechanics and its Applications*, 316(1-4):441–452, 2002.

[9] MPM Den Nijs. A relation between the temperature exponents of the eight-vertex and q-state potts model. *Journal of Physics A: Mathematical and General*, 12(10):1857, 1979.

[10] Bernard Nienhuis. Exact critical point and critical exponents of o (n) models in two dimensions. *Physical Review Letters*, 49(15):1062, 1982.

[11] FY Wang and ZG Dai. Self-organized criticality in x-ray flares of gamma-ray-burst afterglows. *Nature Physics*, 9(8):465, 2013.

[12] V. Pareto. La legge della domanda. *Giornale degli Economisti*, 10(59-68):691–700, 1895.

[13] G. U. Yule. A mathematical theory of evolution, based on the conclusions of dr. j. c. willis. *Phil. Trans. R. Soc. L. B*, 213:2187, 1924.

[14] H. A. Simon. *On a Class of Skew Distribution Functions*. *Biometrika*, 42:425–440, 1955.

[15] G. K. Zipf. *Human Behavior and the Principle of Least Effort.* Addison-Wesley, Cambridge, Massachusetts, 1949.

[16] Takashi Hara and Gordon Slade. Mean-field critical behaviour for percolation in high dimensions. *Communications in Mathematical Physics*, 128(2):333–391, 1990.

[17] Greg Huber, Mogens H Jensen, and Kim Sneppen. Distributions of self-interactions and voids in (1+ 1)-dimensional directed percolation. *Physical Review E*, 52(3):R2133, 1995.

[18] Lene Oddershede, Peter Dimon, and Jakob Bohr. Self-organized criticality in fragmenting. *Physical review letters*, 71(19):3107, 1993.

[19] K. Jacobs. *Stochastic Processes fro Physicists.* Cambridge University Press, 2010.

[20] N. Eldredge. *Life Pulse, Episodes from the story of the fossil record.* Facts on File Publications (New York), New York, 1987.

[21] J. J. Sepkoski. Ten years in the library: new data confirm paleontological patterns. *Paleobiology*, 19:43–51, 1993.

[22] S. Bornholdt, K. Sneppen, and H. Westphal. "longevity of orders is related to the longevity of their constituent genera rather than genus richness.". *Theory in Biosciences:*, 2009.

[23] L. W. Alvarez. Mass extinctions caused by large solid impacts. *Physics Today*, pages 24–33, 1987.

[24] K. J. Kauffman, P. Prakash, and J. S. Edwards. Advances in flux balance analysis. *Current Opinion in Biotechnology*, 14:491–496, 2003.

[25] P. Bak and K. Sneppen. Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.*, 71:4083, 1993.

[26] K. Sneppen, P. Bak, H. Flyvbjerg, and M. H. Jensen. *Evolution as a Self-Organized Critical Phenomenon. Proc. Natl. Acad. Sci. USA*, 92:5209–5213, 1995.

[27] N. Eldredge and S. J. Gould. Punctuated equilibriua: An alternative to phyletic gradualism. In T. J. M Schopf, J. M. Thomas, and S. Francisco, editors, *Models in Paleobiology.* Freeman and Cooper, 1972.

[28] S. J. Gould and N. Eldredge. Punctuated equilibrium comes of age. *Nature*, 366:223–227, 1993.

[29] G. G. Simpson. *Tempo and Mode in Evolution.* Columbia Univ. Press, New York, 1944.

[30] G. G. Simpson. *The Major Features of Evolution.* Columbia Univ. Press, New York, 1953.

[31] H. Flyvbjerg, K. Sneppen, and P. Bak. Mean field model for a simple model of evolution. *Phys. Rev. Lett.*, 71:4087, 1993.

[32] J. de Boer, B. Derrida, H. Flyvbjerg, A. D. Jackson, and T. Wettig. Simple model of self-organized biological evolution. *Phys. Rev. Lett.*, 73:906–909, 1994.

[33] K. Sneppen. *Extremal dynamics and punctuated co-evolution*. *Physica A*, 221:168, 1995.

[34] M. Paczuski, S. Maslov, and P. Bak. Avalanche dynamics in evolution, growth, and depinning models. *Phys. Rev. E*, 53:414–443, 1996.

[35] Y. G. Jin et al. Pattern of marine mass extinction near the permina-triassic boundary in south china. *Science*, 289:432–436, 2000.

[36] A. Trusina, M. Rosvall, and K. Sneppen. Communication boundaries in networks. *Phys. Rev. Lett.*, 94:238701, 2004.

[37] J. B. Axelsen, S. Bernhardsson, and K. Sneppen. One hub-one process: A tool based view on regulatory network topology. *BMC Systems Biology*, 2:25, 2008.

[38] P. Erdös and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:1760, 1960.

[39] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118, 2001.

[40] R. D. Luce and A. D Perry. A method of matrix analysis of group structure. *Psychometrika*, 14:95–116, 1949.

[41] P. W. Holland and S. Leinhardt. *Transitivity in structural models of small groups*. *Comparative Group Studies*, 2:107–124, 1971.

[42] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:409–410, 1998.

[43] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40:35–41, 1977.

[44] L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1:215–239, 1978/1979.

[45] M. W. Hahn and A. D. Kahn. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol*, 22:603–806, 2005.

[46] M. P. Joy, A. Brock, D. E. Ingber, and S. Huang. High-betweenness proteins in the yeast protein interaction network. *J. Biomed. Biotechnol.*, 2005:96–103, 2005.

[47] H. Yu, P. M. Kim, E. Sprecher, V. Trifanov, and M. Gerstein. The importance of bottlenecks in prootein networks: Correlation with gene essentiality and expression dynamics. *Plos. Comp. Biol.*, 3:713–720, 2007.

[48] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA*, 99:7821–7826, 2002.

[49] K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen. Modularity and extreme edges of the interet. *Phys. Rev. Lett*, 90:148701, 2003.

[50] K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen. Modularity and extreme edges of the interet. *Physica A*, 336:163, 2004.

[51] M. Rosvall and C. T. Bergstrøm. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. USA*, 105:1118–1123, 2008.

[52] H. Rieger J. D. Noh. Random walks on comples networks. *Phys. Rev. Letters*, 92:118701, 2004.

[53] M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27:39–54, 2005.

[54] A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science,*, 509:286, 1999.

[55] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabasi. The large scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

[56] Shang-Keng Ma. *Statistical Mechanics*. World Scientific Publishing Co Inc, 1985.

[57] P. Minnhagen and S. Bernhardsson. The blind watchmaker network: Scale-freeness and evolution. *PLoS ONE*, 3:e1690, 2008.

[58] R. Cohen, K. Erez, D. ben Avraham, and S Havlin. "*resilience of the inttener to random breakdowns.*". *Phys. Rev. Lett*, 85:4626–4628, 2000.

[59] K. Christenseni, R. Donangelo, B. Koiler, and K. Sneppen. Evolution of random networks. *Phys. Rev. Letters*, 81:2380, 1998.

[60] M. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, 2002.

[61] M. E. J Newman. The structure and function of complex networks. *SIAM Rev*, 45:167–256., 2002.

[62] L. D. Valdez, C. Buono, P. A. Macri, and L. A. Braunstein. Social distancing strategies against disease spreading. *arXiv:*, 1308, 2009.

[63] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296:910, 2002.

[64] S. Maslov, K. Sneppen, and A. Zaliznyak. Detection of topological patterns in complex networks: correlation profile of the internet. *Physica A*, 333(529-540), 2004.

[65] R. Milo et al. Superfamilies of evolved and designed networks. *Science*, 303:1538–1542, 2004.

[66] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation of escherichia coli. *Nature Genetics*, 22(2002), 2002.

[67] R. Milo, S.S. Shen-Orr, S. Itzkovitz, N. Kashtan, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002.

[68] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA*, 100:11980–11985, 2003.

[69] A. Trusina, S. Maslov, P. Minnhagen, and K. Sneppen. Hierarchy and anti-hierarchy in real and scale free networks. *Phys. Rev. Lett.*, 92:178702, 2004.

[70] J. B Axelsen, S. Bernhardsson, M. Rosvall, K. Sneppen, and A. Trusina. Degree landscapes in scale-free networks. *Phys. Rev. E*, 74:036119, 2006.

[71] J. von Neumann. The general and logical theory of automata. In L. A. Jeffres, editor, *Cerebral Mechanics in behaviour - the Hixon symposium*, pages 1–31, New York, 1951. John Wiley and Sons.

[72] T. C. Schelling. *Models of Segregation. The American Economic Review,*, 59:488–493, 1969.

[73] S. Wolfram. Statistical mechanics of cellular automata. *Reviews of Modern Physics*, 55:601–644, 1983.

[74] D. Helbing I. Farkas and T. Vicsek. Simulating dynamical features of escape panic. *". Nature*, 407:487–490, 2000.

[75] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. USA*, 99:7280–7287, 2002.

[76] E. Bonabeau, M. Dorigo, and G. Theraulaz. Inspiration for optimization from social insect behaviour. *Nature*, 406:39–42, 2000.

[77] N. Wiener and A. Rosenblueth. The mathematical formulation of the problem of conduction of connected excitable elements, specifically the cardiac muscle. *Arch. Inst. Cardiol. Mex*, 16:205–265, 1946.

[78] M. Greenberg and S.P. Hastings. Spatial patterns for discrete models of diffusion in excitable media. *SIAM Journal of Applied Mathematics*, 54:515–523, 1978.

[79] Y. Iwasa et al. Allelopathy of bacteria in a lattice population: competition between colicin-sensitive and colicin-producing strains. *Evolutionary Ecology*, 12(7):785–802, 1998.

[80] M.S.Steinberg. Does differential adhesion govern self-assembly processes in histogenesis? equilibrum configurations and the emergence of a hierarchy among populations of embryonic stemm cells. *Journal of experimental Zoology*, 173:395–433, 1970.

[81] P. B. Armstrong. Cell sorting out: The self-assembly of tissues in vitro. *Information healthcare*, 24:119–149, 1989.

[82] Dave Donaldson. Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934, 2018.

[83] Marc Levinson. *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger-with a new chapter by the author*. Princeton University Press, 2016.

[84] Masahisa Fujita, Paul R Krugman, and Anthony J Venables. *The spatial economy: Cities, regions, and international trade*. MIT press, 2001.

[85] Paul Krugman. Increasing returns and economic geography. *Journal of political economy*, 99(3):483–499, 1991.

[86] Paul Krugman and Anthony J Venables. Globalization and the inequality of nations. *The quarterly journal of economics*, 110(4):857–880, 1995.

[87] Frances Cairncross. The death of distance: How the communications revolution will change our lives. 1997.

[88] Kenneth L Kraemer, Jennifer Gibbs, and Jason Dedrick. Impacts of globalization on e-commerce use and firm performance: A cross-country investigation. *The Information Society*, 21(5):323–340, 2005.

[89] Kim Sneppen and Stefan Bornholdt. Globalization in a nutshell. *Physical Review E*, 98(4):042314, 2018.

[90] Philip McCann. Transport costs and new economic geography. *Journal of Economic Geography*, 5(3):305–318, 2005.

[91] A. Turing. The chemical basis of morphogenesis. *Phil Trans B*, 237:37–72, 1952.

[92] K. Yanagita. *Kagyuko*. Toko Shoin, Tokyo, 1930.

[93] L. Lizana, N. Mitarai, K. Sneppen, and H. Nakanishi. Modeling the spatial dynamics of culture spreading in the presence of cultural strongholds. *Physical Review E*, 83:066116, 2011.

[94] K. Sneppen, A. Trusina, M.H. Jensen, and S. Bornholdt. A minimal model for multiple epidemics and immunity spreading. *Plos One*, 5:e13326, 2010.

[95] F. Uekermann and K. Sneppen. Spreading of multiple epidemics with cross immunization phys. *Rev. E*, 86:036108, 2012.

[96] I. D. Chase. Dynamics of hierarchy formation: the sequential development of dominance relation. *Behaviour*, 80:218–240, 1982.

[97] E. Bonabeau, G. Theraulaz, and J. L. Deneubourg. Phase diagram of a model of self-organizing hierarchies. *Physica A*, 217:373, 1995.

[98] M. Rosvall and K. Sneppen. Modeling self-organization of communication and topology in social networks. *Phys. Rev. E*, 74:16108, 2006.

[99] M. Rosvall and K. Sneppen. Reinforced communication and social navigation generate groups in model networks. *Phys. Rev. E*, 79:026111, 2009.

[100] H. Spencer. *The principles of psychology*. Longmans, London, 1855.

[101] Louise H Taylor, Sophia M Latham, and EJ Mark. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 356(1411):983–989, 2001.

[102] Mark EJ Woolhouse and Sonya Gowtage-Sequeria. Host range and emerging and reemerging pathogens. In *Ending the War Metaphor:: The Changing Agenda for Unraveling the Host-Microbe Relationship-Workshop Summary*, volume 192, 2006.

[103] Nathan D Wolfe, Claire Panosian Dunavan, and Jared Diamond. Origins of major human infectious diseases. *Nature*, 447(7142):279–283, 2007.

[104] R. Axelrod. The dissemination of culture - a model with local convergence and global polarization. *Journal of Conflict Resolution*, 41:203–226, 1997.

[105] K. Klemm, V. M. Eguiluz, R. Toral, and M. San Miguel. *Phys. Rev. E*, 67:045101, 2003.

[106] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social phenomena. *Rev. Mod. Phys*, 81:591–646, 2009.

[107] P. Clifford and A. Sudbury. *Biometrika*, 60:581–588, 1973.

[108] R. A. Holley and T. M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probab.*, 3:643–663, 1975.

[109] S. Galam. Minority opinion spreading in random geometry. *Eur. Phys. J. B*, 25:403–406, 2002.

[110] K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *Int. J. Mod. Phys. C*, 11:1157–1165, 2000.

[111] P. Chen and S. Redner. Majority rule dynamics in finite dimensions. *Phys. Rev. E*, 71:036101, 2005.

[112] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3:87–98, 2000.

[113] S. Huet, G. Deffuant, and W. Jager. *Advances in Complex Systems*, 11:529, 2008.

[114] F. Vazquez, P.L.Krapivsky, and S. Redner. Constrained opinion dynamics: freezing and slow evolution. *Journal of Physics A*, 36:L61–L68, 2003.

[115] H.A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–304, 1940.

[116] Louis Bachelier. *Théorie de la spéculation*. Gauthier-Villars, 1900.

[117] Ingve Simonsen and Kim Sneppen. Profit profiles in correlated markets. *Physica A: Statistical Mechanics and its Applications*, 316(1):561–567, 2002.

[118] Jens Feder. *Fractals*. Springer Science & Business Media, 2013.

[119] Rafal Weron. *Modeling and forecasting electricity loads and prices: a statistical approach*, volume 403. John Wiley & Sons, 2007.

[120] Ingve Simonsen. Volatility of power markets. *Physica A: Statistical Mechanics and its Applications*, 355(1):10–20, 2005.

[121] Mogens H Jensen, Anders Johansen, and Ingve Simonsen. Inverse statistics in economics: the gain–loss asymmetry. *Physica A: Statistical Mechanics and its Applications*, 324(1):338–343, 2003.

[122] Raul Donangelo, Mogens H Jensen, Ingve Simonsen, and Kim Sneppen. Synchronization model for stock market asymmetry. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(11):L11001, 2006.

[123] Mandelbrot BB. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

[124] Robert P Flood and Peter M Garber. Market fundamentals versus price-level bubbles: the first tests. *Journal of political economy*, 88(4):745–770, 1980.

[125] Eng-Tuck Cheah and John Fry. Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of bitcoin. *Economics Letters*, 130:32–36, 2015.

[126] Ch Baek and M Elbeck. Bitcoins as an investment or speculative vehicle? a first look. *Applied Economics Letters*, 22(1):30–34, 2015.

[127] W Brian Arthur. Competing technologies, increasing returns, and lock-in by historical events. *The economic journal*, 99(394):116–131, 1989.

[128] Jay R Ritter. Behavioral finance. *Pacific-Basin finance journal*, 11(4):429–437, 2003.

[129] Jianjun Miao. Introduction to economic theory of bubbles. *Journal of Mathematical Economics*, 53:130–136, 2014.

[130] Irving Fisher. The debt-deflation theory of great depressions. *Econometrica: Journal of the Econometric Society*, pages 337–357, 1933.

[131] Ayumu Yasutomi. The emergence and collapse of money. *Physica D: Nonlinear Phenomena*, 82(1-2):180–194, 1995.

[132] Raul Donangelo and Kim Sneppen. Self-organization of value and demand. *Physica A: Statistical Mechanics and its Applications*, 276(3-4):572–580, 2000.

[133] Stefan Bornholdt. Expectation bubbles in a spin model of markets: Intermittency from frustration across scales. *International Journal of Modern Physics C*, 12(05):667–674, 2001.

[134] Jean-Philippe Bouchaud. Crises and collective socio-economic phenomena: simple models and challenges. *Journal of Statistical Physics*, 151(3-4):567–606, 2013.

[135] Herbert Spencer. *Railway Morals & Railway Policy*, volume 65. Longman, Brown, Green & Longmans, 1855.

[136] Tony Yu-Chen Tsai, Yoon Sup Choi, Wenzhe Ma, Joseph R Pomerening, Chao Tang, and James E Ferrell. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science*, 321(5885):126–129, 2008.

[137] Sandeep Krishna, S Semsey, and Mogens Høgh Jensen. Frustrated bistability as a means to engineer oscillations in biological systems. *Physical biology*, 6(3):036009, 2009.

[138] Marco A Janssen and Wander Jager. Fashions, habits and changing preferences: Simulation of psychological factors affecting market dynamics. *Journal of economic psychology*, 22(6):745–772, 2001.

[139] Serge Galam and Annick Vignes. Fashion, novelty and optimality: an application from physics. *Physica A: Statistical Mechanics and its Applications*, 351(2-4):605–619, 2005.

[140] Kim Sneppen and Namiko Mitarai. Multistability with a metastable mixed state. *Physical review letters*, 109(10):100602, 2012.

[141] Garrett Hardin. The tragedy of the commons. *science*, 162(3859):1243–1248, 1968.

[142] Leo Tolstoy. War and peace. 1869. *War and Peace. 1869. English translation by Rosemary Edmonds, published in*, 1957.

[143] J.L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35:917–926, 1956.

[144] C. T. Bergstrøm and M. Lachman. Shannon information and biological fitness. *Information Theory Workshop, IEEE*, 0-7803-8720-1:50–54, 2004.

[145] E. Kussell, R. Kishony, N-Q. Balaban, and S. Leibler. Bacterial persistence a model of survival in changing environments. *Genetics*, 169:1807–1814, 2005.

[146] E. Kussell and S. Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309:2075–2078, 2005.

[147] S. Maslov and K. Sneppen. Well temperate phage. *preprint, arXiv:1308.1646*, 2013.

[148] Maslov S and Zhang Y-C. Optimal investment strategy for risky assets. *International Journal of Theoretical and Applied Finance*, 1:377–387, 1998.