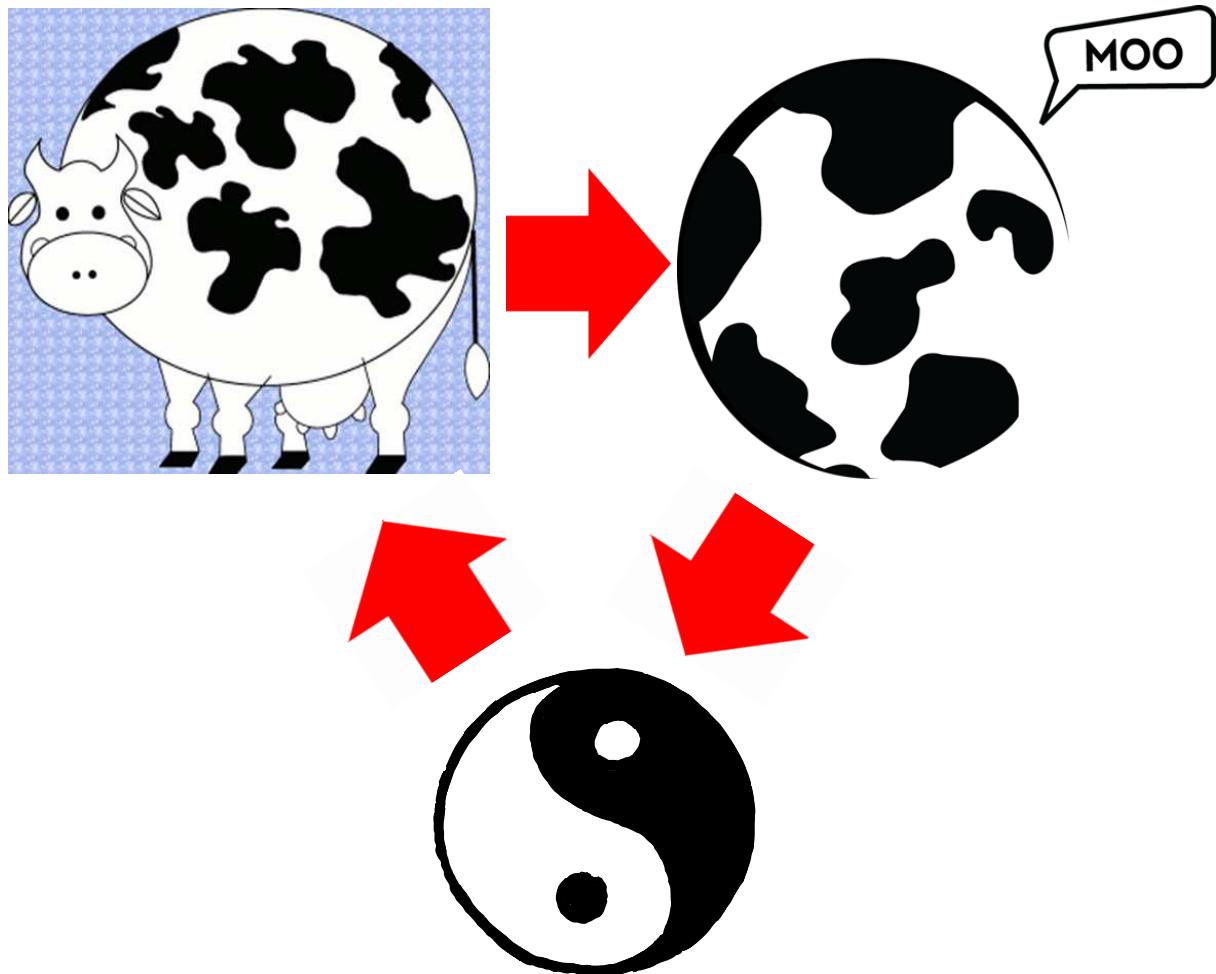


Complex Physics

Kim Sneppen and Jan O. Haerter

September 14, 2023



Contents

1 Statistical Mechanics and Importance Sampling	11
1.1 Entropy and Partition function	11
1.2 Definition of the Ising Model	23
1.2.1 Ferromagnetic and anti-ferromagnetic coupling	24
1.3 Monte Carlo method	26
1.3.1 Importance sampling	27
2 Phase Transitions & Criticality	37
2.1 Definition of phase transitions	37
2.1.1 First order vs. continuous phase transition	38
2.2 Mean field Solutions	45
2.2.1 Intuitive approach	45
2.2.2 Landau theory; “A Spherical Cow model”	53
2.2.3 Potts model.	55
2.3 1D Ising model	60
3 Percolation, Fractals & Fracture	71
3.1 Scaling in context	71
3.1.1 Universality:	75
3.2 Percolation	76
3.2.1 Percolation on a Bethe-lattice	79
3.2.2 Renormalization & Correlation length exponent	87
3.3 Fractals and Fractal Dimensions	91
3.3.1 Large objects with zero density	92
3.3.2 Fragmentation	98
3.4 Directed percolation	103
4 Self Organized Criticality	111
4.1 Introduction	111
4.2 Random walks	112
4.3 Critical branching process	118
4.4 Self Organized Criticality: The Sandpile Paradigm	121
4.5 Evolution as Self Organized Criticality	131

5 Networks	143
5.1 Introduction	143
5.1.1 When Networks are useful	143
5.1.2 Basic concepts	144
5.1.3 Amplification factor	148
5.1.4 Adjacency matrix	152
5.1.5 “Scale free” networks	157
5.1.6 Amplification of “epidemic” signals	158
5.2 Analyzing Network Topologies	162
5.2.1 Randomization: Constructing a proper null model	162
5.2.2 Algorithm generating a synthetic scale-free network	168
5.2.3 A hierarchy measure of networks	169
5.3 Models for Scale free networks	173
5.3.1 Preferential attachment	174
5.3.2 Merging and creation	178
5.4 Appendix: Generating function for the merging process	188
6 Stochastic Dynamics of Discrete Systems	191
6.1 Introduction	191
6.2 Direct Stochastic simulation	192
6.2.1 Event based Stochastic Simulation	195
6.3 Persistently competing states	198
6.3.1 Voter model with cooperativity	200
6.3.2 Information spreading on social scales	205
6.4 Epidemics	209
6.4.1 SIR model	209
6.4.2 Agent based model with superspreading	213
7 Econophysics	225
7.1 Analysis of a Time Series	225
7.2 Fear-Factor model	232
7.3 Modeling Economic Bubbles (Not Pensum)	236
7.4 Bet hedging	240
7.4.1 Bet hedging in random walk markets	240
7.4.2 Bet-hedging with occasional catastrophes	242
8 Dynamic Fronts	249
8.1 Burgers Equation	249
8.2 Noise-driven Interface Dynamics	254
8.2.1 Linear Interface Dynamics	254
8.2.2 Non-Linear Interface Dynamics	262
8.3 Deterministic Interface Dynamics with Chaos	267
8.4 Self-Organized Critical Interfaces	273

9 Diversity	281
9.1 The living world is rich in diversity	281
9.2 Balance versus Dominance	281
9.3 Emergence and Decline of Wrong Paradigms	287
9.4 Diversity from Cyclic Competition	292
9.5 Collapse & Exponential growth	297

Perspectives

Complexity is the science of systems of many interacting parts; systems where the whole is more than the simple sum of its parts. As stated by the condensed matter physicist Phil W. Anderson in a classical article from 1972 [1]:

More Is Different

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.

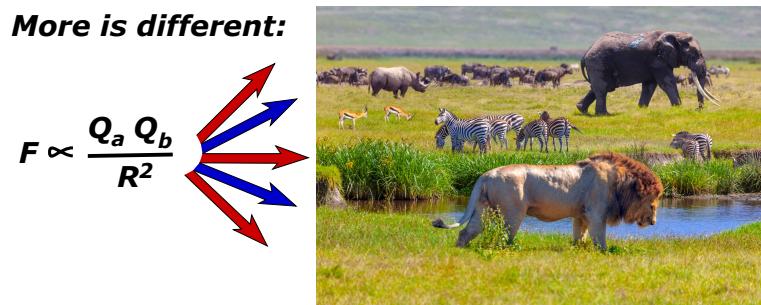


Figure 1: **Life is organized by electrical repulsion and attraction:** Illustration of the immense complexity and diversity that is obtained by repeated use of Coulombs law between charged particles. Apart from the overall gravity, no other fundamental forces are important.

Fractals repeat themselves at different 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 and chapter 2 provides an introduction to **critical phenomena**. Fundamental concepts from equilibrium statistical mechanics are first laid out and it is then discussed in the context of the Ising model. In particular it is 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 3 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 4 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 individual parts are either moving or completely at rest, a dynamics termed “**self-organized criticality**” can emerge, provided that time scales are infinitely separated.
- Chapter 5 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 6 teaches the student about **agent-based models**, stochastic event-based simulations and how to describe self-organization from individual to collective behavior.
- Chapter 7 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 8 (not included here) 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.
- Chapter 9 (not included here) takes on the philosophical interesting issue of Diversity. That is it present a few agent based models where many different types of agents compete for presence/existence.

... 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:

Quotations:

"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

"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

"Simplicity is the ultimate sophistication"

- Leonardo da Vinci

"Life is pretty simple: you do some stuff. Most fails. Some works. You do more of what works. If it works big, others quickly copy it. Then you do something else. The trick is the doing something else."

- Leonardo da Vinci

Chapter 1

Statistical Mechanics and Importance Sampling

1.1 Entropy and Partition function

Entropy of a system is measuring the number of questions needed to specify the state of a system completely, given the available knowledge we already have about the system. Formally the entropy counts the weighted size of the available phase space within bounds set by current knowledge:

$$S = -k_B \langle \ln(p_i) \rangle = -k_B \sum_{i \in \Omega} p_i \ln(p_i) \quad (1.1)$$

Here the sum runs over all states i in the available phase space and the Boltzmann constant $k_B = 1.38 \times 10^{-23} \text{ JK}^{-1}$ makes entropy counted in units of energy divided by temperature. The second law states that entropy in a closed system increases with time (or more formally, does not decrease). This reflects a probabilistic statement of a tendency for a decrease in available knowledge of a closed system with time. If S was simply measured in units of \log_2 , then $-\langle \log_2(p_i) \rangle$ would be the average number of yes-no questions needed to locate the system in its state i . That is, if it is in a state with low probability then one needs many questions, whereas a high probability can be located with few.

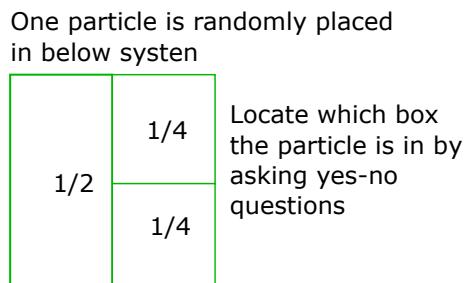


Figure 1.1: **Entropy-information:** Illustration of entropy as information needed to locate a particle in one of several sub-systems.

For example, consider a 3-state system with $p_1 = 0.5$ and $p_2 = 0.25$ and $p_3 = 0.25$ in Fig. 1.1. Asking first whether or not the system is in state 1 will give full localization using 1 question with probability 1/2. In case it was not in state 1, one needs to use one more question to locate it. Thus one asks if it is in state 2, and would have fully localized it whether or not this is the case. Thereby one needs 2 questions with probability 1/2 also, giving an average number of questions of 1.5.

The entropy in statistical physics is closely related to this information, just measured in terms of the Boltzmann constant for historical reasons.

The microcanonical ensemble considers a system in "Isolation" which means that neither particles nor heat enter it. Thus the total energy E will be conserved. Since energy can neither be added to nor left, the available phase space is confined to $\Omega(E)$. For an isolated system, equilibrium is obtained when each microstate within $\Omega(E)$ is equally likely. This typically needs time, as any given system typically starts in some non-equilibrium situation, and subsequently decays towards this ergodic limit.

Mini tutorial: When probabilities change from (1/4,1/4,1/2) to equidistributed (1/3,1/3,1/3) how much does entropy of systems decrease/decrease in units of \log_2 .

For an isolated system in an equilibrium then all states i are equally likely, $p_i = 1/\Omega$ and the entropy is simply given by the size of the phase space Ω :

$$S = k_B \cdot \ln(\Omega) \quad (1.2)$$

Noticeably, if all states are equally likely, the probability of being in a subspace Ω_s of the total phase space Ω is

$$p(\Omega_s) = \frac{\Omega_s}{\Omega} \quad (1.3)$$

and we could partition a system into sub-systems, and then consider the most likely way to partition phase space between the subsystems.

Consider now two independent systems, Fig. 1.2, except that we will allow the exchange of heat between them. Independence of systems 1 and 2 have a number of states given by the product, $\Omega = \Omega_1 \cdot \Omega_2$, the entropy becomes additive, $S = S_1 + S_2$.

The temperature is introduced by considering two containers 1 and 2 that are in thermal contact, so they can exchange energy. The total energy $E_{tot} = E_1 + E_2$ is thereby fixed, whereas the total entropy, $S_{tot} = S_1 + S_2$ should be maximized by transferring energy between 1 and 2. Hence, in equilibrium,

$$\frac{\partial S_{tot}}{\partial E_1} = \frac{\partial S_1}{\partial E_1} - \frac{\partial S_2}{\partial E_2} = 0, \quad (1.4)$$

or

$$\frac{\partial S_1}{\partial E_1} = \frac{\partial S_2}{\partial E_2} \equiv \frac{1}{T}, \quad (1.5)$$

which defines the temperature T , a quantity common to both containers. Note that the quantities T are independent of the size of the system and are therefore referred to as an *intensive* quantity.

Let us now consider the distribution of energy E_1 of a system in contact with a much larger system with energy E_2 . We further assume that these two systems are isolated and in contact with each other. Their total energy is

$$E = E_1 + E_2. \quad (1.6)$$

The total number of states for a constrained value of E_{tot} can be written as

$$\Omega(E) = \Omega_1(E_1) \cdot \Omega_2(E_2) \quad (1.7)$$

With the total entropy, $S_{tot}(E, E_1) = S_1(E_1) + S_2(E - E_1)$, the maximum of $S_{tot}(E, E_1)$ will again occur when the temperature in the two subsystems is equal.

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”. This ensemble assumes that each state i has the same energy and probability, thus variation of energy is not considered and all states count with equal weight, i.e. all states are equally likely as stated by maximizing the entropy.

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 of a system E_1 to be in equilibrium with the surroundings. The microstates are then occupied statistically, by the maximization of entropy for both the system and its surroundings under the total energy constraint.

Let us now assume that small deviations from the energy of a subsystem is possible. To this end, consider the probability for the subsystem to have energy E_1 (and thereby the surrounding larger system 2 to have energy $E_2 = E - E_1$):

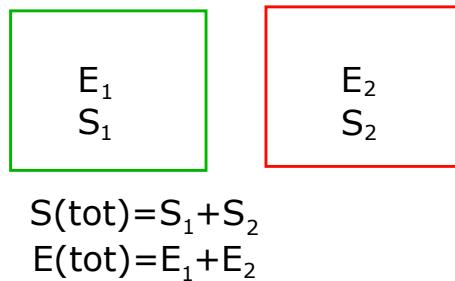


Figure 1.2: **Entropy of independent systems is additive** which can be used to derive the concept of temperature defined as the derivative of entropy with energy.

$$p(E_1) = \frac{\Omega_1(E_1)\Omega_2(E - E_1)}{\sum \Omega_1(E_1)\Omega_2(E - E_1)}. \quad (1.8)$$

With the definition of entropy we can rewrite $\Omega_2(E - E_1)$ has

$$\Omega_2(E - E_1) = \exp\left(\frac{S_2(E - E_1)}{k_B}\right). \quad (1.9)$$

As mentioned, the reservoir is assumed much larger than the subsystem, thus $E_1 \ll E$, and hence

$$S_2(E - E_1) = S_2(E) - E_1 \cdot \frac{\partial S_2}{\partial E_2} + \frac{1}{2}E_1^2 \cdot \frac{\partial^2 S_2}{\partial E_2^2} + \dots \quad (1.10)$$

Noting that the temperature of the system is

$$\frac{1}{T} = \frac{\partial S_2}{\partial E_2}, \quad (1.11)$$

we have that

$$p(E_1) = \text{const} \cdot \Omega_1(E_1) \cdot \exp\left(-\frac{E_1}{k_B T}\right) = \text{const} \cdot \exp\left(\frac{-E_1 + T \cdot S_1}{k_B T}\right)$$

with the $\exp(-E/k_B T)$ part expressing the Boltzmann distribution, i.e. an exponentially diminishing probability with increased energy to a smaller subsystem. In particular, if system 1 only has one state for each energy E_1 (a simple energy ladder), then

$$p(E_1) \propto \exp(-E_1/k_B T). \quad (1.12)$$

This exponential decrease accordingly emerged as a consequence of the cost in entropy S_2 of the surrounding system 2, as E_2 is decreased because the energy E_1 in sub-system 1 increases. This exponential form is important, also because it typically sets a scale of quantities in equilibrium physics. That is, in equilibrium, energy per degree of freedom is narrowly distributed and the

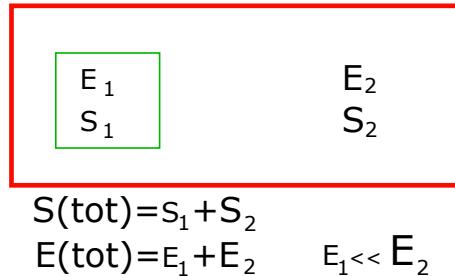


Figure 1.3: **Small sub system** is used to derive Boltzmann distribution because an increase in E_1 comes at the cost of entropy in the surrounding system S_2 .

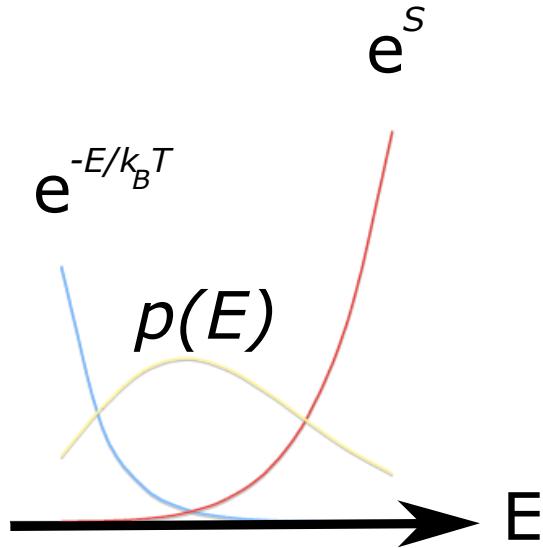


Figure 1.4: **Battle between low E and high S :** Typical behavior of contribution to partition function as a function of energy.

organization of a system into for example clusters also has a scale. The only exceptions are in situations where the entropy part marginally balances the energy part, which happens exactly at phase transitions that we will discuss later.

The central quantity is now the Partition function, defined as the weighted sum of the above states:

$$Z = \sum_i \exp(-\beta \epsilon_i) \quad (1.13)$$

where ϵ_i is the energy of a particular state and where we do not take into account that some states come with a degeneracy factor. Z is used to normalize the probabilities. Z also generalizes the size of the phase space, in the sense that it counts all states with their appropriate weight in the canonical ensemble. That is, a high ϵ_i state counts less because it limits the entropy of the surroundings of our system.

Mini tutorial: $\exp(-E/k_B T) = e^{-1} \cdot e^{-1} \cdot \dots \cdot e^{-1}$ in total $n = E/k_B T$ factors, each representing a factor $1/e^{1/2}$ for gaining one more energy for the considered degree of freedom relative to the probability to losing all its energy.

Taking into account the possible degeneracy of states $\Omega_1(E_1)$ within the system, we have:

$$p(E_1) = \frac{1}{Z} \Omega_1(E_1) \exp(-E_1/k_B T), \quad (1.14)$$

illustrated in Fig. 1.4. $p(E_1)$ is referred to as the *canonical distribution* and, now dropping all subscripts for the subsystem, the partition function becomes:

$$Z = \sum_i e^{S(E_i)/k_B} \cdot e^{-E_i/k_B T} = \sum_i e^{-F_i/k_B T} = \exp(-F/k_B T) \quad (1.15)$$

measuring the statistical weight of all energies i the system could be in with their appropriate degeneracy factor $\exp(S(E_i))$. This partition function replaces the simple counting of states in the microcanonical ensemble with a counting that weights each state with its contributions given how likely it will occur at temperature T in the canonical ensemble.

Z is statistical weight, can be partitioned in macrostates, for example:

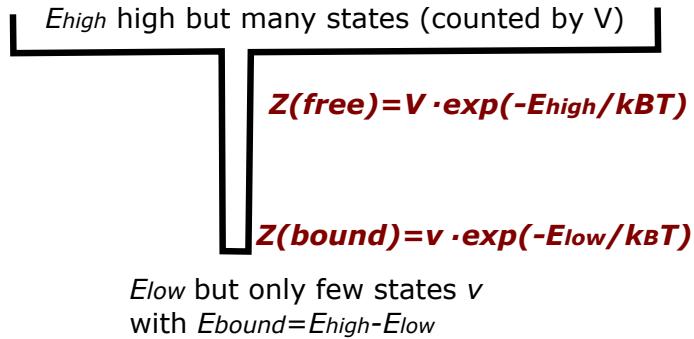


Figure 1.5: **Partitioned function** can be partitioned into parts that are relevant to the problem. Here the bound state can be assigned the weight (partition part) $Z_{\text{bound}} = v \cdot \exp(-E_{\text{low}}/k_B T)$ while the free part is assigned weight $Z(\text{free}) = V \cdot \exp(-E_{\text{high}}/k_B T)$. The important parameters are then V/v and $E_{\text{bound}} = E_{\text{high}} - E_{\text{low}}$, as only the relative size of the two parts of the partition sum matters.

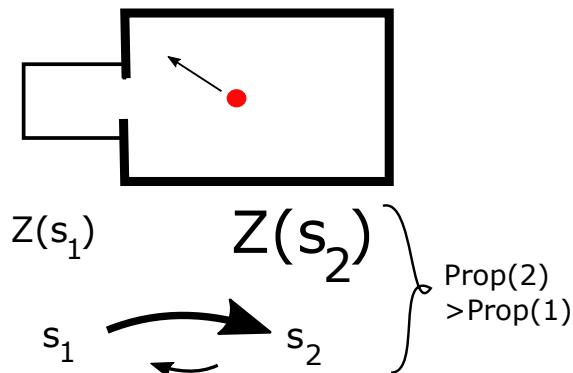
The above equation further defines the free energy of a system in the microcanonical ensemble in terms of the partition function:

$$F = -k_B T \ln(Z) \quad (1.16)$$

This Free energy in the canonical ensemble is the one that the system tends to minimize as one moves toward equilibrium.

Often Z may be partitioned into different contributions, for example by partitioning it into states with different energies, each with an energy, entropy, and free energy. One simple example is the bound states Z_b or free states Z_f (see Fig. 1.5). Here the bound states will have low E and relatively high weight per state, while the free states could be more numerous each with lower statistical weight. This battle between order and entropy is a part of all equilibrium phase transitions.

Mini tutorial: Write an equation that determines the temperature T where the V free states are equally likely to the one bound state in Fig. 1.5.



To balance flux, one may add ordered energy W in state 2 such that $Z(s_1) = Z(s'_1) = Z(s_2) \exp(-W/k_B T)$
 Thus $W = F_1 - F_2$, or difference in F can be turned to
 ordered energy W ... Free energy difference can be used

Figure 1.6: **Partition function** counts effective phase space, which in turn could be used to relate free energy to the possibility of extracting work (e.g. by moving a piston). In the figure, it is much more likely to move from small to big, than the opposite, and as a consequence, it is more likely to be in the big box.

The quantity F is called the free energy because it in principle can be extracted from the system, i.e. the system can perform work on the surrounding by reducing its internal free energy F to a lower state (by expansion for example). That is work can be performed by moving to a lower energy state, or to a higher entropy state (see Fig. 1.6 and questions). For Free energy then:

- Differences in Free energy is the available energy to do work
- The free energy never increases (for the isolated system at constant T)
- Free energy is minimal at equilibrium.

The conversion of the entropy part of the free energy to actual work on surroundings is illustrated in Fig. 1.7.

Where does the energy come from in the Szilard engine?

Qlesson: Statistical Mechanics builds on the concept of entropy maximization, requiring all states in the microcanonical ensemble to be equally likely. The canonical ensemble allows for fluctuations in energy, thereby assigning the Boltzmann probability to each state of a given energy. The canonical ensemble leads to the concept of free energy, which is minimized at equilibrium. A further useful quantity is the partition function, which allows one to compute all physical observables.

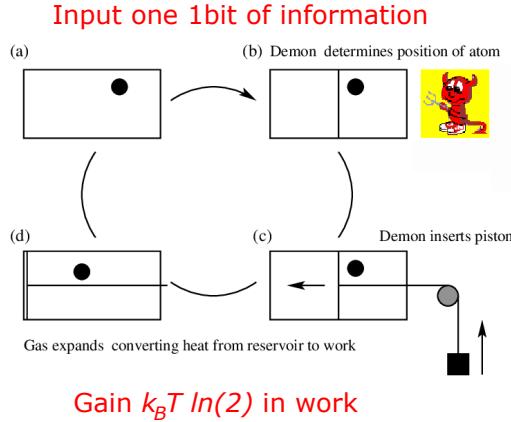


Figure 1.7: **Szilard Engine** illustrates how information can be used to change the entropy of a system, which subsequently can be used to extract energy from a system that exchanges heat with surroundings.

Average Energy and entropy from Partition function

The probability p_i for a state i is

$$p_i = \frac{1}{Z} \exp(-\beta \epsilon_i) . \quad (1.17)$$

where Z was the *partition function*. Z can be useful in expressing observables, e.g. the average energy

$$\langle E \rangle = \frac{\sum_i \epsilon_i \exp(-\beta \epsilon_i)}{Z} = -\frac{\partial}{\partial \beta} \ln Z = k_B \cdot T^2 \cdot \frac{\partial}{\partial T} \ln Z \quad (1.18)$$

and the entropy:

$$\begin{aligned} S &= -k_B \sum_i p_i \ln p_i = k_B \sum_i p_i (\ln Z + \beta \epsilon_i) \\ &= k_B \ln Z + \langle E \rangle / T = \frac{\langle E \rangle - F}{T} = -\frac{dF}{dT} \end{aligned} \quad (1.19)$$

where the last identity can be proven by just differentiating $F = -k_B T \ln(Z)$. 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).

Further, from energy one may calculate the specific heat

$$C \equiv \frac{\partial \langle E \rangle}{\partial T} = \frac{\partial \langle E \rangle}{\partial \beta} \frac{d\beta}{dT} = -\frac{1}{k_B T^2} \frac{\partial \langle E \rangle}{\partial \beta} = \frac{1}{k_B T^2} \frac{\partial^2 \ln Z}{\partial \beta^2} , \quad (1.20)$$

The specific heat can also be calculated from the entropy:

$$\begin{aligned}
 C &= \frac{\partial(F + TS)}{\partial T} \\
 &= \frac{\partial(-k_B T \ln(Z) + TS)}{\partial T} \\
 &= -k_B \ln(Z) - k_B T \frac{d \ln(Z)}{dT} + S + T \cdot \frac{dS}{dT} \\
 &= -\frac{F}{T} + k_B T \frac{d \ln(Z)}{d\beta} \frac{1}{k_B T^2} + S + T \cdot \frac{dS}{dT} \\
 &= \frac{F}{T} - \langle E \rangle \frac{1}{T} + S + T \cdot \frac{dS}{dT} = T \cdot \frac{dS}{dT}
 \end{aligned} \tag{1.21}$$

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”.

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

$$\frac{\partial \langle E \rangle}{\partial \beta} = -\frac{1}{k_B T^2} \frac{\partial}{\partial \beta} \frac{\sum \epsilon_n \exp(-\beta \epsilon_n)}{\sum \exp(-\beta \epsilon_n)} \tag{1.22}$$

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, which for the paramagnet from before, or the Ising model would be the external magnetic field h . Thus one can consider the average magnetization

$$M = - \left(\frac{\partial F}{\partial h} \right)_T, \tag{1.23}$$

and susceptibility (Fig. 1.6)

$$\chi_T = - \left(\frac{\partial^2 F}{\partial h^2} \right)_T, \tag{1.24}$$

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.)

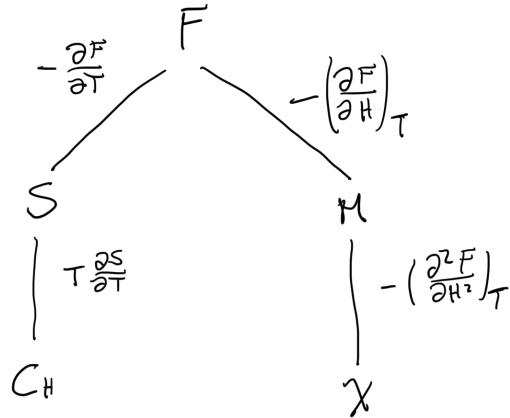


Figure 1.8: **Free energy and possible derivatives:** Here, a magnetic system is assumed, where M represents total magnetization, $H = h$ is the external magnetic field and χ is the magnetic susceptibility.

Paramagnet in microcanonical and canonical ensemble

For a simple example of a discrete system, consider the *paramagnet*, which is given by the Hamiltonian

$$\mathcal{H} \equiv -h \sum_{i=1}^N s_i, \quad (1.25)$$

where the spin variables $s_i = \pm 1$, thus they can take only one of two possible values. The parameter h denotes an external magnetic field. The paramagnet is a simple model for non-interacting spins within a highly anisotropic crystal, where each spin can only be oriented along one spatial direction.

Let us first consider the microcanonical ensemble, where we fix the total energy $E = -h \sum_{i=1}^N s_i$. To evaluate the number of possible configurations for the energy E note that energy E requires a net number of spins $E/h \equiv n$ spins to point "down," i.e., have negative sign. With

$$\begin{aligned} n &= N_\downarrow - N_\uparrow \\ &= N - 2N_\uparrow \end{aligned}$$

we have

$$\begin{aligned} N_\uparrow &= \frac{N-n}{2} \\ N_\downarrow &= \frac{N+n}{2}. \end{aligned}$$

We are now ready to write down the number of states in terms of N and n , i.e.,

$$\Omega(N, n) = \frac{N!}{(N - N_\uparrow)! N_\uparrow!} = \frac{N!}{\frac{N+n}{2}! \frac{N-n}{2}!}.$$

For large particle number N we can again make use of Stirling's approximation $x! \sim (x/e)^x$ to evaluate the entropy $S(N, n) = k_B \ln(\Omega)$:

$$\begin{aligned} \frac{S(N, n)}{k_B} &= N \ln N - N - \frac{N+n}{2} \ln \frac{N+n}{2} + \frac{N+n}{2} - \frac{N-n}{2} \ln \frac{N-n}{2} + \frac{N-n}{2} \\ &= N \ln N - \frac{N+n}{2} \ln \frac{N+n}{2} - \frac{N-n}{2} \ln \frac{N-n}{2}. \end{aligned}$$

The temperature corresponding to a given total energy E is therefore

$$\begin{aligned} \frac{1}{T} &= \frac{\partial S(N, E)}{\partial E} = \frac{\partial S(N, n)}{\partial n} \frac{\partial n}{\partial E} \\ &= \frac{k_B}{2h} \ln \frac{N+n}{N-n}. \end{aligned}$$

Re-arranging, we can solve for n as a function of T , namely,

$$\frac{n(T)}{N} = \frac{E(T)}{Nh} = \tanh \left(\frac{h}{k_B T} \right). \quad (1.26)$$

Let us check if this result makes intuitive physical sense: in the limit of large h , all spins will align and cause $n(T)$ to approach N . In the limit of $h = 0$, spins will orient randomly and there will be no net polarization $n(T)$. In the low-temperature limit for a given fixed h , again all spins will align as the argument of the tanh will be very large. In the high-temperature limit the argument is very small and spins will again orient at random.

Let us now repeat the above problem, using the framework of partition function in the canonical ensemble. The partition function now is a discrete sum over all possible configurations, that is, for N spins that each can take two possible configurations, there are 2^N distinct system configurations. We have

$$\begin{aligned} Z_C(T) &= \sum_{s_1=\pm 1, s_2=\pm 1, \dots, s_N=\pm 1} \exp \left(-\frac{h}{k_B T} \sum_{i=1}^N s_i \right) \\ &= \prod_{i=1}^N \left[\sum_{s=\pm 1} e^{-\frac{h}{k_B T} s} \right] = \left[e^{\frac{h}{k_B T}} + e^{-\frac{h}{k_B T}} \right]^N \\ &= 2^N \cdot \cosh \left(\frac{h}{k_B T} \right)^N. \end{aligned} \quad (1.27)$$

The variable s here serves to represent any of the spins and the factorization is possible since spins are independent. We will soon see more complicated cases, namely the Ising model (Sec. 1.2) where spins are no longer independent.

The energy expectation value is most easily evaluated by the following: since all spins are equivalent, the total energy is just $NE_j(T)$, where $E_j(T)$

is the energy corresponding to a given spin j . Evaluating the energy, we thus have

$$\begin{aligned}
 E(T) = N \cdot E_j(T) &= \frac{1}{Z_C(T)} \sum_{s_1=\pm 1, s_2=\pm 1, \dots, s_N=\pm 1} h \cdot s_j \exp\left(-\frac{1}{k_B T} \sum_{i=1}^N h \cdot s_i\right) \\
 &= Nh \cdot \frac{\sinh\left(\frac{h}{k_B T}\right) \cdot \cosh\left(\frac{h}{k_B T}\right)^{N-1}}{\cosh\left(\frac{h}{k_B T}\right)^N} \\
 &= Nh \cdot \tanh\left(\frac{h}{k_B T}\right), \tag{1.28}
 \end{aligned}$$

where the $\sinh()$ comes from inserting the values $s_j = +1$ and $s_j = -1$ in the sum for each spin. This expression matches the result obtained through the microcanonical ensemble (Eq. 1.26).

Questions

1.1 Entropy:

What is entropy per k_B , S/k_B for a system that can be in 3 states with respective probabilities $1/2$, $1/4$, and $1/4$? What is it if the probability is 0.99 , 0.005 and 0.005 ?

1.2. Partition function:

Consider Fig. 1.5, and calculate the energy E_{bound} that is needed to balance a volume V of the free state such that a particle is equally often in the bound or free state. How does E_{bound} scale with V and T . The volume V is in principle counted in terms of thermal wavelength, which discretizes the space into boxes given by Planck's constant and the temperature.

1.3 Win one, lose all with fifty-fifty probability:

Consider a game where you at each step win 1 unit with probability $1/2$, and lose all with probability $1/2$ (returning to zero units). Calculate the distribution of units and their average sampled over many steps.

1.4 Random energy distribution:

Consider 100 numbers along a line, all set all numbers =1 initially. Take 2 random numbers A and B randomly. Add the two together, and partition them to A' and B' such that $A' + B' = A + B$. Only partition in natural whole numbers, $0, 1, 2, \dots$. Iterate the above 1000000 times and plot the histogram of the obtained values of individual numbers.

1.5 Two electrons.

Consider two single-particle levels with energies $-\epsilon$ and ϵ . In these levels place two electrons (they do not interact). 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 $T \rightarrow \infty$. Hint: Use the definition equation for entropy with probabilities given by Boltzmann distribution.

1.6 Fluctuations.

Show that the fluctuations in the energy are related to the specific heat at constant volume by

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

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

1.2 Definition of the Ising Model

The Ising model is a lattice model where one variable is located at each site of a regular grid [2, 3]. The state of all variables is determined by a Hamiltonian, which couples the variables to a many-particle system. Such models have been successful in the description of critical phenomena, (quantum) magnetism and models for high-temperature superconductivity, as well as 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. In Sec. 1.3, 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, where each site i contains a variable s_i representing the magnetic dipole moment of an atomic spin or simply *spin*, which can be in one of two states,

$$s_i = \begin{cases} +1 & \text{"spin up"} \\ -1 & \text{"spin down"} . \end{cases} \quad (1.29)$$

The Hamiltonian of the Ising model is

$$\mathcal{H} \equiv -J \sum_{\langle ij \rangle} s_i s_j - h \sum_i s_i , \quad (1.30)$$

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 that interacts with the magnetic moment s_i . The magnetization is then defined as the system's macroscopic magnetic moment $M = \sum_i s_i$, that is, the sum over all individual spins.

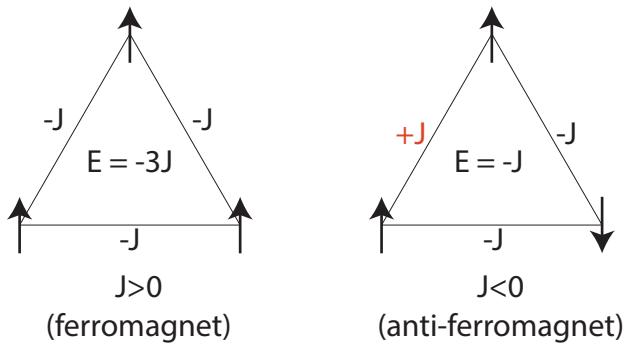


Figure 1.9: **Significance of the sign of coupling:** from order to frustration.

1.2.1 Ferromagnetic and anti-ferromagnetic coupling

It is important to note that the sign of J plays a crucial role in determining the magnetic ordering. 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. 1.2.1).

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 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. 2.3. 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. 2.2.

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

Qlesson: The Ising model is a fundamental paradigm for many interacting systems, not only ferromagnetic materials. It describes the interactions between spins as well as of each spin with an external magnetic field. Under certain conditions, the Ising model can exhibit a second-order phase transition, which can be described by critical exponents.

Questions

1.7 2-spin system Consider the 2 spins Ising system with Hamiltonian

$$H = -s_1 \cdot s_2 \quad (1.31)$$

where each spin can be either -1 or +1. What is the possible states of the system, and their associated energies? Calculate average energy as a function of temperature and give the value of entropy at $T \ll 1$ and at $T \gg 1$.

1.8 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 antiferromagnetic, zero-field spin-1/2 Ising model on a *triangular* lattice

$$\mathcal{H} = J \sum_{\langle ij \rangle} s_i s_j \quad (1.32)$$

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

Find the ground state for the ferromagnetic, zero-field spin-1/2 Ising model on a *triangular* lattice

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j \quad (1.33)$$

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

Consider the Ferromagnetic case, and estimate the critical temperature from one triangle with 3 spins that each can be either +1 or -1. The high

energy and high entropy state is where one spin doesn't match the two others. The low energy and low entropy state is where all 3 spins align. What is the critical value of T where the free energy of the two types of states becomes equal?

1.3 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 that 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, and also in areas far away from lattice models.

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. 1.2, 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 the magnetization per spin $\langle s_i \rangle$, the spin-spin correlation function $\langle s_i s_j \rangle$, or the average energy $\langle \mathcal{H} \rangle$, which are computed by

$$\begin{aligned}\langle s_i \rangle &= \langle s \rangle = \frac{\sum_{\{s\}} s_i \exp(-\beta \mathcal{H})}{\sum_{\{s\}} \exp(-\beta \mathcal{H})} , \\ \langle s_i s_j \rangle &= \frac{\sum_{\{s\}} s_i s_j \exp(-\beta \mathcal{H})}{\sum_{\{s\}} \exp(-\beta \mathcal{H})} , \text{ and} \\ \langle \mathcal{H} \rangle &= \frac{\sum_{\{s\}} \mathcal{H} \exp(-\beta \mathcal{H})}{\sum_{\{s\}} \exp(-\beta \mathcal{H})} .\end{aligned}$$

where the sum \sum runs over all states of the entire system, labeled by $\{s\}$, and where \mathcal{H} is a function of all spins in the system. One could hence imagine simply summing over all configurations and obtaining 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 $J/k_B T \sim 1$. 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 (number of neighbors of each site. The factor 1/2 come from the spin assignment to be +1/2 or -1/2).

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 than the latter, an enormous number even for modest N .

1.3.1 Importance sampling

Now the idea is to sample phase space by importance. This can be thought of as a *Markov chain*, where, in principle, each state of the system can be visited in finite time. That is, the moves we define should be such that the whole of phase space in principle should be visited on an infinitely long walk. In this sense, the transition rules should be such that the walk should be ergodic.

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.

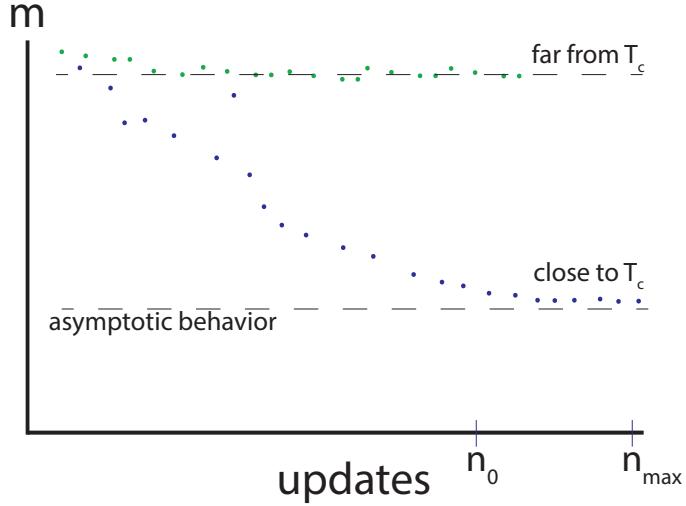


Figure 1.10: **Cartoon of possible Monte Carlo timeseries.** Points show system averages of magnetization m after various system updates for a Monte Carlo simulation with fixed T and parameters J and H , n_0 and n_{max} indicate system updates between which the asymptotic behavior is estimated. A “system 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.

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

$$P_l(t + 1) = P_l(t) \left(1 - \sum_{m \neq l} w_{l \rightarrow m} \right) + \sum_{m \neq l} P_m(t) w_{m \rightarrow l}, \quad (1.34)$$

where $w_{i \rightarrow j}$ 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 $w_{l \rightarrow l} = 1 - \sum_{m \neq l} w_{l \rightarrow m}$, i.e. the probability to remain in configuration l . Note also that $\sum_m w_{l \rightarrow m} = 1$, hence compactly

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

We check that probabilities are normalized correctly:

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

For the stationary solution we therefore have

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

hence

$$\sum_m [P_l^{eq} w_{l \rightarrow m} - P_m^{eq} w_{m \rightarrow l}] = 0. \quad (1.35)$$

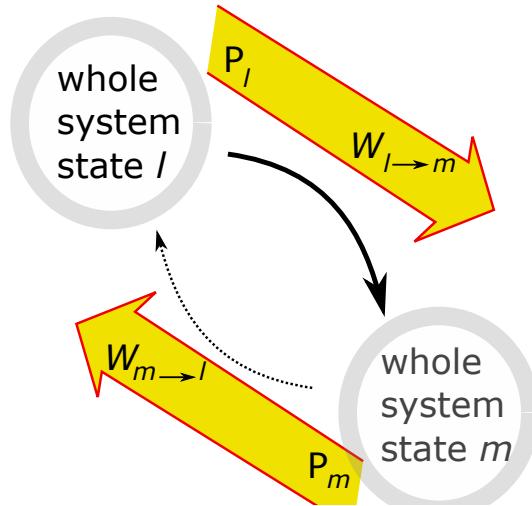


Figure 1.11: **Detailed Balance:** *The flux from l to m is equal to the flux from m to l where l and m are states of the total system. Thus there is a local (detailed) balance between states.*

A simple way to achieve the condition in Eq. 1.35 is to ensure that every term vanishes, that is,

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

This is called *detailed balance*, as it ensures that probability to go from state l to state m is the same as the probability that the system develops in the opposite way (see Fig. 1.3.1). Thus it states that there is no net flux across any links between any pair of states.

$$\frac{w_{l \rightarrow m}}{w_{m \rightarrow l}} = \frac{P_m^{eq}}{P_l^{eq}} = e^{-\beta(E_m - E_l)} \equiv e^{-\beta\Delta E}. \quad (1.36)$$

This is a very important result, as it provides a ratio that no longer requires any knowledge of the partition function Z itself (Z drops out by taking the ratio of the two probabilities P_m^{eq} and P_l^{eq}). If one hence can make use of Eq. 1.36 to define the system equilibrium, the partition function no longer needs to be evaluated.

Requirements for importance sampling

Notably, Eq. 1.36 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.

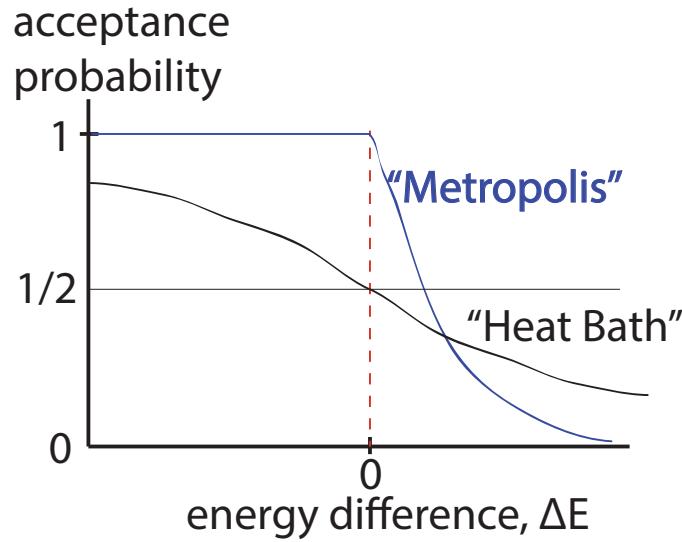


Figure 1.12: Choices for Monte Carlo update function $F(x)$.

A further requirement is *ergodicity*, which in this context means that each state in principle could be visited during a Monte Carlo simulation. In practice, it is sometimes hard to prove that ergodicity is fulfilled. Further, 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. 1.35) is a sufficient starting point to ensure importance sampling. But is it strictly necessary? If so, why? If not, which other (looser) condition would be sufficient? Can you provide a qualitative example of a hypothetical condition, perhaps a schematic illustration?

Sampling Variations

We make the ansatz that the transition probabilities should be a function of only energy differences, i.e. that $w_{l \rightarrow m}$ should be a function of $e^{-\beta \Delta E}$:

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

by symmetry hence

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

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

The detailed balance requirement then implies that

$$\frac{w_{l \rightarrow m}}{w_{m \rightarrow l}} = \frac{F(x)}{F(1/x)} = x, \quad (1.37)$$

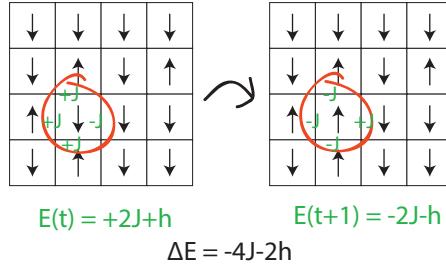


Figure 1.13: **Performing the Monte Carlo step.** Example of a spin flip by which the change of energy is negative for $J > 0$ and $h > 0$. In the Metropolis algorithm, this step would always be accepted. Imagine going the opposite direction: The energy difference would be $\Delta E = +4J + 2h$. The 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.

Furthermore, 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 [4, 5, 6, 7]. This means that one always accepts a suggested change in state that has lower energy, but only accepts an increase in energy with a probability given by the Boltzmann factor (see Figure).
2. $F(x) = \frac{x}{1+x}$, heat bath algorithm.
3. $F(x) = \frac{1}{2} (1 - \tanh(\beta \Delta E / 2))$, 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 $F(x)$?

Practical implementation on a computer.

The simulation is done using a lattice, assigning some spins, and then guessing on a transient time n_0 and a max simulation time n_{max} . The choice on n_0 can be tested by following energy or average magnetization as a function of time until it stabilizes. The n_0 should be after this stabilization, and typically we would then use $n_{max} \gg n_0$.

The basic setup 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, compute $r \equiv e^{-\Delta E/k_B T}$ (*compare* Eq. 1.36), generate a uniformly-distributed real random number q within the interval $[0, 1]$, 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 $\langle A \rangle = \frac{1}{n_{max}-n_0} \sum_{n>n_0} A_n$. In practice, you could speed up the simulation by only averaging between some of the sampled states.

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. Approaches on how to choose the values of n_{max} and n_0 more systematically can be found in the literature [6, 7].

Fig. 1.3.1 shows an example of a Monte Carlo simulation of the spin- $\frac{1}{2}$ Ising model on a 400×400 lattice. This is a rather large system, and we recommend that your simulations uses smaller systems.

The “time-series” 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 $T_c \approx 2.27 J/k_B$, the approach is much slower and substantial temporal fluctuations remain. The spatial plots show the state of 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.

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. $\xi \sim |T_c - T|^{-\nu}$ with distance to the critical point, *see also* the discussion following Eq. 2.4), 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 *correla-*

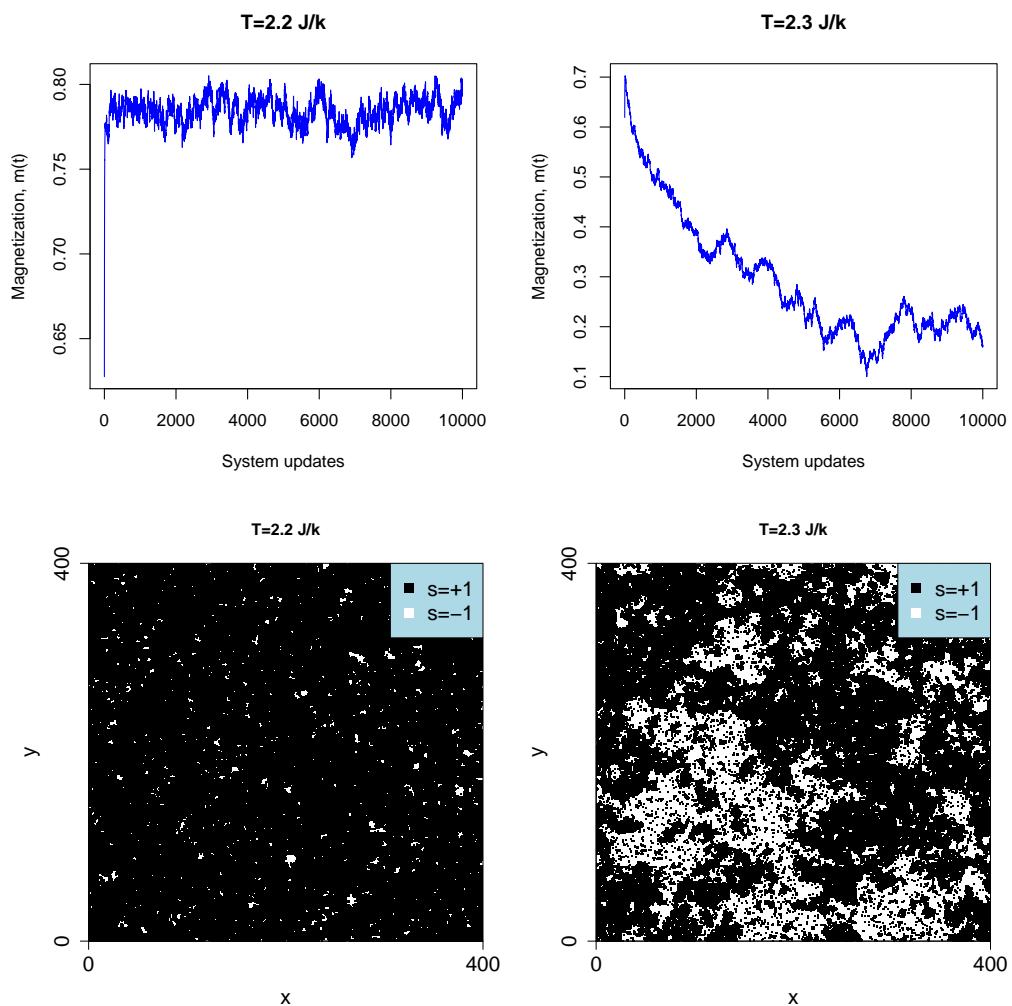


Figure 1.14: **Example of a simulation result.** Computation on a system of 400×400 lattice sites. Timeseries for average magnetization and corresponding spatial spin patterns for two distinct values of T as labeled.

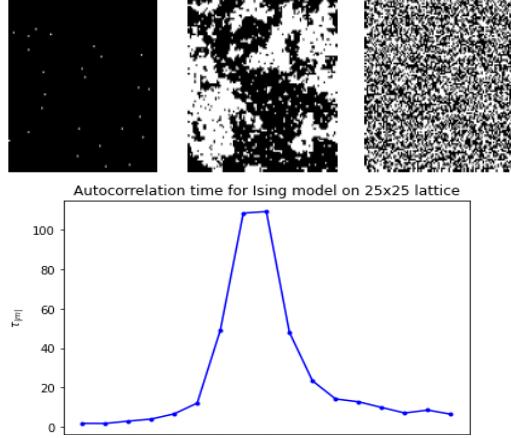


Figure 1.15: **Critical slowing down:** *Simulation time diverges slowest around critical points because spatial correlations are the largest, and therefore slowest to relax.*

tion time τ , which for an observable A can be defined as

$$\tau \equiv \frac{\int_0^\infty t \cdot [A(t) - \langle A \rangle] dt}{\int_0^\infty [A(t) - \langle A \rangle]^2 dt}, \quad (1.38)$$

can be shown to diverge with an exponent z with the size of the clusters, i.e. $\tau \sim \xi^z \sim |T_c - T|^{-z\nu}$, see also Fig. 1.3.1 The number z is 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 [8, 9]. A useful precaution to at least quantify the required sampling time is to measure the *autocorrelation* $C_A(t)$ as the simulation is running:

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

n and t 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, $C_A(t) \sim \exp(-t/\tau_{\text{auto}})$, hence τ_{auto} can be estimated and one can decide how many steps are required for a large enough, independent, sample. For more details on Monte Carlo methods, the reader is referred to the literature [9].

Mini Tutorial: One may in principle sample state space of an Ising model by just selecting random states and their contribution to both the partition

function and for example to the average energy. How likely is it to obtain the correct energy by this method for low temperature and a N=100 spin system?

Qlesson: Monte Carlo methods perform a random sampling of configuration space. By focusing the random sampling on the more likely parts of configuration space, these methods are much more efficient than enumerating all possible options. For equilibrium statistical mechanics systems several requirements have to be fulfilled by the Monte Carlo method, including "detailed balance," "ergodicity," and the assumptions on equilibrium, e.g. that the Boltzmann probabilities are respected.

Questions

1.9 Monte Carlo Simulation.

Consider a 2D spin- $\frac{1}{2}$ Ising model on a square lattice with external field $h = 0$. 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 and $k_B T$ 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 beyond which expectation values should be computed. Define also a function that computes the energy corresponding to a given spin configuration, i.e. $E = -J \sum_{\langle ij \rangle} s_i s_j$. Further, define a function that flips a (random or deterministic) spin, as well as the exponential $x \equiv \exp(-\Delta E/k_B T)$ with ΔE the energy difference between two states l and m and define the acceptance procedure for a transition $a_{l \rightarrow m}$.
2. Make a loop that iterates over the procedure. Starting from a random configuration, carry out the Monte Carlo simulations.
3. Compute the internal energy $\langle E \rangle$ and the magnetization per site $\langle s \rangle$ as function of $k_B T$, 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. 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 .

Lessons:

- Entropy is log of size of unknown state space Ω the system could be in:

$$S = k_B \ln(\Omega)$$

Here k_B is the Boltzmann constant that compensates for the historic “mistake” that temperature T is not measured in units of energy. Thus if a system could be in 2 states with same energy, then its entropy $S = k_B \cdot \ln(2)$.

- Basic stat mec: If a system can be in two states with energy E_1 and E_2 the partition function is $Z = \exp(-\beta E_1) + \exp(-\beta E_2)$, the probability to be in state 1 is $p_1 = \exp(-\beta E_1)/Z$, the average energy is $\langle E \rangle = p_1 E_1 + p_2 E_2$, and its entropy $S/k_B = -p_1 \cdot \ln(p_1) - p_2 \cdot \ln(p_2)$.
- Detailed balance: There exist probabilities $P(i)$ of all states i of the system such that net flux between any two states is zero

$$w(l \rightarrow k)P(l) = w(k \rightarrow l)P(k)$$

where w is transition probabilities.

- The Metropolis algorithm is about calculating averages by sampling along a trajectory formed by stepping in the state space of a system. A move to lower energy is always accepted. A move from lower $E_{present}$ to higher energy E_{new} is accepted with probability $p = \exp(-(E_{new} - E_{present})/k_B T) < 1$. The moves should fulfill detailed balance and ergodicity. The average is obtained by including all steps along the trajectory with equal weight.

Chapter 2

Phase Transitions & Criticality

2.1 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 conductor 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 $M(T) = \sum s_i$ (Fig. 2.1). A first-order transition has a discontinuity in its order parameter when traversing the transition (by varying pressure in the case of gas-liquid). A second-order transition has a discontinuity in the derivative of the order parameter when traversing the critical point.

In Sec. 1.2 we introduced the Ising model¹ as a description of magnetism in some magnetic materials. We remind ourselves of the Ising model Hamiltonian,

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - h \sum_i s_i , \quad (2.1)$$

where s_i is the “spin” at a lattice site i and can take one of the values ± 1 . The parameter 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.

We qualitatively sketch some limits (Fig. 2.3): 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.

¹First studied by Lenz and Ising in 1925, see Brush [10] and Wolf [11] for reviews on the model and its vast applications.

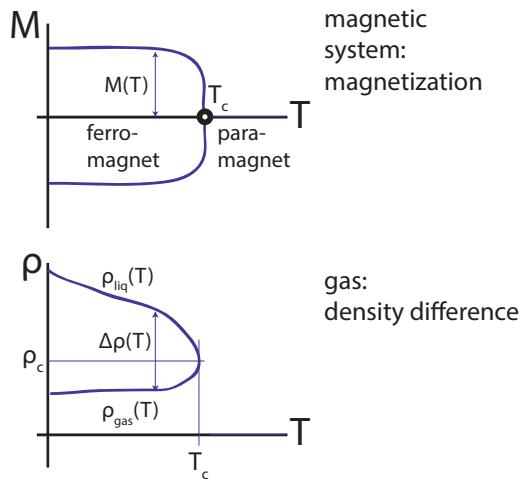


Figure 2.1: **Cartoon of order parameters.** Order parameters describe the state of the system in a reduced way. Order parameter for a zero-external-field ($H = 0$) magnetic system with symmetry regarding spin orientation (top) and a gas, where the order parameter can be defined as the difference in density for the liquid and gas phases.

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?

2.1.1 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. Spins are allowed to flip and create fluctuation around the average internal energy $\langle E \rangle$ — reflecting an exchange with heat and a canonical ensemble. Around $T \sim T_c$, the magnetization varies continuously, hence the transition is *not* of the 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). We will later study the Potts model, which exhibits a first-order transition.

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

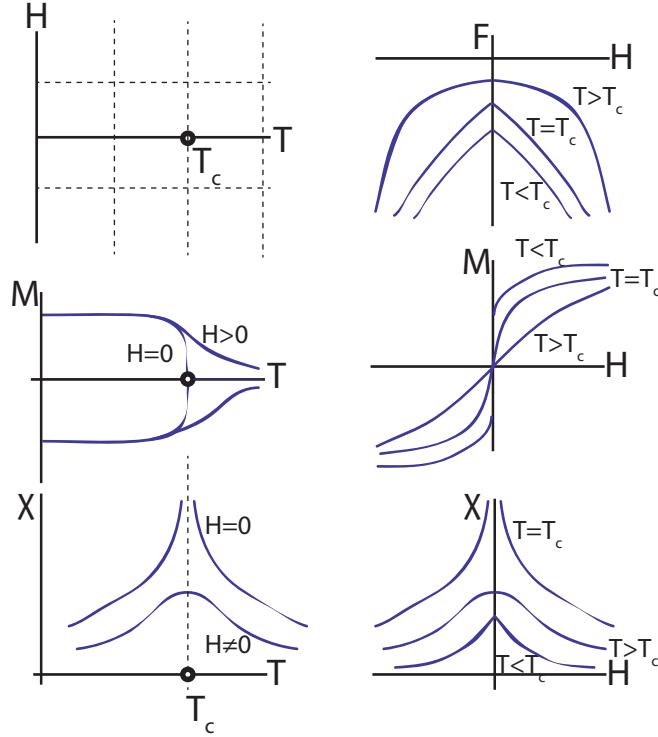


Figure 2.2: Dependencies of variables near the critical point of Ising model.

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

$$\Gamma(\mathbf{r}_i, \mathbf{r}_j) = \langle (s_i - \langle s_i \rangle)(s_j - \langle s_j \rangle) \rangle , \quad (2.2)$$

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

$$\Gamma(\mathbf{r}_i - \mathbf{r}_j) \equiv \Gamma_{ij} = \langle s_i s_j \rangle - \langle s \rangle^2 . \quad (2.3)$$

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

$$\Gamma(\mathbf{r}) \sim r^{-\tau} \exp(-r/\xi) , \quad (2.4)$$

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. $\xi \rightarrow \infty$ and

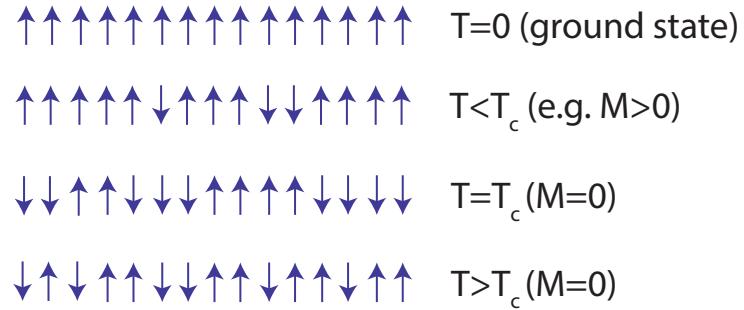


Figure 2.3: **Illustration of spin ordering.** Cartoon of spin ordering for the 2-d Ising model for increasing temperature (top to bottom).

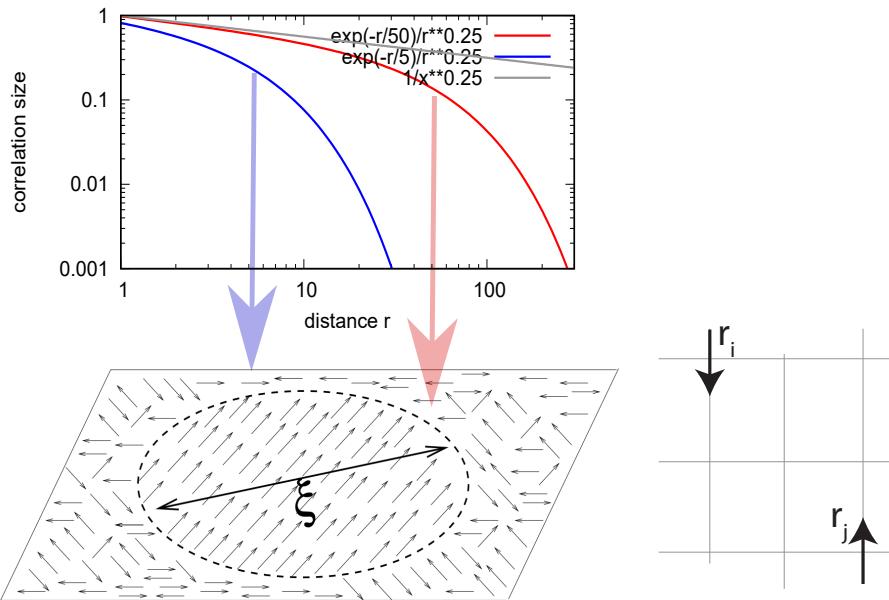


Figure 2.4: Spatial Correlations in 2-d Ising model: The correlation depends on distance with a power law, $\Gamma \propto 1/r^{0.25}$. However, this is only true up to a cutoff that is set by distance to critical temperature $T_c - T$. The red curve is a factor 10 closer to T_c than the blue one.

$\exp(-r/\xi) \rightarrow 1 = \text{const.}$ Indeed, experiments as well as some exactly soluble models show that near criticality $T \rightarrow T_c$, the value of the correlation function decays as a power law with distance, i.e.

$$\Gamma(\mathbf{r}) \sim \frac{1}{r^{d-2+\eta}}. \quad (2.5)$$

In this equation, $\eta = \nu + 2 - d = 0.25$ depends on some of the system properties and is an example of a so-called *critical exponent*. The cutoff function depends

on distance to critical point as

$$\xi \propto 1/(T_c - T) \quad (2.6)$$

thus halving the distance to critical point doubles the correlation length ξ .

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

$$\chi_T = k_B T \frac{\partial^2}{\partial h^2} \ln Z \quad (2.7)$$

$$= \frac{1}{k_B T} (\langle M^2 \rangle - \langle M \rangle^2) \quad (2.8)$$

$$= \frac{1}{k_B T} \langle (M - \langle M \rangle)^2 \rangle \quad (2.9)$$

$$= \frac{1}{k_B T} \langle \sum_i (s_i - \langle s_i \rangle) \sum_j (s_j - \langle s_j \rangle) \rangle \quad (2.10)$$

$$= \frac{1}{k_B T} \sum_{ij} \Gamma_{ij}, \quad (2.11)$$

where the total magnetization M was written as the sum over all spins ($M = \sum_i s_i$) and $Z = \sum_{states} e^{\beta(J \sum_{nn} s_i s_j + h \sum_i s_i)}$. This is hence a relation between spontaneous microscopic fluctuations in the system in equilibrium (described by the correlation function) and the response of the system as a whole to external perturbations as determined by the susceptibility — a macroscopic observable. 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:

$$N \sum_j \Gamma_{0j} \sim N \int dr \Gamma(r) r^{d-1} \sim \chi_T. \quad (2.12)$$

Hence, for correlations to remain (*compare* Eq. 2.5), 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?

Definitions of critical exponents

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

$$t \equiv \frac{T - T_c}{T_c}, \quad (2.13)$$

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 mean field	3-dimensional vector	-0.12	0.36	1.39	4.8	0.71	0.04
		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 2.1: **Critical exponents for several models.** For the 4-d ising model ref [12] estimated from simulations that $\beta = 0.5$, $\nu = 0.5$ and $\alpha = 0$.

termed "reduced temperature". The reduced temperature t , sometimes referred to as the "control parameter" or "tuning parameter," hence changes sign from negative to positive, when the absolute temperature is increased (or "tuned") from the origin and $t = 0$ when $T = T_c$. In terms of t a critical exponent is generally defined as the limit

$$\lambda = \lim_{t \rightarrow 0} \frac{\ln |F(t)|}{\ln |t|}, \quad (2.14)$$

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

Commonly used critical exponents are

$$C_H \sim |t|^{-\alpha} \text{ zero-field specific heat} \quad (2.15)$$

$$M \sim (-t)^\beta \text{ zero-field magnetization} \quad (2.16)$$

$$\chi_T \sim |t|^{-\gamma} \text{ zero-field isothermal susceptibility} \quad (2.17)$$

$$H \sim |M|^\delta sgn(M) \text{ critical isotherm } (t = 0) \quad (2.18)$$

$$\xi \sim |t|^{-\nu} \text{ correlation length} \quad (2.19)$$

$$\Gamma(\mathbf{r}) \sim \frac{1}{r^{d-2+\eta}} \text{ pair-correlation function at } T_c. \quad (2.20)$$

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 system. 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 1945 (Fig. 2.5). The diagram shown is the actually measured version of the one in Fig. 2.1. 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. 2.5)!

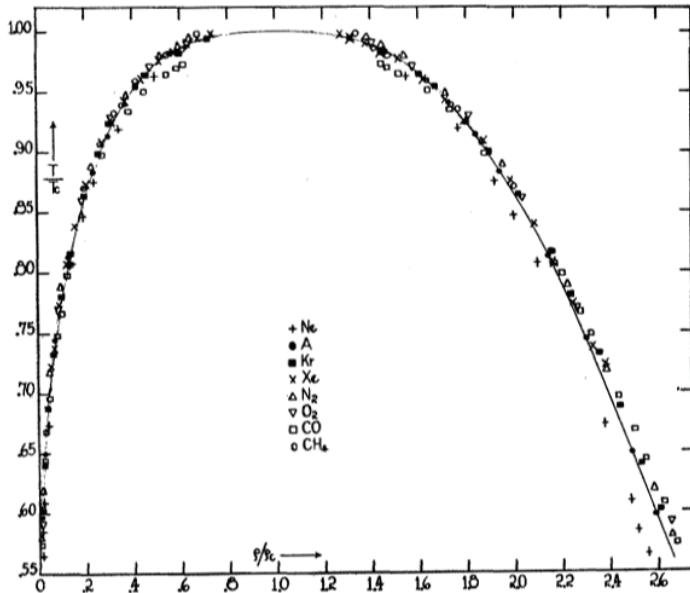


Figure 2.5: Measurement on eight fluids of the coexistence curve. The solid line is a fit to a cubic equation, i.e. to the choice $\beta = 1/3$, where $\rho - \rho_c \sim (-\epsilon)^\beta$ [13].

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. 2.1). 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.

Qlesson: Phase transitions are singularities in the free energy or one of its derivatives. Phase transitions can be quantified by an order parameter, e.g., magnetization for the ferromagnet, or density for the liquid-gas transition. For continuous phase transitions the singularity can be characterized by critical exponents.

Questions

2.1 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.

2.2 Paramagnet²

A paramagnetic solid contains a large number N of non-interacting, spin-1/2 particles on fixed lattice sites. This substance is placed in a uniform magnetic field $H = h$.

- (i) Write down an expression for the partition function of the solid, neglecting lattice vibrations, in terms of $x = h/k_B T$.
- (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) $M = N \tanh x$, $\chi = N/(k_B T \cosh^2 x)$, $S = Nk(\ln 2 + \ln(\cosh x) - x \tanh x)$).

2.3 Rushbrooke inequality. As the different observables depend on each other, then also, the corresponding critical exponents are related. Consider the specific heats at constant field H (or h) and constant magnetization M , respectively:

$$C_H \equiv T \left(\frac{dS}{dT} \right)_H , \quad (2.21)$$

$$C_M \equiv T \left(\frac{dS}{dT} \right)_M , \quad (2.22)$$

²Yeomans: Problem 2.2

as well as the magnetic susceptibility

$$\chi_T \equiv \left(\frac{\partial M}{\partial H} \right)_T . \quad (2.23)$$

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:

$$\chi_T(C_H - C_M) = T \left(\frac{\partial M}{\partial T} \right)_H^2 . \quad (2.24)$$

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

2.2 Mean field Solutions

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. 2.3). 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. 2.2).

In the following, we give two approaches.

2.2.1 Intuitive approach

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

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - h \sum_i s_i . \quad (2.25)$$

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. 2.25

$$\begin{aligned} s_i s_j &= (m + (s_i - m))(m + (s_j - m)) \\ &= m^2 + m(s_i - m) + m(s_j - m) + (s_i - m)(s_j - m) . \end{aligned} \quad (2.26)$$

Assuming that fluctuations are small (generally a poor assumption), we neglect the second-order fluctuation term (4th term in Eq. 2.26). This is the basis of the mean-field assumption, i.e. that expectation value of the product of a spin fluctuation with another spin fluctuation 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

$$E_{MF} = -J \sum_{\langle ij \rangle} (-m^2 + m(s_i + s_j)) - h \sum_i s_i . \quad (2.27)$$

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. 2.27 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.

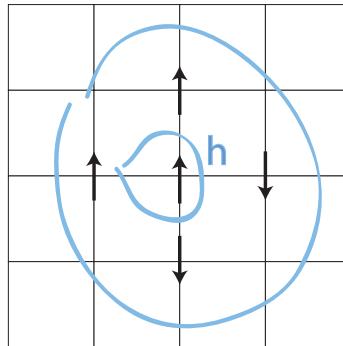


Figure 2.6: **Mean field assumption:** The surroundings of a given site act as an effective “mean field” estimated from the average across the lattice. Thereby correlations between individual sites are ignored. The equation is closed by assuming that the site itself contributes to the mean m field given by the mean of its neighbors.

Mean field partition function and critical temperature

We can now write the mean field partition function as

$$Z_{MF} = \sum_{\{s_j\}} \exp(-\beta E_{MF}) \quad (2.28)$$

$$\begin{aligned} &= \exp\left(-\beta \frac{NJzm^2}{2}\right) \sum_{s_1=\pm 1} \cdots \sum_{s_N=\pm 1} \prod_{j=1}^N \exp(\beta(Jzm + h)s_j) \\ &= \exp\left(-\beta \frac{NJzm^2}{2}\right) \left[\sum_{s_1=\pm 1} \exp(\beta(Jzm + h)s_1) \right] \cdots \left[\sum_{s_N=\pm 1} \exp(\beta(Jzm + h)s_N) \right] \\ &= \exp\left(-\beta \frac{NJzm^2}{2}\right) [e^{\beta(Jzm+h)} + e^{-\beta(Jzm+h)}] \cdots [e^{\beta(Jzm+h)} + e^{-\beta(Jzm+h)}] \\ &= \exp\left(-\beta \frac{NJzm^2}{2}\right) [2 \cosh(Jzm\beta + h\beta)]^N. \end{aligned} \quad (2.29)$$

It is now straightforward to compute the magnetization per site:

$$m = -\frac{1}{N} \frac{d \ln(Z)}{dh} = \frac{1}{N} \sum_{j=1}^N \langle s_j \rangle = \langle s_i \rangle \equiv \langle s \rangle,$$

where naturally value $\langle s_i \rangle$ for any site $i \in \{1, \dots, N\}$ is equal. This equation could also be obtained from exploring minima in Free energy $dF/dm = 0$ (or $d \ln(Z)/dm = 0$). Hence

$$\begin{aligned} m &= \frac{\exp\left(-\beta \frac{NJzm^2}{2}\right) [2 \cosh(Jzm\beta + h\beta)]^{N-1} [2 \sinh(Jzm\beta + h\beta)]}{\exp\left(-\beta \frac{NJzm^2}{2}\right) [2 \cosh(Jzm\beta + h\beta)]^N} \\ &= \tanh(Jzm\beta + h\beta). \end{aligned} \quad (2.30)$$

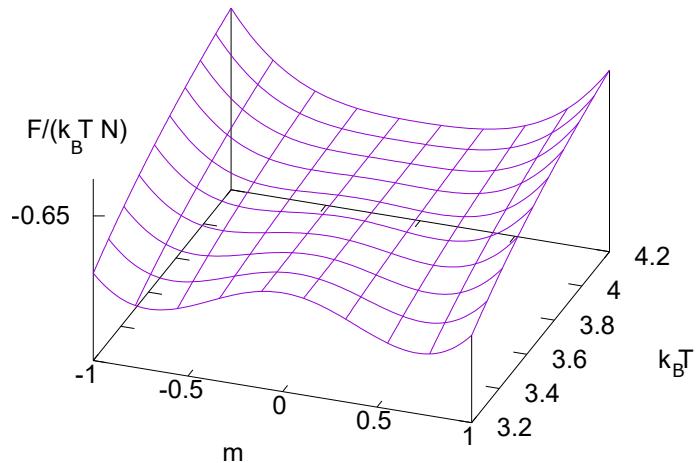


Figure 2.7: **Free energy in Mean field Ising model ($h=0$):** Illustration of the appearance of two solutions when $k_B \cdot T$ is below a critical temperature. Results for $J \cdot z = 4$

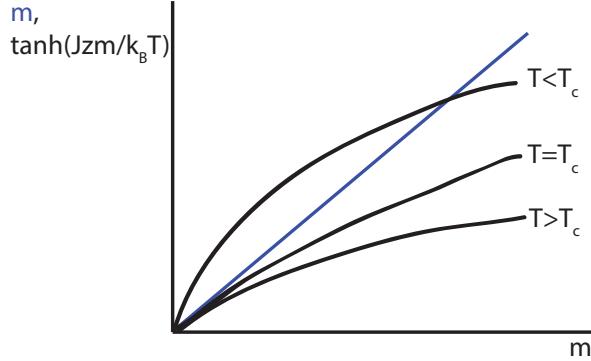


Figure 2.8: **Self-consistency condition:** Curves of $\tanh(Jzm/k_B T)$ for various values of T as well as the curve m .

This expression deserves a comment: the expectation value of one of the spins was computed by weighting the two terms in the expression

$$e^{\beta(Jzm+h)} + e^{-\beta(Jzm+h)}$$

by the two values of spin, ± 1 , respectively, yielding a factor of

$$(+1) e^{\beta(Jzm+h)} + (-1) e^{-\beta(Jzm+h)} = 2 \sinh(Jzm\beta + h\beta)$$

in the numerator. Notably, Eq. 2.30 is a self-consistent implicit expression for m : first, we had assumed that the average magnetic field produced by the neighbors of any given spin i can be described by m , second, based on this assumption we computed the corresponding partition function, and finally, from this partition function we computed the expectation value of the spin at site i , $\langle s_i \rangle$. The self-consistent expression (Eq. 2.30) thus completes this line of thought.

Can Eq. 2.30 produce singular behavior near the critical temperature T_c ? To investigate this, we first require the value of T_c , which is obtained at zero external magnetic field, that is, setting $h = 0$. For $h = 0$, Eq. 2.30 yields the simplified self-consistent equation,

$$m = \tanh(Jzm\beta). \quad (2.31)$$

It is not possible to isolate m from this equation to have a closed-form solution. However, if we plot both sides of Eq. 2.31 on the same graph, we are able to appreciate possible solutions. Both m and $\tanh(Jzm\beta)$ are sketched in Fig. 2.2.1 for several values of temperature. Inspecting the different curves, note that Eq. 2.31 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 cases will hence occur at the *critical* $\beta_c \equiv (Jz)^{-1}$. Alternatively, the mean-field critical temperature is

$$T_c \equiv \frac{Jz}{k_B} .$$

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 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. That is, the mean field approximation works as if the dimension was infinite. Intuitively this reflects the fact that the distances across the lattice decreases with dimension, making the basic assumption about the neighbors being representative of the whole more accurate.

Mini Tutorial: There are two solutions to Eq. 2.31, where $h = 0$. What is the difference between the two? What would happen to these two, if $h \neq 0$?

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. 1.1). 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

$$t \equiv \frac{T - T_c}{T_c} . \quad (2.32)$$

The mean field free energy can be computed from Eq. 2.29 by taking the logarithm of the mean field partition function Z_{MF} , yielding

$$f_{MF} = -\frac{k_B T}{N} \ln Z_{MF} = \frac{Jzm^2}{2} - k_B T \ln [2 \cosh(Jzm\beta + h\beta)] . \quad (2.33)$$

It is now useful to work in dimensionless units by dividing Eq. 2.33 through by zJ and introducing the rescaled quantities $h' \equiv h/zJ$ as well as $\theta = T/T_c = k_B T/Jz$. For later use, note that in these units $t = \theta - 1$. We then have

$$\begin{aligned} \frac{f_{MF}}{zJ} = \frac{F}{zJ \cdot N} &= \frac{m^2}{2} - \frac{T}{T_c} \ln 2 - \frac{T}{T_c} \ln \left(\cosh \left(\frac{T_c(m+h')}{T} \right) \right) \\ &= \frac{m^2}{2} - \theta \ln 2 - \theta \ln \cosh \left(\frac{m+h'}{\theta} \right) \end{aligned} \quad (2.34)$$

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

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.

The exponent β (not to be confused with resiprok temperature) describes the temperature dependence of magnetization $m \sim (-t)^\beta$ in the vicinity of T_c , i.e. in the limit of $t \rightarrow 0$. 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. 2.2.2), 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. 2.34 in a Taylor series, yielding

$$\frac{f_{MF}}{zJ} = \frac{m^2}{2} - \theta \ln 2 - \theta \ln \left[1 + \frac{1}{2} \frac{m^2}{\theta^2} + \frac{1}{24} \frac{m^4}{\theta^4} + \mathcal{O}\left(\frac{m^6}{\theta^6}\right) \right]. \quad (2.35)$$

Using the additional expansion $\ln(1+x) = x - \frac{x^2}{2} + \dots$ the final logarithm in Eq. 2.35 simplifies to yield

$$\begin{aligned} \frac{f_{MF}}{zJ} &= \frac{m^2}{2} - \theta \ln 2 - \theta \left[\frac{1}{2} \frac{m^2}{\theta^2} - \frac{1}{12} \frac{m^4}{\theta^4} + \mathcal{O}\left(\frac{m^6}{\theta^6}\right) \right] \\ &= \frac{m^2}{2} \left(1 - \frac{1}{\theta} \right) - \theta \ln 2 + \frac{m^4}{12\theta^3} + \mathcal{O}\left(\frac{m^6}{\theta^6}\right). \end{aligned} \quad (2.36)$$

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

$$\frac{1}{zJ} \frac{\partial f_{MF}}{\partial m} = m \left(1 - \frac{1}{\theta} \right) + \frac{1}{3} \frac{m^3}{\theta^3} = 0. \quad (2.37)$$

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

$$\frac{1}{zJ} \frac{\partial^2 f_{MF}}{\partial m^2} = \left(1 - \frac{1}{\theta} \right) + \frac{m^2}{\theta^3}. \quad (2.38)$$

Apart from the solution $m = 0$ Eq. 2.37 leads to

$$m^2 = 3(1 - \theta)\theta^2 = 3(-t)\theta^2,$$

where we have introduced the reduced temperature $t \equiv (T - T_c)/T_c$. Hence, for small $|t|$ but $t < 0$ the Eq. 2.39 has the two solutions

$$\begin{aligned} m(t) &= \pm \sqrt{3} \theta |t|^{1/2} = \pm \sqrt{3 \cdot \frac{T^2}{T_c^2} \cdot \frac{T_c - T}{T_c}} \quad \text{for } T < T_c \\ m(t) &= 0 \quad \text{for } T > T_c \end{aligned} \quad (2.39)$$

We now check the second derivative (Eq. 2.38) for any of the real solutions and find that for $T < T_c$ the second derivative is positive for nonzero m and negative for $m = 0$. For $T > T_c$ 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 the isothermal susceptibility: $\chi_T \sim |t|^{-\gamma}$ (*compare* Eq. 2.17). To compute the susceptibility

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

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. 2.30, namely

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

where we use that critical temperature is $\beta_c Jz = 1$. First, let us Taylor expand Eq. 2.40 to linear order in h , to yield

$$m = \tanh(\beta Jzm) + h\beta(1 - \tanh^2(\beta Jzm)), \quad (2.40)$$

and expand the hyperbolic tangent out as:

$$\tanh(x) = x - \frac{x^3}{3} + \mathcal{O}(x^5). \quad (2.41)$$

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. 2.40, we obtain

$$m = \beta Jzm - \frac{1}{3}(\beta Jzm)^3 + h\beta - h\beta(\beta Jzm)^2. \quad (2.42)$$

When applying the derivative w.r.t. h on both sides of the equation and taking the limit $h \rightarrow 0$, we have (Notice that the last term gives two contributions, where the one proportional to h vanishes as $h \rightarrow 0$):

$$\begin{aligned} \chi_T &= \frac{\partial m}{\partial h} \Big|_{h=0} \\ &= \beta Jz\chi_T - (\beta Jz)^3 m^2 \chi_T + \beta - \beta(\beta Jz)^2 m^2 \end{aligned} \quad (2.43)$$

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

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

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

$$\lim_{t \rightarrow 0, t < 0} \chi_T = \frac{1}{k_B T_c} \frac{1}{(-2t)} . \quad (2.44)$$

Super-critical case: $T > T_c$. In Eq. 2.43 it is now appropriate to disregard the terms involving m^2 , since, after taking the derivative regarding h , the limit $h \rightarrow 0$ is taken and m takes its zero-field value, namely $m = 0$. Hence, we then have

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

which yields

$$\lim_{t \rightarrow 0, t > 0} \chi_T = \frac{1}{k_B T_c} \frac{1}{t} = \frac{1}{J z t} . \quad (2.45)$$

From Eqs 2.44 and 2.45 it is now easy to read off the critical exponent γ corresponding to the temperature dependence of χ_T near the critical temperature. Given the susceptibility critical exponent definition ($\chi_T \sim |t|^{-\gamma}$, compare Eq. 2.17), in both cases,

$$\gamma = +1 .$$

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

Mean-field critical exponent for 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, $h = 0$). 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} and the variation in magnetization $m(t)$ for $T < T_c$ (i.e. $\theta < 1$):

$$\begin{aligned} \frac{f_{MF}}{zJ} &= \frac{3t\theta^2}{2} \left(1 - \frac{1}{\theta} \right) - \theta \ln 2 + \frac{9t^2\theta^4}{12\theta^3} \\ &= -\frac{3(\theta-1)^2\theta}{2} - \theta \ln 2 + \frac{3(\theta-1)^2\theta}{4} = -\frac{3(\theta-1)^2\theta}{4} - \theta \ln 2 \end{aligned}$$

we are now in a position to proceed with entropy S and specific heat c_H as derivatives of the free energy, i.e.

$$S = -\frac{\partial f_{MF}}{\partial T} = -\frac{\partial f_{MF}}{\partial \theta} \frac{\partial \theta}{\partial T} = \frac{zJ}{T_c} \left(\frac{3}{4}(3\theta^2 - 4\theta + 1) + \ln 2 \right)$$

Thus with $zJ = k_B T_c$ then S/k_B becomes a number without units and only a function of θ . The specific heat becomes (remember that $d\theta/dT = 1/T_c$):

$$c_H = T \cdot \frac{\partial S}{\partial T} = \theta \cdot k_B \cdot \frac{3}{2} (3\theta - 2) \sim \frac{3}{2} k_B \text{ at } \theta \sim 1$$

Since $c_H \sim \text{const}$ close to T_c (does not diverge with $\theta \rightarrow 1$), 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 $\alpha_{MF} = 0$.

Question

2.4) Calculate numerically the specific heat for the mean-field version of the Ising model (at $h = 0$). Determine the order of the phase transition. Please remember that the average energy is given in terms of the partition function as $\langle E \rangle = -\frac{d\ln(Z)}{d\beta}$, which can be expressed analytically in terms of m at each value of temperature. Calculate m numerically from the self consistently equation $m = \tanh(\beta J m)$ by Newton method to the difference function $Func = m - \tanh() = 0$. Then calculate the specific heat numerically by $C = d\langle E \rangle / dT$.

2.2.2 Landau theory; “A Spherical Cow model”

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

$$\mathcal{F}(m) = \mathcal{F}_0 + a_2 m^2 + a_4 m^4 + \text{higher order terms} \quad (2.46)$$

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. 2.35), obtained from the microscopic partition function, yields a similar temperature dependence as Eq. 2.46 (consider question 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} :

$$\frac{d\mathcal{F}(m)}{dm} = 0 = 2\tilde{a}_2tm + 4a_4m^3 = m(2\tilde{a}_2t + 4a_4m^2),$$

yielding $m = 0$ and

$$|m| = \sqrt{t\tilde{a}_2/2a_4}, \quad (2.47)$$

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 .

$$\begin{aligned} C_H &= \frac{\partial \langle E \rangle}{\partial T} = \frac{1}{k_B T^2} \frac{\partial^2 \ln Z}{\partial \beta^2} = -\frac{1}{k_B T^2} \frac{\partial^2(\beta F)}{\partial \beta^2}, \\ &= -\frac{1}{k_B T^2} \frac{\partial}{\partial \beta} \left(F + \beta \frac{\partial}{\partial \beta} F \right) = -\frac{1}{k_B T} \left(2 \frac{\partial F}{\partial \beta} + \frac{\partial^2 F}{\partial \beta^2} \right) \end{aligned}$$

and focusing on divergence in double differentiation, and use that T is proportional to $t = 1 - T/T_c$, and $d/d\beta = (t^{-2}/T_c) \cdot d/dt$.

Using Eq. 2.47 in the free energy (Eq. 2.46), we have for $t < 0$

$$\mathcal{F} = \mathcal{F}_0 - \frac{\tilde{a}_2^2 t^2}{4a_4} + \mathcal{O}(t^3), \quad (2.48)$$

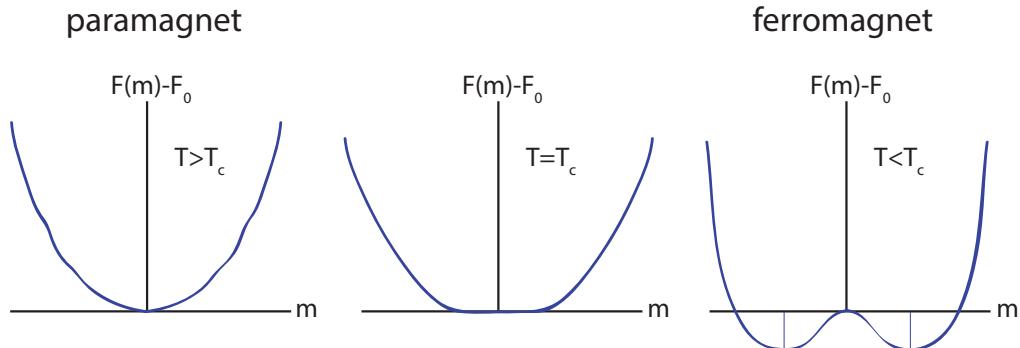


Figure 2.9: **Landau free energy for different values of temperature.** Zero external field ($h = 0$). For $T > T_c$ and $T = T_c$ the minimum of the free energy is located at $m = 0$. For $T < T_c$, there are two minima that are symmetrically located at finite magnetization.

hence, the specific heat tends to a constant as $t \rightarrow 0^-$. 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

$$\mathcal{F} = \mathcal{F}_0 - hm + \tilde{a}_2 tm^2 + a_4 m^4, \quad (2.49)$$

Minimizing w.r.t. m gives

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

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).

Qlesson: Mean-field theory assumes an overall magnetization that is "felt" by each spin in the lattice, i.e., the spin "sees" only an average field that it interacts with. The concept of pairwise interaction is lost, leading to shortcomings in describing spin-spin correlation functions.

2.2.3 Potts model.

One simple extension of the Ising model are so-called Potts model, (R.B. Potts. "Some Generalized Order-Disorder Transformations". Mathematical Proceedings. 48, 106 (1952)). In this model, s_i can take a number q that can be more than two values. Energy is only assigned when neighboring sites have the same value, i.e.

$$\begin{aligned} \epsilon(i, j) &= -\delta_{(s_i - s_j)}, \text{with Hamiltonian} \\ \Rightarrow H &= -\sum_{nn} \delta(s_i - s_j) \end{aligned} \quad (2.50)$$

where δ_x specifies the delta function which is unity when $x = 0$ position i and j have same state s . Such Potts models can describe e.g. opinion dynamics in a population, where "agreement" of opinion could cause a negative value of energy. The Potts model qualitatively depends on the number of possible spins q : As q increases the possible entropy associated with non-aligned neighbors increases, and the system may undergo a first-order phase transition. For $q = 2$ the Potts model is equivalent to the Ising model (see F.Y. Wu, (1982)).

Mean field of Potts Model

The Potts model includes more states for each spin, thereby increasing the entropy of the disordered state. The larger the number of states q the more

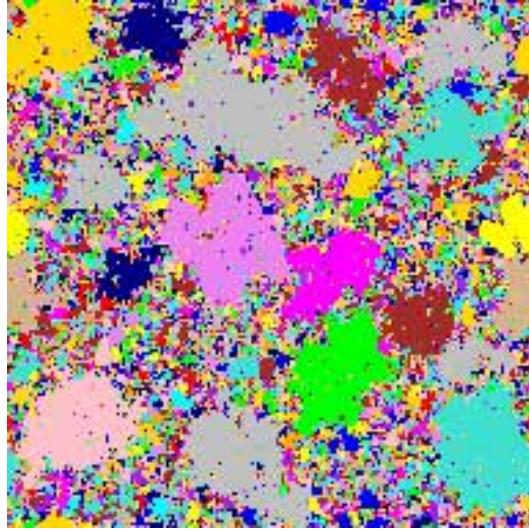


Figure 2.10: 16 state Potts model

entropy, and the more drastic the difference between the ordered state and a disordered state. This in turn gives the Potts model both quantitative and qualitative differences from the Ising model. For $q = 2$ it is equivalent to the Ising model with its second-order phase transition, while we here will demonstrate that larger q open for a first-order phase transition. The derivation here is taken from the review "The Potts model" by F.Y Wu, Review Modern Physics 54, (1982).

The Hamiltonian was

$$H = -\frac{1}{N} \sum_{v>u} \delta_{s(v)-s(u)} \quad (2.51)$$

where the sum runs over all N variables in the system. that is each energy interacts with everyone, but then with each interaction energy is reduced as system size increases. The fraction of spins in state i is now denoted x_i , i between 1 and q . Thus

$$\sum_{i=0}^{q-1} x_i = 1$$

The energy, entropy, and free energy per spin is:

$$\frac{E}{N} = -(1/2) \sum_{i=0}^{q-1} x_i^2 \quad (2.52)$$

$$\frac{S}{k_B N} = - \sum_{i=0}^{q-1} x_i \ln(x_i) \quad (2.53)$$

$$\beta \frac{F}{N} = \sum_{i=0}^{q-1} \left(x_i \ln(x_i) - \frac{1}{2} \beta x_i^2 \right) \quad (2.54)$$

Now we are investigating the transition between all states being equally populated and a single dominating state, parametrized by an order parameter s between 0 and 1:

$$\begin{aligned} x_0 &= \frac{1}{q} (1 + (q - 1)s) \\ x_i &= \frac{1}{q} (1 - s) \quad \text{for all } i = 1, 2, \dots, q - 1 \end{aligned} \quad (2.55)$$

where $s = 1$ corresponding to $x_0 = 1, x_i = 1$ (1-state dominate completely, with zero entropy, $S/N = 0$) and $s = 0$ corresponding to $x_0 = x_i = 1/q$ (messy states with entropy $S/N = k_B \ln(q)$).

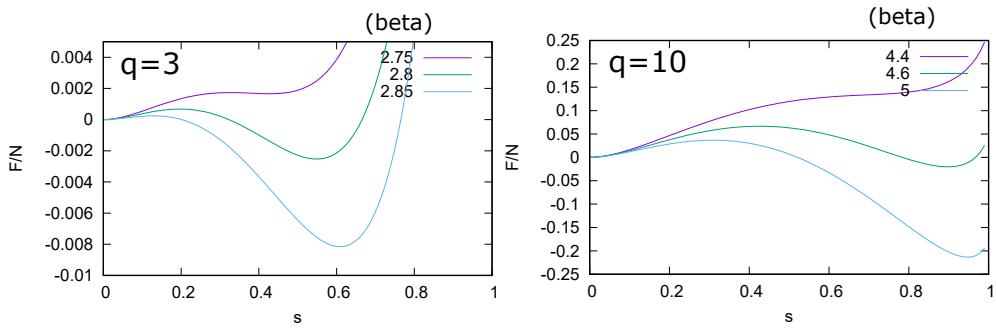


Figure 2.11: Illustration of Free energy behavior of mean field potts model for $q = 3$ and $q = 10$ from eq. 2.56. In both cases, there are some values of β where we have a double-well potential, implying that there are alternative stable states, a hallmark of a First-order transition.

Now we calculate the Free energy:

$$\begin{aligned} \beta \left(\frac{F(s)}{N} - \frac{F(s=0)}{N} \right) &= + \frac{1 + (q - 1)s}{q} \cdot (\ln(1 + (q - 1)s) - \ln(q)) \\ &\quad + (q - 1) \frac{1 - s}{q} \cdot (\ln(1 - s) - \ln(q)) \\ &\quad - \frac{1}{2} \beta \left(\frac{(1 + (q - 1)s)}{q} \right)^2 - \frac{1}{2} \beta \cdot (q - 1) \left(\frac{(1 - s)}{q} \right)^2 \\ &\quad + \frac{1}{2} \beta q \cdot \left(\frac{1}{q} \right)^2 \end{aligned}$$

where the $q - 1$ factor in the second line is due to the $q - 1$ states that each has density $(1 - s)/q$. The last line considers the $s = 0$ energy of the fully disordered state. Rewriting a little:

$$\begin{aligned} \beta \left(\frac{F(s)}{N} - \frac{F(s=0)}{N} \right) &= + \frac{1 + (q - 1)s}{q} \cdot \ln(1 + (q - 1)s) \\ &\quad + (q - 1) \frac{1 - s}{q} \cdot \ln(1 - s) - q \ln(q) \\ &\quad - \frac{1}{2} \beta \cdot \left(\frac{q - 1}{q} s^2 \right) \end{aligned} \quad (2.56)$$

This is plotted for a few cases in the Figure above, illustrating the possibility for alternative stable states.

Expansion for small s : To get some analytics on the observed bistability, we expand the \ln to the third order in s around $s = 0$, using $\ln(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3$ where we consider 1'st, 2'nd and 3'rd order in s (you may also Taylor expand directly using Wolfram alpha):

$$\begin{aligned} First &= \frac{(q-1)s}{q} - s \cdot \frac{(q-1)}{q} = 0 \\ Second &= \frac{(q-1)}{2q} \cdot (q-\beta) \cdot s^2 \\ Third &= -\frac{1}{6}(q-1)(q-2) \cdot s^3 \end{aligned} \quad (2.57)$$

For $q = 2$ one sees that the second-order term switches from negative to positive as T increases beyond $1/q = 1/2$, i.e. critical temperature $\beta_c = q = 2$. The full expression for Free energy is:

$$\beta \left(\frac{F(s)}{N} - \frac{F(s=0)}{N} \right) = (q-\beta) \cdot s^2 - \frac{1}{6}(q-1)(q-2) \cdot s^3 + \dots \quad (2.58)$$

where the emergence of a negative term for $q > 2$ implies that there have to be minimum free energy at high s values for some temperatures. Thus at these temperatures, there will be two solutions to the minimum free energy, one a s equal to zero and one at finite s . This is illustrated in the Figure. When there are two solutions of s for a given temperature, s_1 and s_2 and thus the energy difference between them is associated with Latent heat. One of these solutions will be $s = 0$. The other solution comes from $dF/ds = 0$ and $F(s) = F(s = 0)$.

The position of the minimal fee energy at $s > 0$ can be calculated by dF/ds using the full expression in eq. 2.56:

$$\begin{aligned} \frac{dF/N}{ds} &= \frac{(q-1)}{q} \ln(1 + (q-1)s) + \frac{(1 + (q-1)s)(q-1)}{q(1 + (q-1)s)} \\ &\quad - \frac{q-1}{q} \ln(1-s) - \frac{(q-1)(1-s)}{q(1-s)} - \beta \frac{q-1}{q} s \\ &= \frac{(q-1)}{q} (\ln(1 + (q-1)s) - \ln(1-s) - \beta s) = 0 \\ \Rightarrow \beta s &= \ln\left(\frac{1 + (q-1)s}{1-s}\right) \end{aligned} \quad (2.59)$$

For $q = 2$ we thus get $s = \tanh(\beta \cdot s/2)$ which is the mean-field solution to the Ising model.

The first order phase transition for $q > 2$ Potts model also implies that there is hysteresis around its critical point, which is illustrated in Fig 2.12 (for 2 dimensions q needs to exceed 4 to give first-order transition). Thus there

exists a disordered state at sub-critical temperature ($T < T_c$), which is metastable against a collapse into the ordered state that of course have a lower free energy (marked with red in Fig. 2.12).

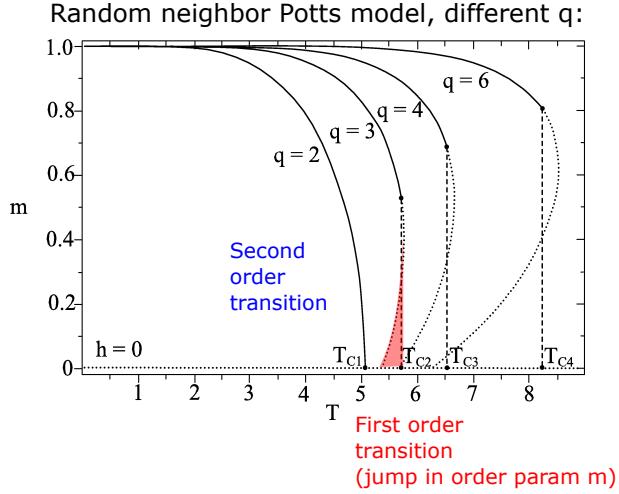


Figure 2.12: **First order transition in Potts model:** magnetization (fraction of spins in majority state) as a function of temperature in a simulation of Potts model [14].

Questions

2.5) Specific heat and susceptibility in mean-field Ising model. (reproducing results above)

Follow the steps in Sec. 2.2.1 to obtain the mean field-specific heat and susceptibility near $T = T_c$, i.e. $t = 0$. 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 $h \rightarrow 0$ is taken before the limit $t \rightarrow 0$). Can you obtain the critical exponent γ (in $\chi_T \sim |t|^{-\gamma}$)?

2.6 Landau free energy from mean-field free energy.

(a) By expanding out the mean-field free energy (Eq. 2.33) to fourth order in m ,

show that the resulting expression is symmetric regarding the transformation $m \rightarrow (-m)$ 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.

2.7 Plot the free energy as a function of s for the Potts model for $q = 2$, $q = 3$, $q = 10$ and $q = 100$ (eq. 2.56) for different temperatures to capture the behavior around the phase transition. Consider the $q = 10$ case and plot the s values for which $dF/ds = 0$ as a function of temperature ($k_B T = 1/\beta$). Calculate numerically the Energy difference between the alternative stable states.

2.8 Redo the Taylor series expansion in eq. 2.57.

2.3 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. 2.3). The corresponding 1D Hamiltonian is

$$\mathcal{H}_N = -J \sum_{i=0}^{N-1} s_i s_{i+1} - h \sum_{i=0}^{N-1} s_i , \quad (2.60)$$

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

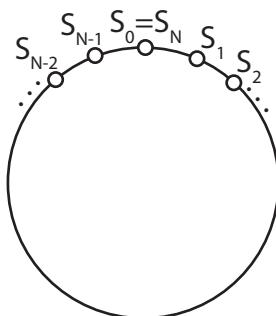


Figure 2.13: **1D Ising model.** Each of the N sites i has two nearest neighbors. Note the periodic boundary conditions, which are enforced by demanding $S_0 = S_N$, i.e. the 1D system becomes a closed loop of N sites.

Partition function

The partition function for the N sites is

$$Z_N = \sum_{s_0=\pm 1, s_1=\pm 1, \dots, s_{N-1}=\pm 1} \exp(\beta J(s_0s_1 + s_1s_2 + \dots + s_{N-1}s_0) + \beta h(s_0 + s_1 + \dots + s_{N-1})) , \quad (2.61)$$

where the notation $\{s\}$ means that all configurations of the different s_i are summed over in Eq. 2.61, 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

$$\begin{aligned} Z_N = & \sum_{s_0=\pm 1, s_1=\pm 1, \dots, s_{N-1}=\pm 1} \underbrace{e^{\beta Js_0s_1 + \beta h \frac{s_0+s_1}{2}}}_{T_{0,1}} \underbrace{e^{\beta Js_1s_2 + \beta h \frac{s_1+s_2}{2}}}_{T_{1,2}} \dots \\ & \cdot \underbrace{e^{\beta Js_{N-1}s_0 + \beta h \frac{s_{N-1}+s_0}{2}}}_{T_{N-1,0}} . \end{aligned} \quad (2.62)$$

Noticeably, each of the factors in the argument of \exp in Eq. 2.62 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

$$\begin{bmatrix} e^{\beta(J+h)} & e^{-\beta(J)} \\ e^{-\beta(J)} & e^{\beta(J-h)} \end{bmatrix} \quad (2.63)$$

where the rows correspond to the two values of $s_i = \pm 1$ and columns correspond to the two values of $s_{i+1} = \pm 1$. Hence, Eq. 2.62 can be interpreted as N matrix multiplications, which is easy to see if remembering that the matrix multiplication of $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times k}$, which yields a matrix $M \in \mathbb{R}^{m \times k}$ is defined as

$$\begin{aligned} M &\equiv AB , \text{ whereby} \\ M_{mk} &\equiv \sum_n A_{mn} B_{nk} . \end{aligned} \quad (2.64)$$

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

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

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 the characteristic polynomial

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

or

$$\begin{aligned} e^{2\beta J} - e^{-2\beta J} - 2\lambda e^{2\beta J} \cdot \cosh(\beta h) + \lambda^2 &= 0 \\ 2 \cdot \sinh(2\beta J) - 2e^{2\beta J}\lambda \cdot \cosh(\beta h) + \lambda^2 &= 0 \\ (\lambda - e^{\beta J} \cosh(\beta h))^2 + 2 \cdot \sinh(2\beta J) - e^{2\beta J} \cosh^2(\beta h) &= 0 \end{aligned}$$

This give the two solutions for the eigenvalues:

$$\begin{aligned} \lambda_{1/2} &= e^{\beta J} \left(\cosh \beta h \pm \sqrt{\cosh^2(\beta h) - 2e^{-2\beta J} \sinh(2\beta J)} \right) \\ &= \left(e^{\beta J} \cosh \beta h \pm \sqrt{e^{2\beta J} \sinh^2(\beta h) + e^{-2\beta J}} \right) \\ &= e^{\beta J} \pm e^{-\beta J} \quad \text{for } h = 0 \end{aligned}$$

where we uses $\cosh^2(x) = \sinh^2(x) - 1$ and $2 \cdot \sinh(2x) = e^{2x} - e^{-2x}$

Free energy

In the thermodynamic limit, that is, for $N \rightarrow \infty$, the free energy per site is easy to compute. Assuming that the eigenvalues are listed in decreasing magnitude, i.e. $\lambda_1 > \lambda_2$,

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

In the thermodynamic limit it can be used that $\lambda_1 > \lambda_2$, and thus the ratio $\left(\frac{\lambda_2}{\lambda_1} \right)^N \rightarrow 0$. Hence the free energy per site, f , 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, $\beta \rightarrow \infty$, and thus

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

which is the ground state energy of a single spin.

What is f in limit of infinite temperature.

Knowing f one can also compute the magnetization as a derivative (Energy (system state) = $-J \sum s_i s_j - h \sum s_i$:

$$m = (1/\beta) \cdot \frac{\sum_{\text{states}} (\sum s_i) \exp(\beta(\sum s_i s_j + h \sum s_i))}{\sum_{\text{states}} \exp(\beta(\sum s_i s_j + h \sum s_i))} = (1/\beta) \cdot \frac{d \ln(Z)}{dh}$$

namely (using $f = -k_B T \ln(Z)$):

$$\begin{aligned} m &= -\frac{\partial f}{\partial h} = k_B T \cdot \frac{1}{\lambda_1} \cdot \frac{\partial \lambda_1}{\partial h} \\ &= \frac{e^{\beta J} \sinh(\beta h) + (e^{2\beta J} \sin(2\beta h)) / (\sqrt{e^{2\beta J} \sinh^2 \beta h + e^{-2\beta J}})}{(e^{\beta J} \cosh \beta h + \sqrt{e^{2\beta J} \sinh^2 \beta h + e^{-2\beta J}})} \\ &= \frac{\sinh(\beta h) + \sinh(2\beta h) / \sqrt{\sinh^2(\beta h) + e^{-4\beta J}}}{(\cosh(\beta J) + \sqrt{\sinh^2(\beta h) + e^{-4\beta J}})} \end{aligned}$$

After some manipulations, it is found that (using $\sinh(2x) = 2\sinh(x)\cosh(x)$, redo and check yourself)

$$m = \frac{\sinh(\beta h)}{[\sinh^2(\beta h) + e^{-4\beta J}]^{1/2}}. \quad (2.65)$$

It is interesting to explore limiting behavior of $m(T)$:

- consider fixed $J > 0$ and $h \neq 0$ and allow $\beta \rightarrow \infty$, thus approaching the zero-temperature (ground state) limit. It is hence found that $|m| \rightarrow 1$, where the sign is determined by the sign of h . Conversely, in the high-temperature limit, where $\beta \rightarrow 0$, we have that $m \rightarrow 0$, i.e., all magnetization vanishes.
- consider fixed $J > 0$ and $\beta > 0$, and allow $h \rightarrow 0$. We now find that $m \rightarrow 0$, for any (fixed) choice of $J > 0$ and $\beta > 0$. Thus, there is no spontaneous magnetization for any finite temperature in the 1D Ising model. This stands in stark contrast to the 2D Ising model, where we showed in the mean field approximation, that spontaneous magnetization exists when $T < T_c$. Conversely, in the limit of $h \rightarrow \infty$, $|m| \rightarrow 1$, thus, for any given temperature a sufficiently large field h can polarize all spins.

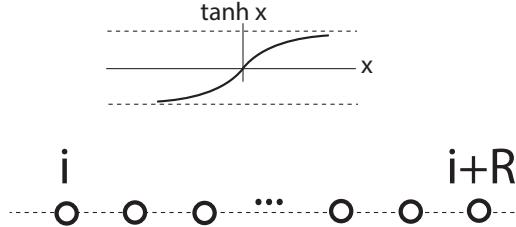


Figure 2.14: **Spin-spin correlation function.** Spin chain with two spins separated by R sites, $\tanh x$, with $x = \beta J$, is the correlation function between two neighboring spins.

Correlation function

Re-consider the zero-field spin-spin correlation function, namely

$$\begin{aligned} \langle s_i s_{i+l} \rangle &\equiv \frac{1}{Z} \sum_{s_1=\pm 1, s_2=\pm 2, \dots, s_N=\pm 1} s_i s_{i+l} e^{\beta J \sum_k s_k s_{k+1}} \\ &= \frac{1}{Z} \lim_{J_k \rightarrow J, \text{ for } k \in \{1, \dots, N\}} \sum_{\text{all states}} s_i \underbrace{s_{i+1} s_{i+1}}_1 \underbrace{s_{i+2} s_{i+2}}_1 \dots \\ &\quad \cdot \underbrace{s_{i+l-1} s_{i+l-1}}_1 s_{i+l} e^{\beta \sum_k J_k s_k s_{k+1}} \end{aligned}$$

where we inserted a number of factors equal to 1. This each can be re-capitulated by differentiating $Z = \sum_{\text{all states}} \exp(\beta \sum_k J_k s_k s_{k+1})$ with respect to variables $J_1, J_2, J_3 \dots$ that all are going to be varied around $J_i \sim J$. Each differentiation gives one factor $\cosh(\beta J)$ in partition sum because there is contribution from $s_I = +1$ and $s_i = -1$. Recasting to differentiation one obtains:

$$\begin{aligned} \langle s_i s_{i+l} \rangle &= \frac{1}{Z} \lim_{J_k \rightarrow J, \text{ for } k \in \{1, \dots, N\}} \sum \frac{1}{\beta^l} \frac{\partial^l Z}{\partial J_i \partial J_{i+1} \partial J_{i+2} \dots \partial J_{i+l-1}} \\ &= \frac{[\sinh(\beta J)]^l [\cosh(\beta J)]^{N-l}}{[\cosh(\beta J)]^N} \\ &= [\tanh(\beta J)]^l. \end{aligned} \tag{2.66}$$

where the $\sinh()$ factors come from the s_i prefactors that take +1 and -1 values for each of the l sites between the i site and the $i+l$ site. Hence, for any $\beta > 0$ and $J > 0$ the spin-spin correlation function is reduced by some factor less than unity for each step taken away from the origin i . This implies that the correlations decay exponentially with distance. Now we want to express the

characteristic length of this decay. We re-write Eq. 2.66

$$\begin{aligned}\langle s_i s_{i+l} \rangle &= \exp(l \ln(\tanh(\beta J))) \\ &= \exp\left(-l \underbrace{\ln\left(\frac{1}{\tanh(\beta J)}\right)}_{1/\xi}\right) = e^{-l/\xi},\end{aligned}\quad (2.67)$$

where $\xi \equiv -\frac{1}{\ln(\tanh(\beta J))}$ is the correlation length. This only diverges for $\beta \rightarrow \infty$, representing the zero-temperature (ground state) limit. This is another way of realizing that the 1D Ising model does not show any long-ranged order for finite temperatures. For any finite temperature, the exponential decay of correlations would require the susceptibility to remain finite.

There is another, intuitive, way to compute the zero-field ($h = 0$) correlation between two spins that are separated by l lattice sites. Consider first the correlation of any two neighboring spins, that is, 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 comes from equal contributions from aligned and anti-aligned spins in the denominator. This result was then 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 l lattice sites instead:

$$\Gamma(l) = \langle s_i s_{i+l} \rangle, \quad (2.68)$$

and realize that the expectation value in Eq. 2.68 will not change by inserting products $s_j s_j = 1$, hence

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

where the last equation expresses independent alignment between one pair and the next pair along the 1-dimensional lattice. 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.

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

This recovers the result from above.

Finally, the finite correlation length can approximately be obtained from an entropy argument, where each shift in spin direction costs energy $2J$ but allows for an entropy $S \sim k_B \ln(\xi)$ where ξ would be the length of a typical segment with spins in the same state. Thus energy cost and entropy gain balances when $k_B T \ln(\xi) \sim 2J$.

Qlesson: The 1D Ising model can be solved analytically, e.g., using the transfer matrix method, whereby the free energy and the observables can be written down as exact expressions. The 1D Ising model does not show any finite-temperature phase transition, whereas the 2D Ising model does.

Questions

2.9 Helix Coil Transition: The helix-coil transition can be treated as an analog to the 1-d Ising model with a slightly different transition matrix. It is parametrized in terms of two quantities (Zimm (1959), Sheraga (1973)): an entropy cost term σ that counts the loss in conformational entropy to form the first bond, and an energy gain term that counts the energy for each subsequent addition of a bond. In more detail, let us normalize the statistical weights relative to the coil (c) state, i.e. we assign to it a free energy $F_c = 0$ and therefore a statistical weight $\exp(-F_c/k_B T) = 1$. To initiate a helix then costs a free energy $F_{ch} = F_{hh} - T S_{ch}$ with a corresponding statistical weight $\exp(-F_{ch}/k_B T) = \sigma s$; to continue a helix costs a free energy F_{hh} , with a statistical weight $\exp(-F_{hh}/k_B T) = s$. From the above, $\sigma < 1$ always, while s can be < 1 (then there is no transition to a helix state), or > 1 . Thus:

σs = statistical weight to initiate helix in coil region

s = statistical weight to continue helix The statistical weights are associated

with the 4 possible transitions as one considers the possible ways to distribute the two states along a still longer polymer:

$$\begin{aligned} c \rightarrow h: & \sigma s \\ c \rightarrow c: & 1 \\ h \rightarrow h: & s \\ h \rightarrow c: & 1 \end{aligned}$$

Given a system of size N , we denote $Z_c(N)$ the statistical weight that the last state in the sequence is in the coil state. The statistical weight that the last state is in a helix state is denoted $Z_h(N)$. Then

$$Z(N) = Z_c(N) + Z_h(N)$$

is the total partition sum for the system of size N . Now the basic trick is to use the transition probabilities above and decompose the statistical weight of a system of size $N + 1$ as:

$$\begin{aligned} Z_c(N+1) &= Z_c(N) + Z_h(N) \\ Z_h(N+1) &= s\sigma Z_c(N) + s Z_h(N) \end{aligned}$$

Show that the largest eigenvalue and associated free energy is

$$\lambda_{max} = \frac{1}{2} \left(s + 1 + \sqrt{(1-s)^2 + 4\sigma s} \right).$$

$$F = -k_B T \ln(Z) = -k_B T N \ln\left(\frac{1}{2}(1+s+\sqrt{(1-s)^2+4\sigma s})\right)$$

The partition sum can also be expressed as

$$Z = \sum \Omega_{jk} \sigma^j s^k$$

Show from this that $d\ln(Z)/d\ln(s) = (s/Z) \cdot dZ/ds$ can be used to calculate $\langle k \rangle$ which is equal to the number of helix sites $N\langle \text{helix} \rangle$ in our system of length N . Show that the average helix content of the polymer is:

$$\langle \text{helix} \rangle = \frac{\langle k \rangle}{N} = \frac{1}{N} \frac{d \ln(Z)}{d \ln(s)} = \frac{s}{2\lambda_{max}} \left(1 + \frac{s+2\sigma-1}{\sqrt{(1-s)^2+4\sigma s}} \right)$$

Calculate the average number of helix initiations.

2.10 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. 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.)

2.11 No finite temperature phase transition in 1D Ising model. +
Using the free energy and the concept of domain walls (Fig. 2.3), 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.

2.12 No finite temperature phase transition in 1D Ising model.

The free energy of the $h = 0$ ferromagnetic Ising model for a state of fixed energy E_Ω can be calculated from

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

Here, Ω indicates possible multiplicity of states of the same internal energy. F_Ω can be considered as the combined canonical free energy contributed of all states with the same internal energy.

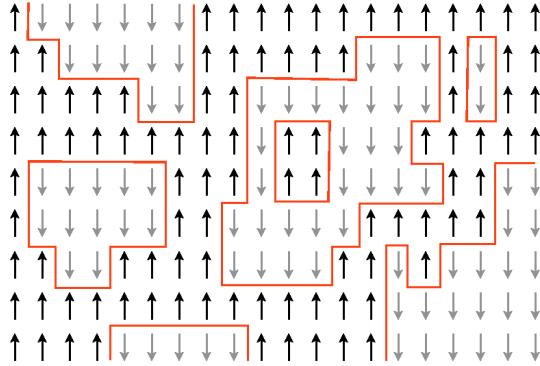


Figure 2.15: **Examples of domain walls in a 2D Ising model.** Domain walls are shown as red lines.

(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. 2.3 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).

2.13 Exponent relations.³

Starting from the scaling for the singular part of the free energy

$$f^{(s)}(t, h) \sim b^{-d} f^{(s)}(b^{y_1} t, b^{y_2} h) \quad (2.69)$$

where $t \equiv (T - T_c)/T_c$ and $h = h_0/k_B T$ show that $\alpha = 2 - d/y_1$, $\beta = (d - y_2)/y_1$,

³Yeomans, problem 8.2

$\gamma = (2y_2 - d)/y_1$ and $\delta = y_2/(d - y_2)$ and hence confirm that $\alpha + 2\beta + \gamma = 2$, and $\gamma = \beta(\delta - 1)$.

2.14 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.

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 the pattern becomes random, and you will end up with a featureless mix of up and down spins.

2.15. Monte Carlo Simulation of Potts model:

Simulate the Potts model with $q=10$ on a 2-d lattice and find the critical temperature T_c , above which only the disordered state exists. Also find the temperature interval T_{min} to T_c where one alternates between an ordered state and a disordered state where all spins are about equally represented.

Lessons:

- Ising model Hamiltonian for a Ferrogmagnet is (in units of J)

$$H = - \sum_{i,j} s_i s_j$$

where sum is over all neighbor sites in the lattice, and each spin s_i takes values $+1/2$ or $-1/2$.

- Mean Field Ising model is obtained by assuming that spin at a given site interact with the average spins of the entire lattice, $H \sim H_{mean-field} = - \sum_i s_i \langle s \rangle$. The partition function can then be factorized, with each factor having a contribution from the up spin plus the contribution for the down spin.

- Phase transition occur at a critical point, where system becomes scale free with exponents that only depend on symmetries and not on details of interactions.
- 1-d Ising model have no phase transition, and instead fragment into smaller and smaller segments of parallel spins as temperature is increased. This is because of the entropy gained by making domains.

Chapter 3

Percolation, Fractals & Fracture

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

– Johann Wolfgang von Goethe

3.1 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 $\exp(-E_i/k_B T)$. For example, if one spin in the 1-dimensional Ising model is aligned with its neighbors and has ground state energy 0, it will have the relative probability $\exp(-2J/k_B T)$ to be switched in the opposite direction. Such systems predict that the probability of finding any particular state to be occupied exponentially decreases with the state's energy.

In general, an exponential decrease in probability implies that the probability distribution $P(k)$ of a observable k has a typical scale a :

$$P(k) \propto \exp(-k/a). \quad (3.1)$$

For example, consider a system where energy increases linearly with the size a . In this case, the probability of observing various sizes fulfills

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

Therefore the probability for finding something large, say, $10 k$ or $20 k$ becomes vanishing small — and finding something double as large becomes exceedingly smaller again.

In equilibrium, this type of distribution arises when the probability of concentrating one more unit of energy on one variable is e^{-1} , and this remains true independently of how much energy one has already concentrated.

It factorizes to very small probabilities indeed. Exponential distribution as the normal Boltzman distribution makes unusually large concentration of energy essentially impossible.

Exponential distributions and objects with characteristic sizes are often seen in the real world. Think for example about rain droplets. Or think 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 with the binding energy of the ordered state just exactly balances the exponential associated with the entropy of the disordered state (*compare* Fig. 2.3 and corresponding exercise). But this is a special situation because the temperature will have to be externally tuned to be at the critical point.

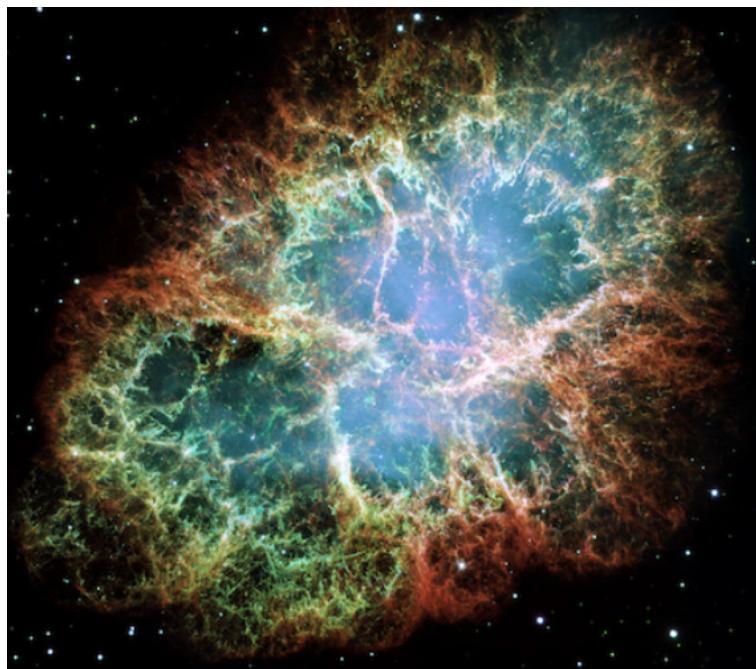


Figure 3.1: Remnants of a supernova that formed the Crab nebula. Fractals and power laws are ubiquitous in nature, for example in turbulence or fragmentation. *Image: NASA, ESA, J. Hester, A. Loll (ASU) Acknowledgment: Davide De Martin (Skyfactory).*

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.) The probability to obtain an Energy E is estimated to scale as $P(E) \sim 1/E^{1.6 \pm 0.06}$ for solar x-ray bursts [15].

- Most people have a few 100 followers on their internet activity, but some have millions.
- Gene regulatory networks, where most proteins only associate with a few others, but some are central hubs and associate with most others.
- Turbulent liquids can have large vortexes 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. 3.1.

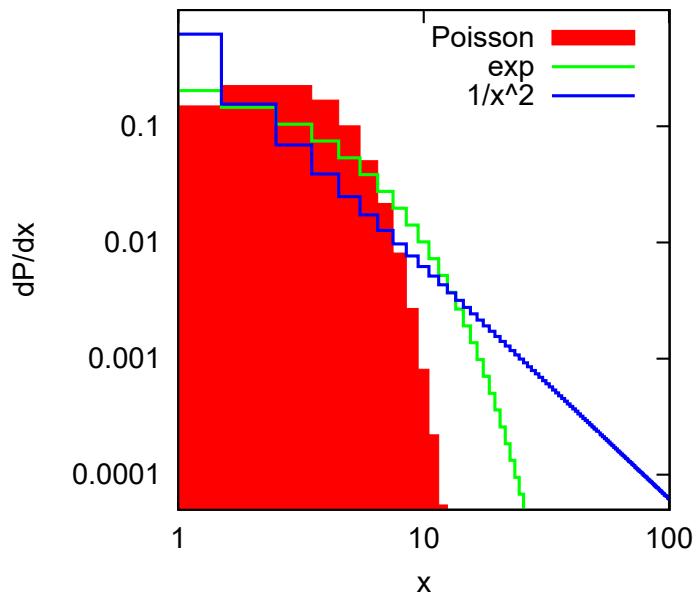


Figure 3.2: **Idealized distributions:** Comparison between a power law with exponent 2, that is, $\sim 1/k^2$, the Poisson distribution ($p(x) \propto a^x/x!$) and an exponential distribution $p(x) \propto \exp(-x/a)$. Notice that both the x and y axes are logarithmic.

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

$$P(k) = \frac{1}{k^\gamma} , \quad (3.2)$$

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. 3.2. 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 complicated way using the rank distribution of sizes, see Fig. 3.3. 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., $= 1/(\gamma - 1)$).

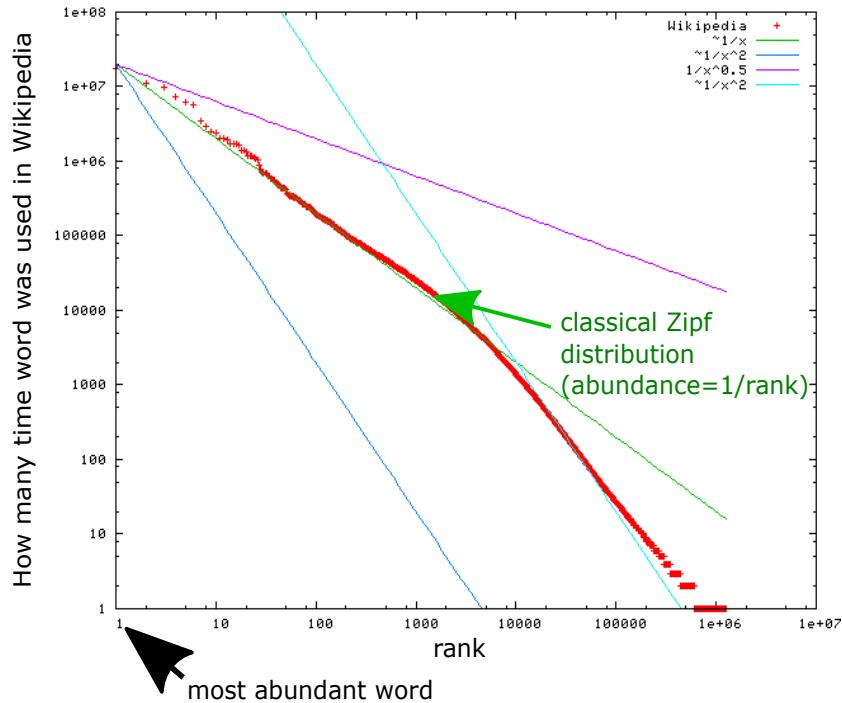


Figure 3.3: **Rank ordered distribution of the number of times different words are used in Wikipedia.** Thus there are ten words that were used more often than 2,000,000 times, and 1,000 words that were used more often than 20,000 times. (Identify a word of rank 1000 and see how many times this was used. All words with lower ranks 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, $(P(s \cdot k_1)/P(k_1)) =$

$p(s \cdot k_2)/p(k_2)$ for all scales k_1 and k_2) hence

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

where $f(k) = \log(P(k)/P(1))$ and $f(1) = 0$. Thus $f(k)$ fulfills the two defining requirements for being a logarithm ($f(ab) = fa + fb$ and $f(1) = 0$):

$$\log(P(k)/P(1)) = \gamma \cdot \log(k) \Rightarrow P(k) \propto k^\gamma \quad (3.3)$$

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:

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

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

3.1.1 Universality:

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 percolation as a static problem with some analogy to the Ising model with power laws close to a critical point. In subsequent chapters, we will see other mechanisms that generate power laws, including in particular self-organized criticality, “rich gets richer” and merging processes. In all cases, we are driven by the belief that there is universality.

Scaling and scaling phenomena are closely related to universality: The observation that the large-scale free features are liberated from microscopic details. This is expressed in terms of scaling exponents that characterize how one scale relates to another, without resorting to the microscopic scale. In Ising model the exponents at the critical point does not depend on whether one uses a triangular, hexagonal, or square lattice. The critical point does, but how for example the magnetization changes as one moves away from the critical point does not. In the percolation problem discussed in this chapter, the exponent of the cluster size distribution of density of the largest cluster similarly does not depend on lattice details. However, the position of the critical point does.

Questions:

3.1) Consider the distribution [16, 17, 18]

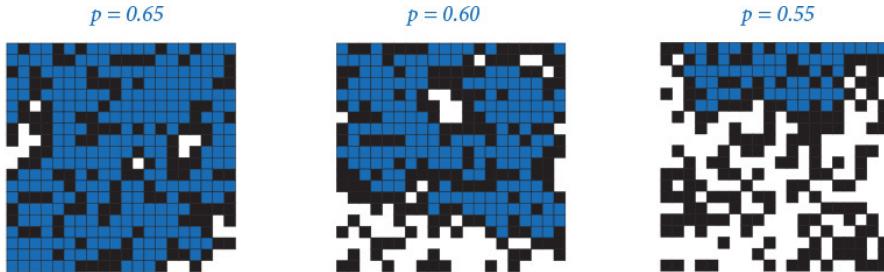
$$p(s) \propto 1/s^\tau$$

as a distribution of 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 the respective contribution to average wealth in the distribution.

Qlesson: Exponent 2 is special, and it is also the most common in nature/ sociology/ economy, it is the original Zipf law exponent [19].

3.1B) Prices law states that for an organization of N persons, then \sqrt{N} do 1/2 the work. Lotka's law, from "The Frequency Distribution of Scientific Productivity, 1926) states $P(s) \propto 1/s^2$. For $P(s) = 1/s^2$ then $y = 1/x - 1/M$ is a fraction of the population that does more than threshold work $s = x$ and less than some upper limit M . In order that this population fraction makes 1/2 of all work $\int_x^M s ds / s^2 = \frac{1}{2} \int_1^M s ds / s^2$ or $\ln(M/x) = (1/2)\ln(M)$ or threshold work $x = \sqrt{M}$. The fraction of population above this threshold is $y = 1/x - 1/M \sim 1/\sqrt{M}$ and Price states that this fraction should be $1/\sqrt{N}$ which is true if the max work anyone can do is $M = N$. Thus Lotka's law is consistent with Price law, provided we look at distributions of employers where occasionally 1 person can do all the work.

Qlesson: Alternatively, the more usual 20% of people doing 80% of work, $\int^x sp(s)ds = (1/5) \int^N sp(s)$ giving a cutoff x that is translated into to part of population $s = \int_x^N p(s)ds$. With $p(s) = 1/s^{\text{gamma}}$ then prove $\gamma = 3/2 + \sqrt{5}/2$.



Percolation is less probable as the site vacancy probability p decreases

Figure 3.4: **Percolation** Blue color indicate water, that can only move along white/blue cells. Here we imagine water is put on all sites and on top. When the path of water is connected from top to bottom, there is percolation, and the water can run through the medium.

3.2 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, see Fig. 3.4. 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. 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.

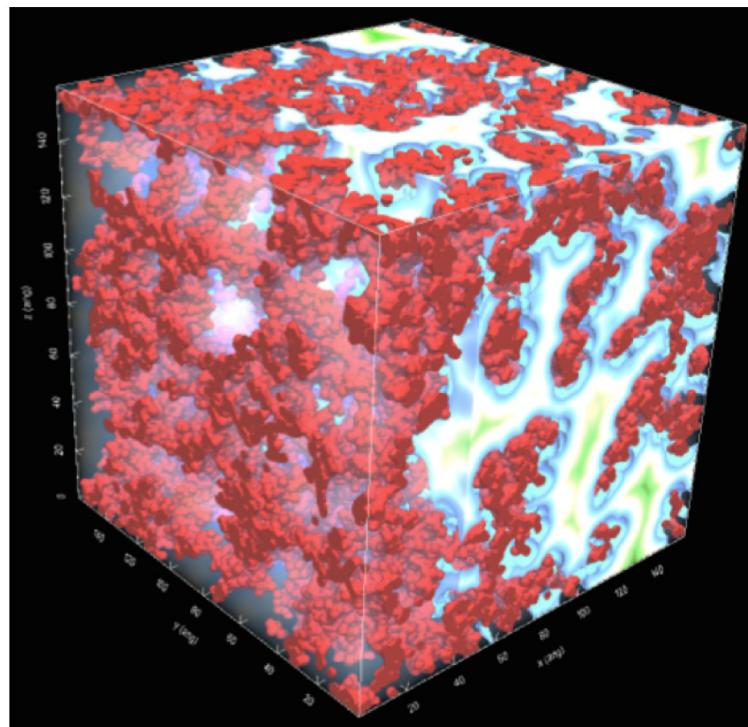


Figure 3.5: Percolation in nanoporous silicate Color indicating distance to nearest part of solid. From Malte Sørensen's lecture notes on percolation, Springer.

Let us first consider a simple percolation example on a two-dimensional square lattice (Fig. 3.6). 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 that are directly or indirectly connected by bonds. Each of these clusters is 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. 3.6 using Python, which in turn will allow us to gain some intuition for this type of problem.

Questions:

- 3.2)** 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

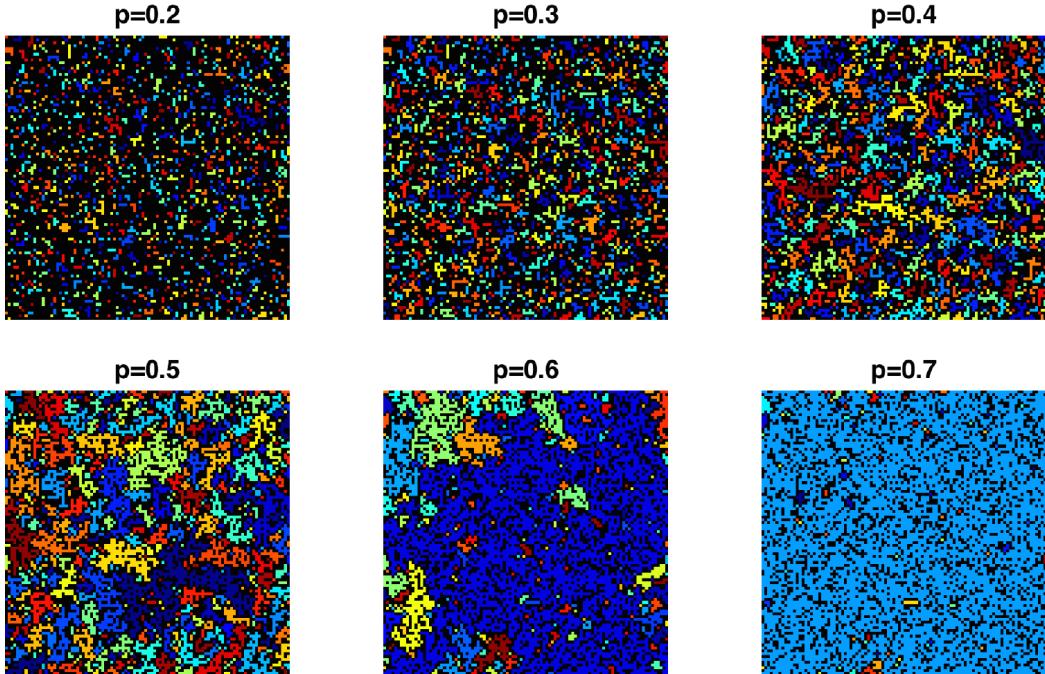


Fig. 1.4 Plot of the clusters in a 100×100 system for various values of p .

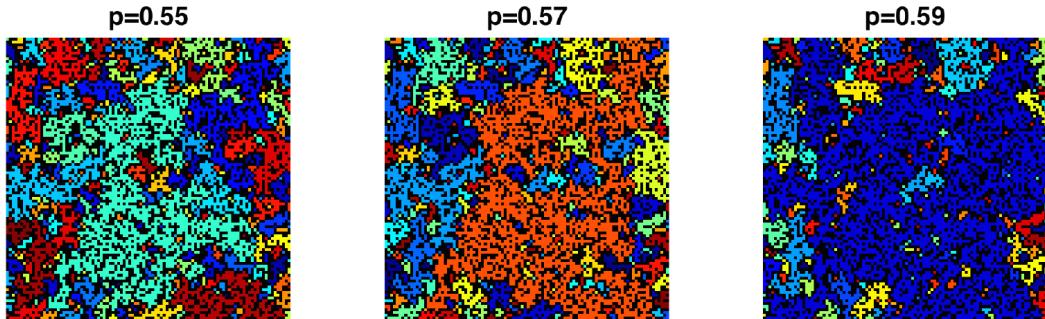


Figure 3.6: Clusters in the site percolation model. In the example shown, sites are organized on a two-dimensional square lattice and each site is occupied with probability 0.55. Neighboring sites of equal color belong to the same cluster. From Malte Sørensen's lecture notes on percolation, Springer.

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.

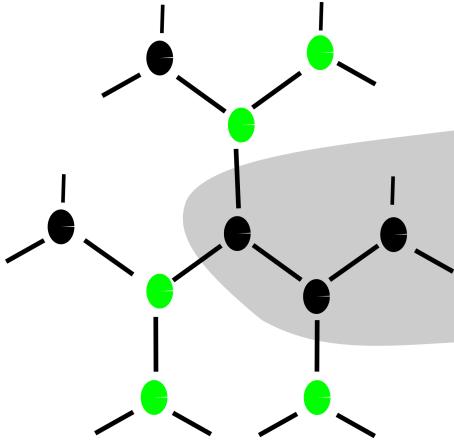


Figure 3.7: **Bethe lattice with occupied (black) and empty sites (green).** The lattice continues out to infinity without any closed loop, i.e. branches are all separate. The black sites form clusters, that may percolate to infinity, if the density of the black sites (the probability of being occupied) 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).

3.2.1 Percolation on a Bethe-lattice

Percolation is a classical problem in physics, with multiple applications [20, 21, 22]. Fig. 3.6 outlines a classical percolation problem, of a liquid percolating a porous material. The percolation process depends critically on the fraction of conducting space to blocked space. This critical behavior is what we will investigate in this chapter. We will do this by first considering percolation on idealized geometries, defined by the 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. 3.7. Each branch originating from a given site contains $z - 1$ (two, in the example) new sub-branches at the $z = 3$ neighboring sites. Let us first consider 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

$$N(\leq x) = 1 + 3 \cdot (1 + 2 + 2^2 + \dots + 2^{x-1}) = 1 + 3 \cdot \sum_{i=0}^{x-1} 2^i = 3 \cdot 2^x - 2, \quad (3.4)$$

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

$$N_D(\leq x) \propto x^D \quad (3.5)$$

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 Bethe-lattice?

The critical probability p_c for percolation on a Bethe lattice can be calculated by considering an occupied site and imagining a virus that spreads to neighbor sites. Subsequently, new sites are then invaded. Consider one such new site: The critical branching is happening when this site connects to just one subsequent site further out in the network. This one new site can be one of $z - 1$ neighbor sites (back propagation does not count). This means that

$$p_c \cdot (z - 1) = 1, \text{ hence } p_c = \frac{1}{z - 1} = \frac{1}{2} \quad (3.6)$$

Again, inspect the Bethe lattice with $z = 3$ in Fig. 3.7 to confirm that $p_c = 0.5$ 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 is, when the limit $\lim_{p \rightarrow p_c} \mathcal{Q}$ is approached?

Average cluster: First consider the scaling of the mean cluster size $S(p) \propto |p_c - p|^{-\gamma}$ to which **an occupied site** belongs (scaling for $p \rightarrow p_c$, 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:

$$S(p) = 1 + 3 \cdot T. \quad (3.7)$$

Here T is the average contribution from one of the sub-branches and 3 is the coordination number z of the Bethe lattice. This is illustrated in Fig. 3.8. This contribution can be determined self-consistently from

$$T = p \cdot (1 + 2T), \quad (3.8)$$

because T only receives a contribution from the first site of the sub-branch if this is occupied (with probability p), multiplied by the contribution from the two subsequent sub-branches when the first site is occupied.

$$T = \frac{p}{1 - 2p} \text{ for } p < p_c = 1/2. \quad (3.9)$$

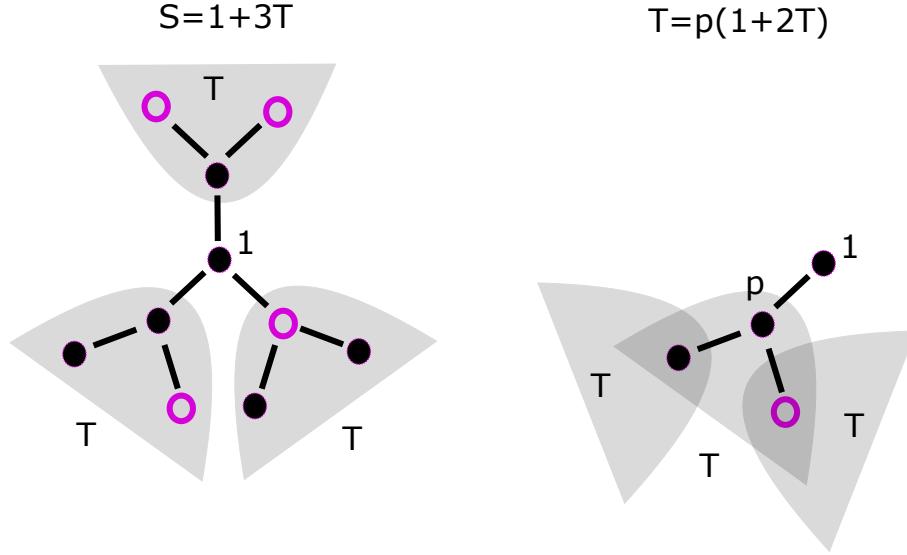


Figure 3.8: **Recursion relation for average cluster size** Left side shows initial cluster size starting from 1 unit of occupied lattice. The right panel shows recursion following one of the 3 branches.

Inserting into Eq. 3.7, we therefore obtain the mean cluster size as

$$S(p) = 1 + 3T = \frac{1 + p}{2(p_c - p)} \propto (p_c - p)^{-1}, \quad (3.10)$$

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.

Distribution of cluster sizes: Now we are nearly in a position to calculate the intuitive characteristics of the near percolating system, namely the distribution of cluster sizes, $n_s(p)$. That is we will count the number of clusters of size s , and see how it varies with the percolation parameter 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

$$n_s(p) = \sum_t g_{s,t} p^s (1-p)^t, \quad (3.11)$$

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 $g_{s,t}$.

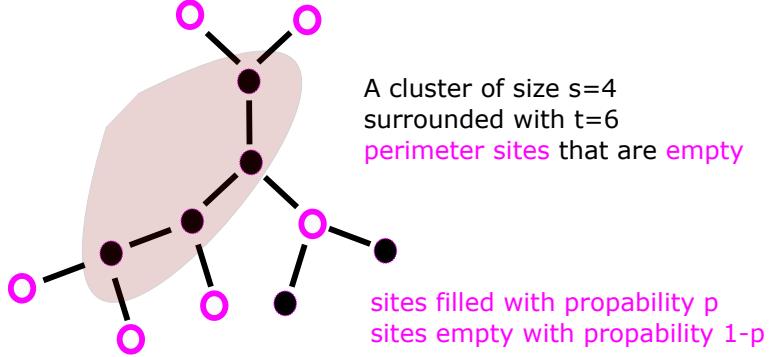


Figure 3.9: **A cluster with boundary sites:** For Bethe lattice we always have Always $t=s+2$, since each added filled site would remove one previous boundary site and add two new.

For a Bethe lattice with $z = 3$ the perimeter number $t = 2 + s$, see Fig. 3.9 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 loses 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,

$$t = 2 + s . \quad (3.12)$$

Accordingly, the configuration count $g_{s,t}$ in the above sum only contains non-zero values when $t = 2 + s$:

$$n_s(p) = g_{s,2+s} \cdot p^s \cdot (1 - p)^{2+s} , \quad (3.13)$$

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 $g_{s,s+2}$:

$$\begin{aligned} \frac{n_s(p)}{n_s(p_c)} &= \left(\frac{1-p}{1-p_c} \right)^2 \left(\frac{p}{p_c} \cdot \frac{1-p}{1-p_c} \right)^s \\ &= \left(\frac{1-p}{1-p_c} \right)^2 \exp \left(s \cdot \ln \left(\frac{p}{p_c} \cdot \frac{1-p}{1-p_c} \right) \right) \\ &= \left(\frac{1-p}{1-p_c} \right)^2 \exp \left(s \cdot \ln(4p - 4p^2) \right) . \end{aligned}$$

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 2.1.1). Denote the argument of the natural logarithm in Eq. 3.14 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

$f(p) = 1 - 4 \cdot (p - p_c)^2$ (its actually exact). We therefore get

$$\begin{aligned} \frac{n_s(p)}{n_s(p_c)} &\sim \left(\frac{1-p}{1-p_c} \right)^2 \cdot \exp(s \cdot \ln(1 - 4 \cdot (p - p_c)^2)) \\ &\sim \left(\frac{1-p}{1-p_c} \right)^2 \cdot \exp\left(-\frac{s}{s_{p-p_c}}\right) \end{aligned} \quad (3.14)$$

Thus

$$s_{p-p_c} \equiv \frac{-1}{\ln(1 - 4 \cdot (p - p_c)^2)} \sim \frac{1}{4} \cdot (p - p_c)^{-2} \propto |p - p_c|^{-1/\sigma}. \quad (3.15)$$

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. 3.14. 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 . That means that clusters larger than the threshold $1/(p_c - p)^2$ are exponentially unlikely, while the frequency of clusters smaller than this threshold is dominated by power law scaling.

In Fig. 3.10 we investigate the $n_s(p)/n_s(p_c)$ highlighting a so-called rescaling plot that puts focus on the cut-off function. By considering this ratio we are of course not able to address the behavior of $n_s(p_c)$ itself; that would have required a calculation of the $g_{s,s+2}$ instead. To calculate this critical distribution we just assume that it is a power law, and then use our prior derived knowledge of the scaling of the average size of clusters.

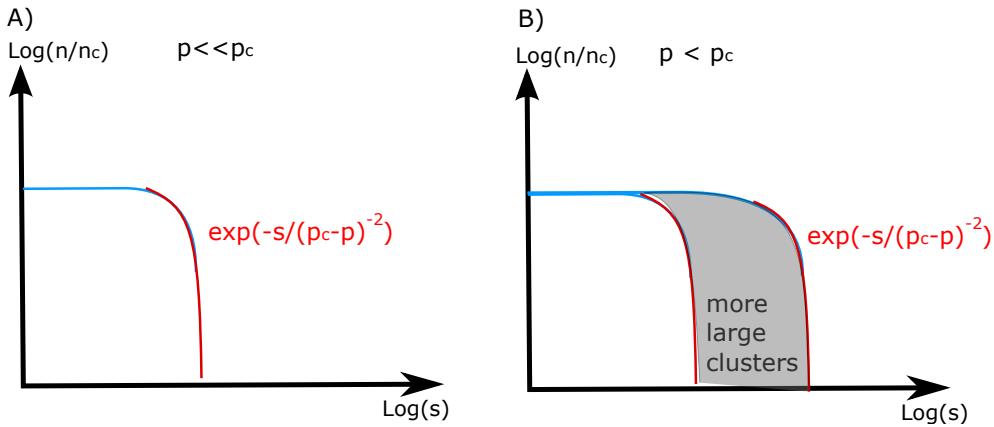


Figure 3.10: **Distribution of cluster sizes normalized by critical one:** A) show re-normalized distribution $n_s(p)/n_s(p_c)$ for p small, while B) show distribution for p much closer to p_c . Grey area marks added new large clusters. By dividing with the unknown critical distribution the plot highlights the diverging cutoff as it diverges as $(p - p_c)^{-2}$.

Change coordination number from $z = 3$ to $z = 4$ and recalculate the above scaling of maximal cluster size with p approaching p_c (how does relation be-

tween t and s change?)

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 $s_{p \rightarrow p_c}$ 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 has to come up with a distribution that connects 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 these diverging scales we need a scale-free distribution, that is we need a scale-free distribution of the cluster sizes.

Mini tutorial: How likely is it to find a cluster of size s relative to 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

$$n_s(p) = n_s(p_c) \cdot \exp(-s/s_{p \rightarrow p_c}) \propto s^{-\tau} \cdot \exp(-s/s_{p \rightarrow p_c}). \quad (3.16)$$

We have hence incorporated the cutoff and assume a power law distribution for all cluster sizes below this cutoff. That is, we assume that the only relevant scale is the cutoff scale set by $p_c - p$.

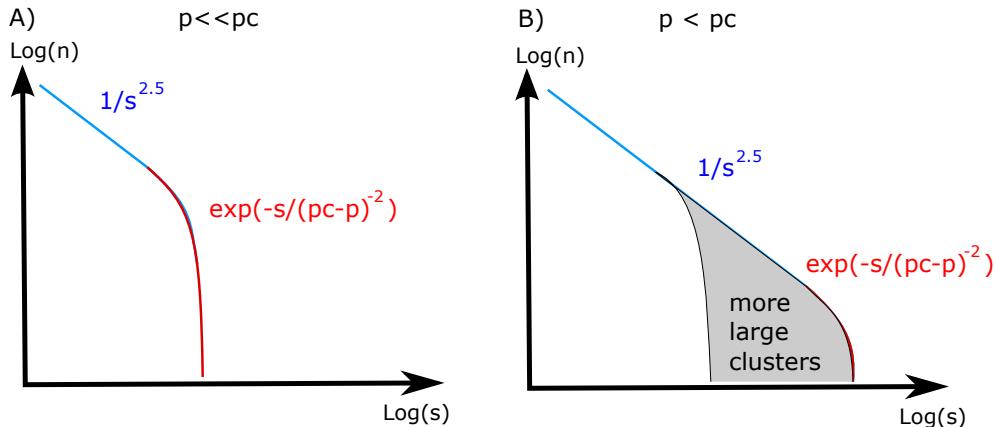


Figure 3.11: **Distribution of cluster sizes:** A) show distribution for p small, while B) show distribution for p much closer to p_c . Grey area marks new large clusters. If one gradually increases p one would see these larger clusters coming from the merging of smaller clusters ($p \rightarrow p_c$). Notice the cutoff diverge as $(p - p_c)^{-2}$ meaning that when we half distance to $p_c = 0.5$ then cutoff is moved up by a factor 4.

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 $p \rightarrow p_c$. Now assuming the above power law with cutoff $s_{p_c-p} = s_\Delta$. 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 $s n_s(p)$. When summing over all clusters the average cluster size becomes:

$$\begin{aligned} S(p) &\propto \sum_{s=1}^{\infty} s^2 n_s(p) / \sum_{s=1}^{\infty} s^1 n_s(p) \\ &\propto \sum_{s=1}^{\infty} s^2 n_s(p) \propto \int_1^{\infty} s^{2-\tau} \cdot \exp(-s/s_\Delta) \cdot ds \end{aligned} \quad (3.17)$$

$$(3.18)$$

where the denominator can be ignored when $\tau > 2$ (because the average is dominated by the lower end of the integral). We can now let the integration go to zero because we expect exponent $\tau < 3$ (if τ becomes larger equal to 3 then the integral at zero would diverge towards zero):

$$S(p) \approx s_\Delta^{3-\tau} \cdot \int_0^{\infty} z^{2-\tau} \cdot \exp(-z) \cdot dz = const \cdot s_\Delta^{3-\tau}. \quad (3.19)$$

where we use the continuous limit because we are concerned with 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. 3.10). Therefore

$$2\tau - 6 = -1 \rightarrow \tau = \frac{5}{2}, \quad (3.20)$$

thus obtaining the cluster size distribution:

$$n_s(p) \propto \frac{1}{s^{5/2}} \cdot \exp(-s \cdot |p - p_c|^2). \quad (3.21)$$

See Fig. 3.11. 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

$$\tau = 3 - \sigma \cdot \gamma \quad (3.22)$$

is obtained. In general the scaling of cluster sizes (eq. 3.21) is one example of a power law distribution augmented by a cut-off function that define the behavior 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:

$$n(s) = s^{-\tau} \cdot f(s \cdot |p - p_c|^\sigma) \quad (3.23)$$

$$\rightarrow s^\tau \cdot n(s) = f(x) \text{ with } x \equiv s \cdot |p - p_c|^\sigma. \quad (3.24)$$

This means that one could 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 plot would then directly reveal the function $f(x)$.

This numerical approach is called *data collapse* and allows us to estimate both p_c , σ , and naturally the cutoff function f . Also by scaling out the main power law dependence $1/s^\tau$ the y-axis of the plot can be greatly compressed and eventual deviations from expected τ can be precisely quantified.

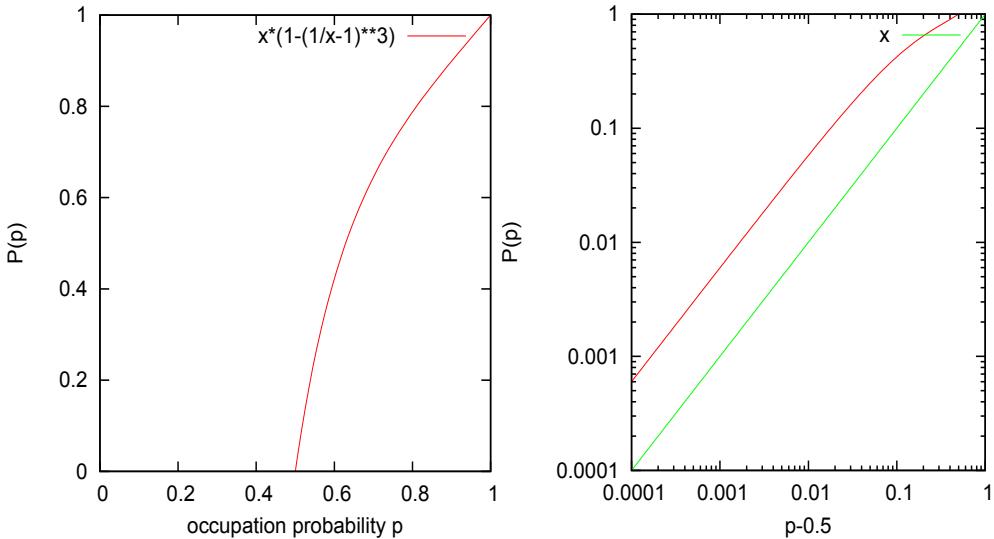


Figure 3.12: Scaling above p_c for Bethe Lattice: Strength $P(p)$ of the infinite cluster in Bethe lattice with $z = 3$, where P is the fraction of sites contained in the infinite cluster and p is the occupation probability. The right-hand side shows a typical scaling plot (note the double-logarithmic axes), allowing one to extract the behavior as one moves very close to $p_c = 0.5$ (for question 2.2).

Apart from the above exponents, γ , σ and τ that characterize various aspects of the size distribution of clusters (average, cutoff, and slope) there are two other exponents that are often used in percolation. The first is something that characterizes the density of the clusters, in fact, it is defined via the density of the infinite cluster for $p > p_c$:

$$P(p) = (p - p_c)^\beta \quad (3.25)$$

Formally $P(p)$ is the probability that a random point belongs to the infinite cluster. $P(p)$ is also called the strength of the infinite cluster. β typically takes

values between 0 and 1, a larger β means that its density at a given $p - p_c$ is smaller (because $p - p_c < 1$). The $\beta = 1$ exponent for the Bethe lattice is derived in a question and the associated scaling of $P(p)$ is shown in Fig. 3.12. Mini tutorial: If cluster size distribution is $p(s)$, what is then the distribution for cluster sizes when selecting random points and then counting cluster sizes associated with each of these points?

3.2.2 Renormalization & Correlation length exponent

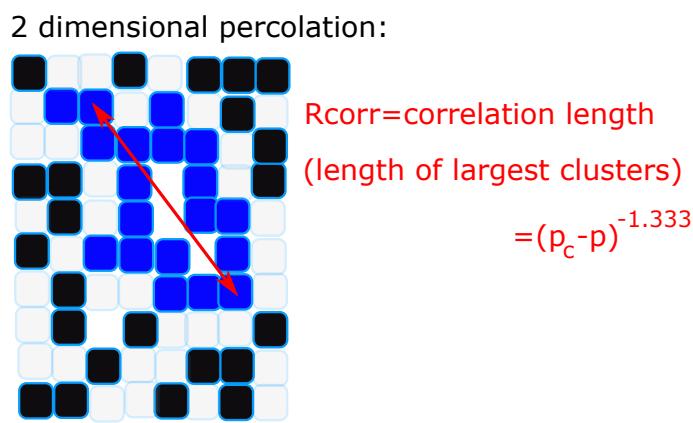


Figure 3.13: **Correlation length illustration:** The blue cluster is the largest cluster and the correlation length is the distance across it. Notice that in the shown percolation we use the 4-neighborhood criteria: Only sides above and beside each other are connected; sites that are diagonal from each other are not connected.

Beyond the critical exponents discussed above, there are other critical exponents in percolation — the correlation length exponent ν . The correlation function $g(r)$ is the probability that one occupied point is within the same cluster as another occupied point when these points are separated by a linear distance r :

$$g(r) \propto \exp(-r/R_{corr}) , \quad (3.26)$$

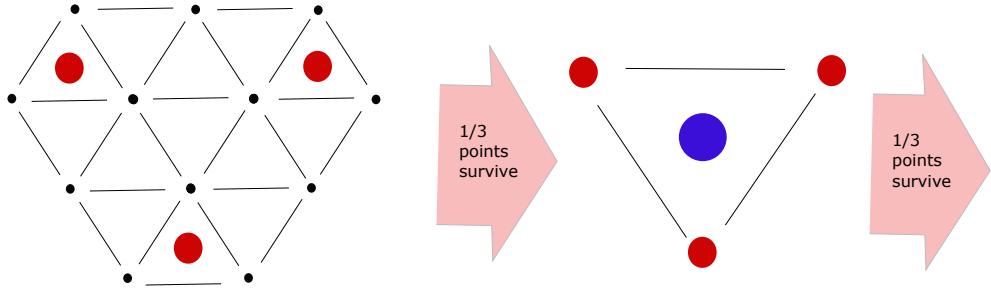
In Fig. 3.13 we illustrate the correlation length emphasizing that it is the linear dimension of the largest clusters at the given value of p . Here, for pedagogical reasons, “exp” is used as a name for a function with a scale (in fact the actual cutoff function has another form). The correlation length represents the maximal distance between two sites that belong to the same cluster. The correlation length scales as

$$R_{corr} \propto |p - p_c|^{-\nu} . \quad (3.27)$$

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⁾.

¹Percolation is characterized by a number of exponents, $1/\sigma$, γ , ν , τ and β (see question

A) Renormalization of triangular lattice, close to pc:



B) Renormalization transformation



Figure 3.14: Renormalization of 2-D site percolation on triangular lattice: At each step one moves to a larger scale by coarse-graining 3 lattice points into one, assigning this supersite to be occupied or unoccupied determined by the majority rule.

In Fig. 3.14 we argue for the correlation length exponent ν of about $4/3$ using a simple renormalization argument. In the figure, we imagine subsequent coarse-graining of 3 nodes into one, while at the same time changing p to a slightly different p while correlation length will decrease by an appropriate factor $b = \sqrt{3}$ (Because the density of new points is a factor 3 smaller, the corresponding length scale is the square root of this factor larger). Coarse graining 3 cells to one by majority rule (panel B), with each having probability p to be occupied implies that the new unit is occupied with renormalized probability given by the renormalization transformation

$$p' = p^3 + 3p^2(1-p)$$

This we solve for a fixed point $p' = p$ and obtain $p = 1/2$ as the interesting fixed

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.

point corresponding to the critical transition. If $p > 1/2$ then p' will diverge even more from $1/2$ and subsequent iterations will lead to fixed point $p''' = 1$. If $p < 1/2$ then p' will be even smaller and subsequent iterations will lead to fixed point $p''' = 0$. The sequence of p, p', \dots, p''' is called the *renormalization group flow*. The correlation length in old lattice was $R_{corr} = |p_c - p|^{-\nu}$ which should be related to correlation length in coarser lattice $R'_{corr} = |p_c - p'|^{-\nu}$ by a simple rescaling $R'_{corr} = R_{corr}/b$. That is, we have the relation

$$\begin{aligned} |p_c - p|^{-\nu}/b &= |p_c - p'|^{-\nu} \Rightarrow \\ b &= (dp'/dp)^\nu \Rightarrow \\ \nu &= \frac{\ln(b)}{\ln([dp'/dp]_{p=1/2})} = \frac{\ln(\sqrt{3})}{\ln(\sqrt{3}/2)} = 1.355 \end{aligned} \quad (3.28)$$

where we in last derivation use $dp'/dp = -6p^2 + 6p = -6/4 + 6/2 = 3/2$ when $p = 1/2$. The obtained exponent compares remarkably well with the exact result of $\nu = 4/3 = 1.33333$.

Now the scaling exponents are expected to be independent of details such as the type of lattice. But the renormalization transformation will, of course, depend on the lattice and will change in the case of a normal square lattice where coarse-graining is in blocks of four.

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 3.1: **Exponents for the Ising model and for percolation in two, three and infinite dimensions (Bethe lattice).** Notice that for the Ising model the exponent relates to temperature, $|T_c - T|$, whereas one use $|p_c - p|$ is the variable in percolation. In the Ising model the order parameter was the magnetization, in percolation it is the probability of belonging to the largest cluster. Chin-Kun Hu, 1984 suggested that the Ising model is related to bond percolation with bond probability $p = 1 - \exp(-2J/k_B T)$. For the Bethe lattice the correlation length exponent ν is assumed to be the same as for high dimensional percolation ($= 1/2$), see also argument by [23].

Questions:

3.3) Go through the following argument associated with 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 move into the high-density region at $p > p_c$. This would be analogous to exploring 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.

$$\begin{aligned} P(p) &= (\text{Probability that site occupied}) \\ &\quad \cdot (\text{Probability that at least one neighbour leads to infinity}) \\ &= p \cdot (1 - Q^3) \end{aligned}$$

where Q is the probability that an arbitrary neighbor site is not connected to infinity. For $z = 3$ then:

$$\begin{aligned} Q &= (\text{Probability site empty}) \\ &+ (\text{Probability site occupied}) \cdot (\text{Probability no subbranch leads to infinity}) \\ &= (1 - p) + p \cdot Q^2 \rightarrow \\ Q &= \frac{1}{2p} \cdot (1 \pm \sqrt{(2p - 1)^2}) \rightarrow \\ Q &= 1 \text{ or } Q = \frac{1}{p} - 1 \end{aligned} \tag{3.29}$$

Prove this by insertion of the 2 solutions! The strength $P(p)$ is therefore

$$P(p) = p \cdot \left(1 - \left(\frac{1}{p} - 1\right)^3\right) = \frac{p^3 - (1-p)^3}{p^2}$$

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. 3.12). Thus, it becomes more and more difficult to 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").

3.4) Compare exponents in Table 3.1. 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 the corresponding percolation?

What does it mean that the correlation length exponent is smaller for the Ising model than for the corresponding percolation?

(QLesson: $(p_c - p)$ raised to a big negative exponent means that objects diverge to a larger size than if the exponent were small ($1/(p_c - p)^2 > 1/(p_c - p)$). Also, the cluster size distribution becomes steeper with dimension.)

3.5) Argue for a renormalization transformation $p' = p^4 + 4 p^3 \cdot (1-p) + 2 p^2 \cdot (1-p)^2$ for site percolation on a square lattice (draw which combinations of occupied sites among 4 sites in a square that should be occupied to lead to one occupied site at the next level). Derive the critical point and the correlation length exponent.

QLesson: The renormalization transformation and its convergence depend on the lattice. The square lattice makes it unclear what should be considered occupied after coarse-graining 4 sites into one. One renormalization step leads to gaining some of the original connecting paths and losing others. These mistakes are larger for the square lattice than for the triangular lattice.

Fractals are everywhere:

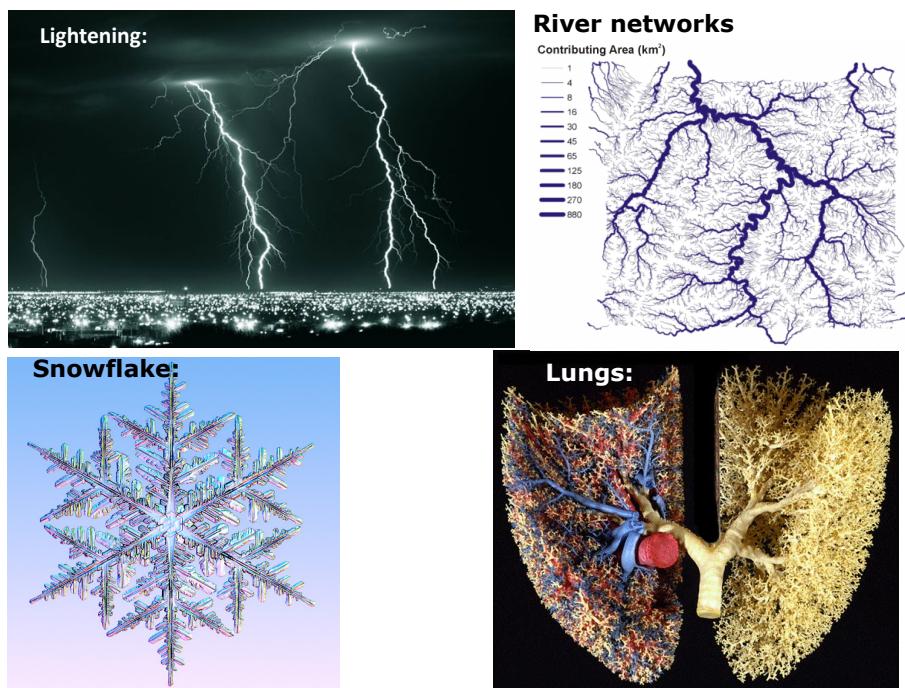


Figure 3.15: **Fractal examples:** Figure illustrates examples of fractals found in our surrounding nature and in ourselves. Fractals are objects with dimensions in between our usual integer dimensions, objects that fill a vanishing part of available space. For example allowing an efficient connection between many points in space and an exit point (river networks or lung network, or blood vessel network, etc.). Often fractals are hierarchical.

3.3 Fractals and Fractal Dimensions

Fig. 3.15 show examples of fractals, strange objects between dimensions: Lines of finite extension but infinite length, points of lines of excitations that fill zero fraction of space. We will first introduce fractals and discuss them in

connection with the spatial structure of the clusters in percolation. These clusters are in fact so full of holes that the density inside the very large clusters is approaching zero, see 3.13.

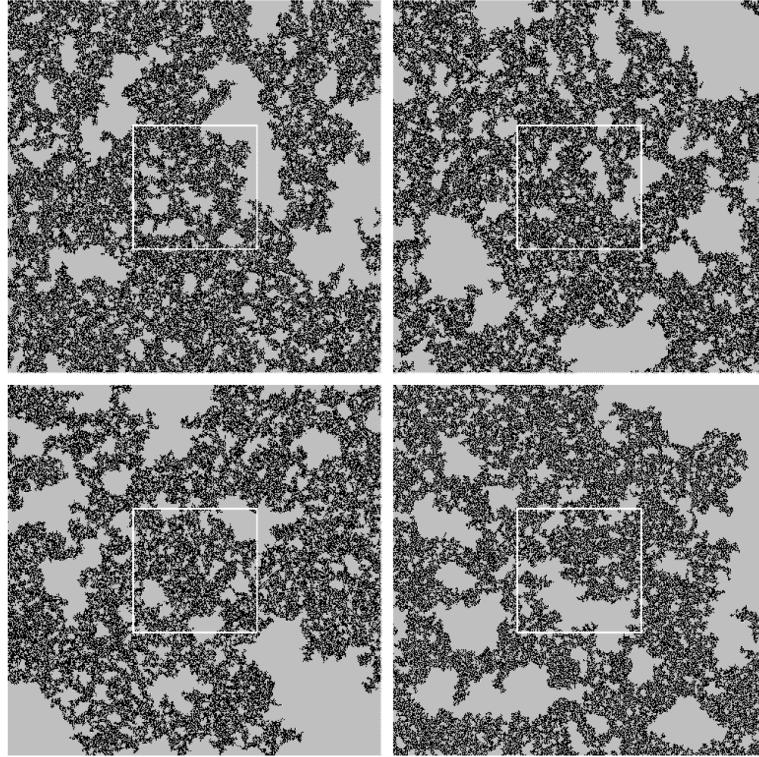


Figure 3.16: Infinite cluster in percolation: Figure shows percolating cluster at the critical point, in subsequent levels of detail. We will see how to calculate the fractal dimension of this object. From introduction to percolation theory by Armin Bunde and Jan .W. Kantelhardt.

3.3.1 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

$$M(\ell) \propto \ell^D \quad (3.30)$$

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=3. Similarly in 2 or 1 dimensional objects as seen Fig. 3.17. However, a object may often have holes and irregular boundaries, as for example with snow crystals. In that case

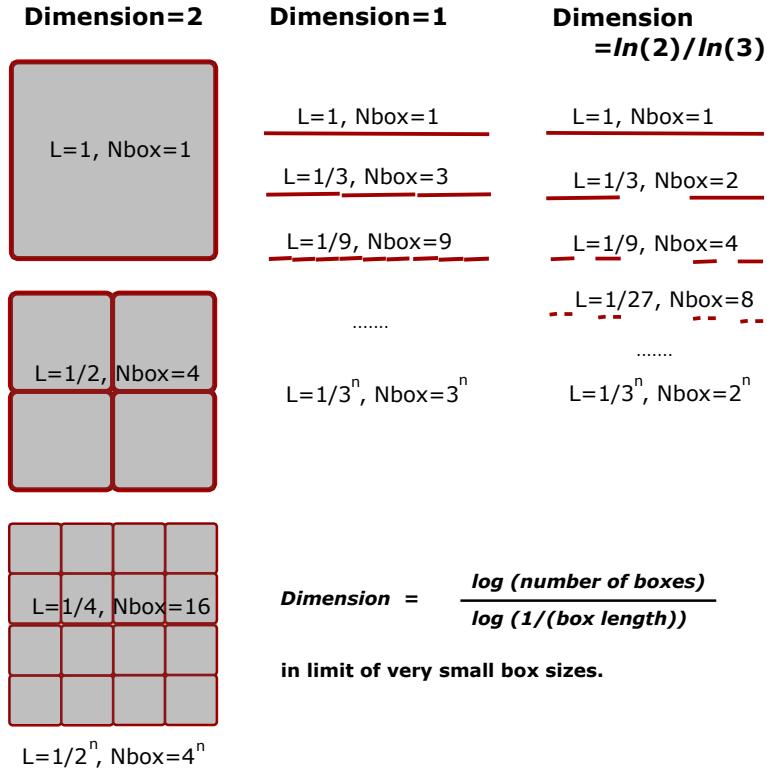


Figure 3.17: Dimensions of objects

the *fractal dimension* of the object could be smaller than three (or whatever dimension it is embedded in).

In some real world fractals, one measures them with smaller and smaller “measuring stick” $\epsilon \rightarrow 0$. 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) can be calculated by the scaling of the number of boxes

$$N(\epsilon) \propto \frac{1}{\epsilon^D} \quad (3.31)$$

needed to cover the object (the set of points) as function of the used box size ϵ . This type of box covering is illustrated in Fig. 3.19. Given $N(\epsilon)$ for a range of different box sizes ϵ , the dimension is then calculated as

$$D = \lim_{\epsilon \rightarrow 0} \left(\frac{\log(N(\epsilon))}{\log(1/\epsilon)} \right). \quad (3.32)$$

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.

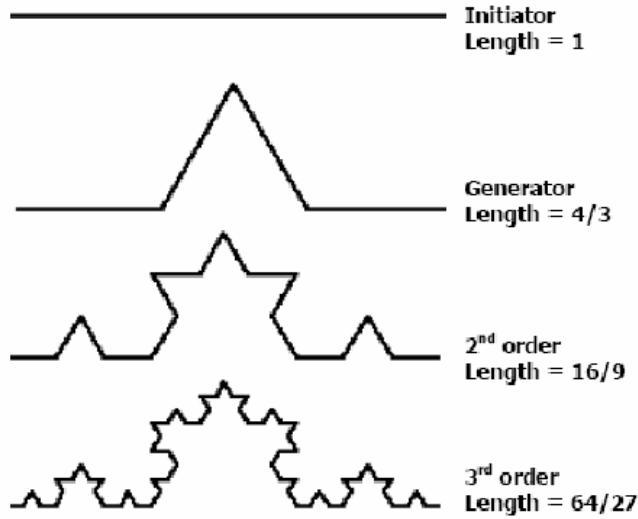


Figure 3.18: **Koch curve.** Illustration of how the length of the line increases as one considers finer and finer detail: When the "measuring stick" is three times smaller, the total length is four times larger. Thus after t steps one need 4^t boxes each of length $(1/3)^t$. This gives dimension $D = \ln(4^t)/\ln(1/(1/3)^t) = \ln(4)/\ln(3) = 1.26$.

This is illustrated for the Koch curve in Fig. 3.18, where one mathematically construct an infinitely wiggly line with dimension $D = 1.26$.

Notice that the limit $\epsilon \rightarrow 0$ means that when calculating D one naturally focus on the part of the object that has highest density. Imagine that the object under investigation consists of two sub-parts, one with dimension D_1 and one with dimension D_2 . Then

$$N(\epsilon) = \frac{1}{\epsilon^{D_1}} + \frac{1}{\epsilon^{D_2}} \quad (3.33)$$

which is dominated by highest value of D when $\epsilon \rightarrow 0$ (and this scaling is independent on the relative weight of the two sub-parts).

To put the dimension concept in perspective, then the coast of Great Britain (Fig. 3.19) is characterized by dimension $D = 1.2$, while the more convoluted coast of Norway is closer to $D = 1.5$.

Let us now return to percolation and its largest cluster at critical conditions. If we were to describe the dimension of this largest cluster, we might investigate larger and larger lattices, and explore the scaling of the percolating cluster. The fractal dimension of a 2-dimensional percolating cluster at critical conditions is scaling of the number of sites (small boxes) that is needed to

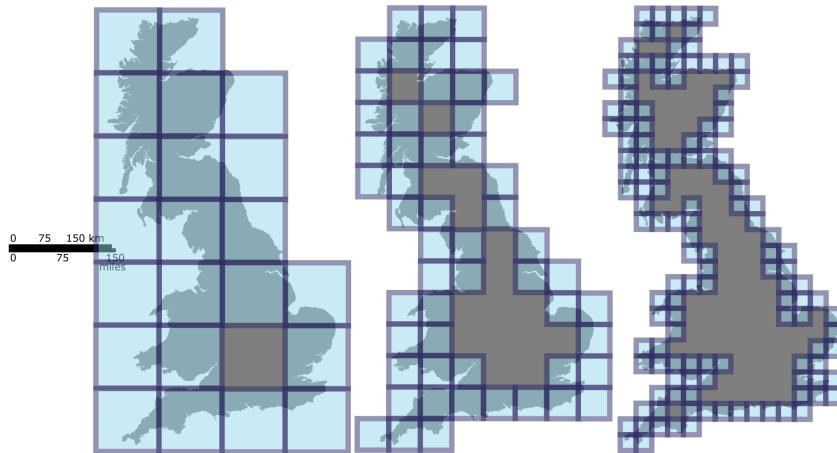


Figure 3.19: **Fractal dimension of Great Britain.** Illustration of the box-counting method on the coast of Great Britain, taken from Wikipedia, Prokofiev - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=12042116>

cover it at length scale ℓ :

$$M \propto \ell^D$$

where D is the fractal dimension. This scaling relation between mass and linear dimension can also be expressed in terms of known exponents.

Mini Tutorial:

when $p > p_c$ then the above scaling relation is only true up to a maximal length scale. What is this length scale? What happens above this length scale?

Consider 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

$$M \propto (p_c - p)^{-1/\sigma}. \quad (3.34)$$

reflecting the cutoff in the size distribution we also examined earlier. Similarly, the linear dimension of the largest cluster is given by its correlation length

$$R_{corr} \propto (p_c - p)^{-\nu} \Rightarrow (p_c - p) \propto R_{corr}^{-1/\nu}. \quad (3.35)$$

Thus we get the mass of the largest cluster in terms of its linear extension

$$M \propto R_{corr}^{1/\nu\sigma}. \quad (3.36)$$

The (fractal) dimension of the largest cluster at p_c therefore is

$$D = \frac{1}{\nu\sigma} = \frac{91}{48} \approx 1.90 < 2 \quad (3.37)$$

at the critical point. Thus at larger scales, the critical cluster will approach zero density. Only very few points will belong to the system-spanning cluster.

Close the p_c there will also be smaller clusters that are quite large. Locally they will look similar to the largest cluster and will thus also each be characterized by the dimension D . In contrast, then the dimension of all sites in all clusters clearly have $D_{tot} = d = 2$ because the density of occupied points is finite ($= p$).

Mini tutorial:

What is the density of the percolating cluster at p_c in an infinitely large two-dimensional system?

That the fractal dimension of the percolating cluster is 1.9 implies that the density of this large cluster scale with length scale l as $\rho = l^{1.9}/l^2 = l^{0.1}$. Thus when looking at critical percolation on the length scale

- $l = 10^{10}$ the density of the percolating cluster is $\rho = 0.10$
- $l = 10^{20}$ the density of the percolating cluster is $\rho = 0.01$
- $l = 10^{30}$ the density of the percolating cluster is $\rho = 0.001$

Thus the density approaches zero. This approach to zero means that the density at scale $l = 10^{30}$ is 0.1 when measured with sticks of length $l_s = 10^{20}$. That is this is the density of l_s boxes that are filled inside the $l = 10^{30}$ system. Similarly the density at scale of $l = 10^{20}$ is 0.1 when measured with sticks of length $l_s = 10^{10}$. In this way, each scale looks the same when viewed on the corresponding scale, but when these subsequent images are put together, the density on the large scale drops as it separates from the smallest scale.

The mass and dimension of the largest cluster M also connect to the exponent β . β is defined in terms of the density of the infinite cluster when $p > p_c$. Above p_c there will be a correlation length, below which the percolating structure looks fractal. It turns out that this correlation length scale is in the same way above and below p_c . This allows us to calculate the mass for distances up to the size set by the correlation length (eq. 3.27), with densities deduced from the exponent β :

$$M \sim R_{corr}^d P(p) \propto R_{corr}^d (p - p_c)^\beta \propto R_{corr}^{d-\beta/\nu} \quad (3.38)$$

where d is the embedding dimension and we use eq. 3.27 to express everything in terms of the correlation length. Notice that naturally $d = 2$ for percolation in two dimensions. 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

$$\nu = (\beta + 1/\sigma)/2 . \quad (3.39)$$

Thus, for a larger cluster size cutoff $1/\sigma$ one also has a larger correlation length ν .

Mini tutorial:

Given that $\beta = 5/36$ and $\sigma = 36/91$ for 2-dim. percolation, calculate ν .

The percolating cluster in a 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 sub-part (called "backbone") are reported to have dimension $D_{\text{backbone}} = 1.13$ (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 a dimension of less than 1²)

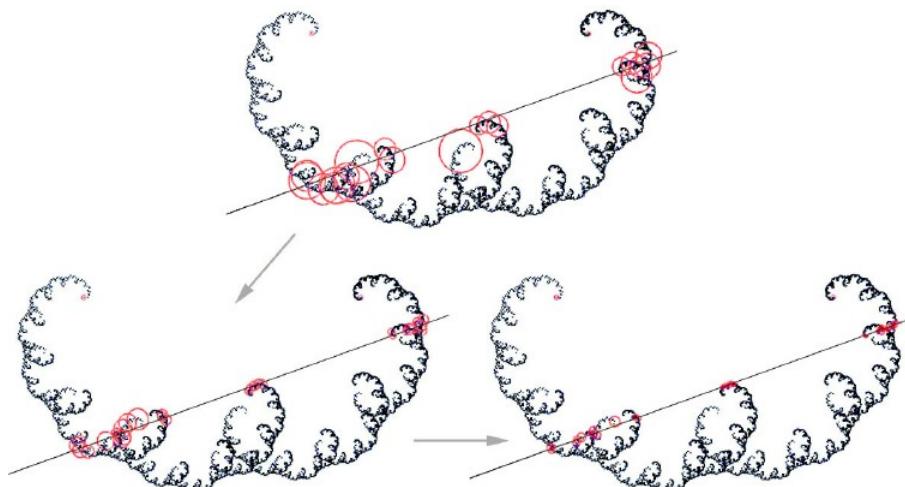


Figure 3.20: **Intersecting fractals.** Fractal subset of a fractal, here a fractal with dimension D_A is intersecting with another fractal $D_B = 1$. Fig. From Jozsef Vass (2013).

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

$$D(A \text{ and } B) = D(A) + D(B) - D_{\text{space}} \quad (3.40)$$

where dimension $D_{\text{space}} = d$ is the space they both are embedded independently inside (S. Miyazima and H.E. Stanley, Phys. Rev. B 35, 8898, 1987). See fig.

²The number of these so-called red bonds scales as $n_{\text{red}} \propto \frac{1}{p-p_c}$ for $p > p_c$ (Coniglio, 1982) and thus scale with correlation length $R \propto (p - p_c)^{-\nu}$ as $n_{\text{red}} \propto R^{1/\nu}$, i.e. $D_{\text{red}} = 1/\nu = 3/4$ for two-dimensional percolation.

3.20. This is proven by covering the total space with $1/\epsilon^{D_{space}}$ boxes. Of these many boxes, $1/\epsilon^{D(A)}$ are part of object A, while $1/\epsilon^{D(B)}$ boxes are part of object B. Thereby the probability that one box contain something from object A, respectively object B is

$$\begin{aligned} P(A) &= \frac{(1/\epsilon^{D_A})}{(1/\epsilon^{D_{space}})} \\ P(B) &= \frac{(1/\epsilon^{D_B})}{(1/\epsilon^{D_{space}})} \end{aligned} \quad (3.41)$$

Assuming independent coverage, the intersection fraction is multiplying probabilities $P(A \text{ and } B) = P(A) \cdot P(B)$, or:

$$\frac{\epsilon^{D_{space}}}{\epsilon^{D(A \text{ and } B)}} = \frac{\epsilon^{D_{space}}}{\epsilon^{D(A)}} \cdot \frac{\epsilon^{D_{space}}}{\epsilon^{D(B)}} \quad (3.42)$$

from which eq. 3.40 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 $D(dust) = 1.5 + 1 - 2 = 0.5$. We will see later that the distance between these dust particles are power law distributed³.

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 dimension D . What is D of this “rough” mountain range?

3.3.2 Fragmentation

Let us now consider the above equations from the perspective of a simple model for fragmentation. In some approximation, this can be rephrased as an initial impact that causes of a lot of cracks, followed by subsequent merging of straight cracks. This merging or absorption of cracks reflects the fact that when two cracks meet then the first crack stops the propagation of the second

³In general a colored random walk in one dimension with Hurst exponent H (see econophysics chapter, $\Delta W^2 \propto t^{2H}$) 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 $D = 2$

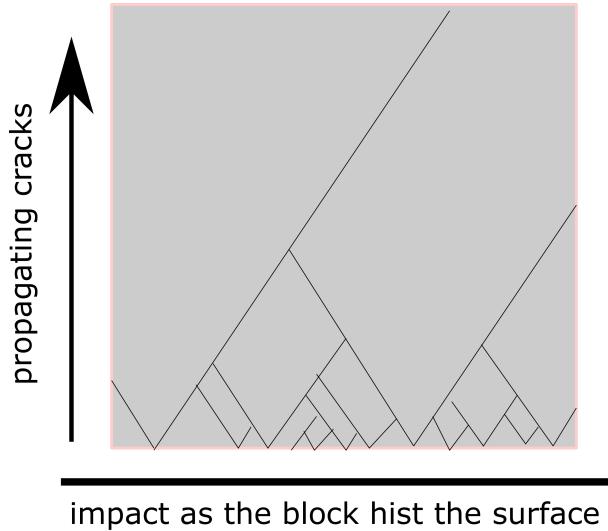


Figure 3.21: **Simple fragmentation model.** Crack propagation on a plane, but created by impact on its side. Cracks density declines as $1/t^{0.5}$ where t is distance from the impact surface.

one. This is illustrated in Fig. 3.21, with similar dynamics in some cellular automata models, see Fig. 3.22.

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. 3.21, 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 is counted as one point. Because we assume that one initiates cracks with constant density along the impact surface and each fragment has a point on 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 [24]. I.e. assume that

$$n(s) = \frac{1}{s^\tau} \cdot f\left(\frac{s}{L^D}\right) \quad (3.43)$$

where f is some cutoff function that drops to zero when the cluster becomes so big that its linear dimensions are comparable with the whole system L

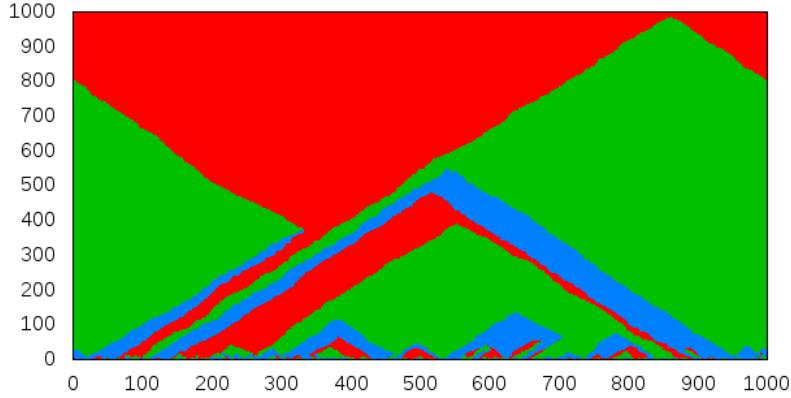


Figure 3.22: Coarsening in a rock-paper-scissor dynamics in one dimension. One initially seeds each site with one of the 3 different species types, a “rock”, “scissor” or a “paper”. They can grow into each other as defined by the normal rules of the rock-paper-scissor game. Populations of one species can be separated by other species, forming antagonistic boundaries that move either left or right. Boundaries between species move left or right and are 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 an invading “paper”.

(Because D is the dimension of the cluster, then s feels the boundary exactly when it reaches the size L^D).

We can now express:

- Total cluster mass = (average cluster mass) \times (number of clusters)

Thus 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

$$\begin{aligned}
 L^{D_{num}} \cdot \int_1^{L^D} s \cdot n(s) ds &= L^{D_{tot}} \\
 L^{D_{num}} \cdot [s^{2-\tau}]_1^{L^D} &= L^{D_{tot}} \\
 L^{D_{num} + (2-\tau)D} &= L^{D_{tot}} \\
 D_{num} + (2-\tau) \cdot D &= D_{tot} \\
 \tau = 2 - \frac{D_{tot} - D_{num}}{D}, \tag{3.44}
 \end{aligned}$$

where we assumed that only the upper end of the integral counted, i.e. that $\tau < 2$. If τ is larger, one instead has to focus on the lower end of the integral.

For the fragments in Fig.3.21, then $D = 2$, $D_{tot} = 2$ and $D_{num} = 1$ giving

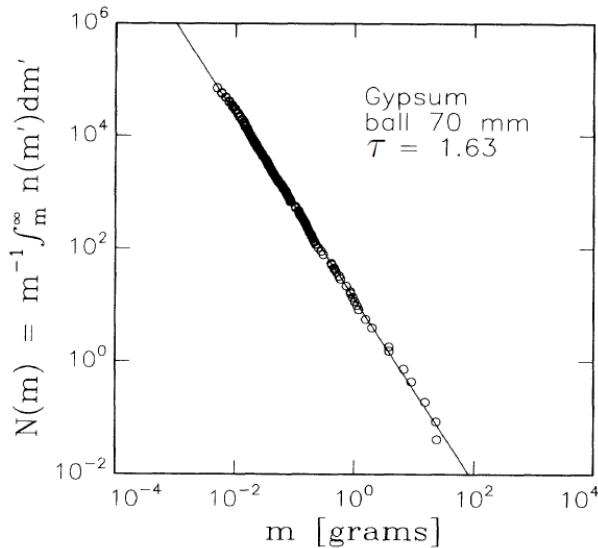


Figure 3.23: **Fragment size distribution.** Measured from fragmentation of gypsum blocks that were dropped on the floor.

fragment size distribution

$$n(s) \propto \frac{1}{s^{3/2}} \quad (3.45)$$

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

$$n(s) \propto \frac{1}{s^{5/3}} . \quad (3.46)$$

This is close to the fragments size distribution $1/s^{1.63}$ that was measured by [25], see Fig. 3.23.

Noticeably, meteors are distributed with a 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, assuming that asteroids larger than 10km impact at a rate of once per 64 million years. Thus it is an extremely uncertain estimate.

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?

The above dimensions can also be assigned to percolation, see Fig. 3.24, but in that case the size distribution $\tau > 2$ which makes the above integration depend on lower bound, and not the upper end of the integral. This complicates the approach.

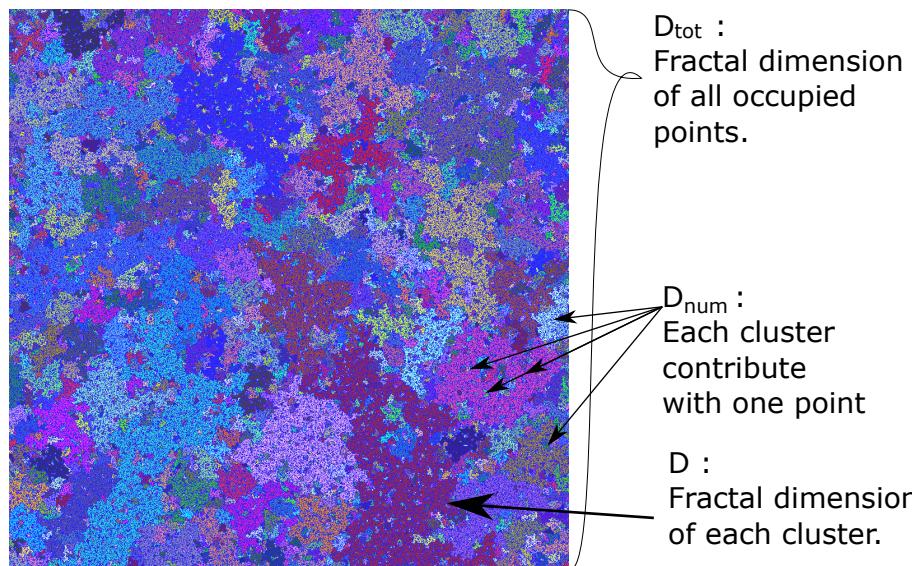


Figure 3.24: **Dimensions for cluster sizes.** A set of clusters can be assigned different fractal dimensions, which together can be used to calculate cluster sizes. Notice that $D_{num} = 2$ for percolation because there is a finite chance that each point belong to a cluster of size 1. This chance is equal $p \cdot (1 - p)^4$ as the size one cluster demand that the 4 nearest neighbors need to be unoccupied.

Questions:

- 3.6)** Consider dust on a line, with points distributed with dimension $D_{num} = D_{dust}$. Thus the line could be subdivided into segments of dimension 1, each associated with one “dust particle”. 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).

- 3.7)** What is the fractal dimension of the intersection of a line and a two-dimensional percolating cluster at the critical value p_c ? What is the distribution of distances between subsequent intersections?

Qlesson: Fractal dimension is obviously (?) smaller than 1. Why?

- 3.8)** Formulate a cellular automata 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)

3.9) 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 the lower limit. In that case, the largest cluster doesn't contribute to the integral compared to all the smaller ones. The mass of all but the largest cluster is $L^{D_{tot}} - L^D$. Now argue for the identity

$$\begin{aligned} L^{D_{tot}} - L^D &= L^{D_{num}} \int s^{1-\tau} f(s/L^D) ds \\ &= L^{D_{num}} \cdot (const - L^{D \cdot (2-\tau)}) \end{aligned}$$

stating that the largest cluster is given by $D_{tot} = D_{num}$ with a "const" factor that depends on the small-scale cutoff. The subtracted part comes from the small correction due to the upper end of the integral, i.e. the largest cluster. Argue that when $\tau > 2$ then $D_{tot} = D_{num}$ and

$$\tau = 1 + \frac{D_{num}}{D} \quad (3.47)$$

For 2-d percolation, $D_{num} = D_{tot} = 2$ also follows from the fact that there is a finite density of clusters of size 1; I.e. the probability that a site belongs to a cluster of size 1 is $p \cdot (1-p)^4$ which is indeed finite. Equation first derived by Greg Huber.

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.

3.4 Directed percolation

Directed percolation (DP) is the first dynamic process in these notes. DP 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 has such an active neighborhood, the probability of becoming active, at the 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 noninfected state. On the other hand, if p is large, then all sites will eventually become active (infected) nearly all the time.

Directed percolation is also often associated with the following stochastic equation (Reggeon field theory):

$$\frac{dn}{dt} = \frac{d^2n(x,t)}{dx^2} + b \cdot n(x,t) - d \cdot n(x,t)^2 + \eta(x,t) \quad (3.48)$$

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

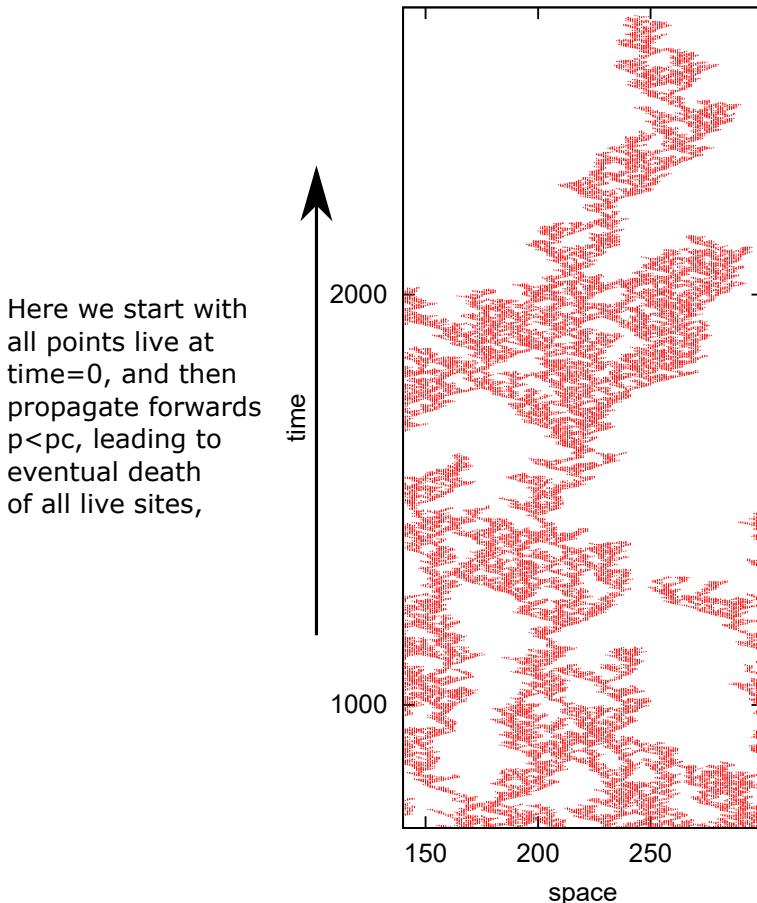


Figure 3.25: Directed percolation. Each site can give offspring to itself and its two nearest neighbors. This process has two correlation lengths, one along the time axis $T(p) \propto 1/|p_c - p|^{1.7}$ and a shorter one along the spatial dimension $X(p) \propto 1/|p_c - p|^{1.1}$. In addition, the process is characterized by one density exponent β , with the density of active sites at long times after initiation, $\rho \propto |p - p_c|^\beta$, $\beta = 0.27$. This is also the scaling for the probability that a random point belongs to the infinite cluster.

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 creates 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 depends

on the lattice, but the exponents characterizing the phenomenon are 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 (ν_{\parallel}) than in the spatial direction (ν_{\perp}).

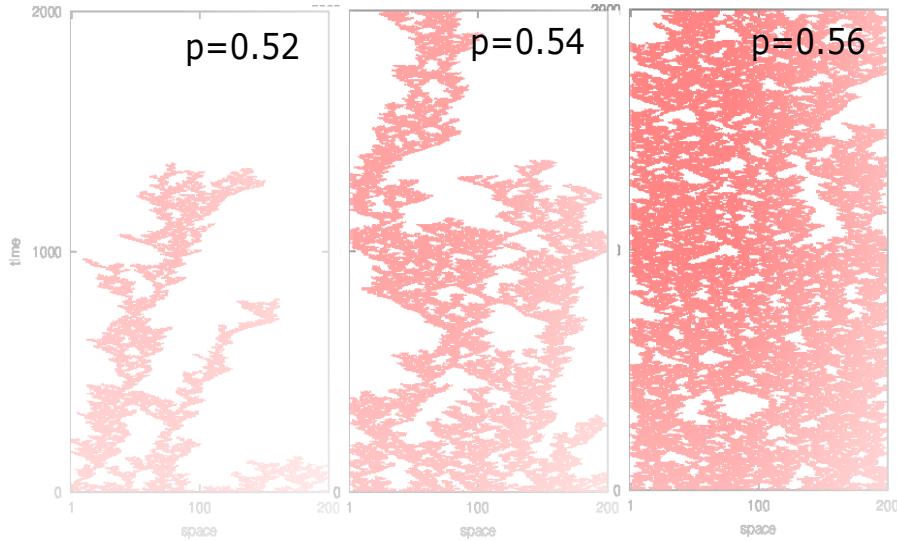


Figure 3.26: **Directed percolation around p_c :** Each site can give offspring to itself and its two nearest neighbors with probability p . There is a threshold value of p for which propagation will continue.

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

$$t_{corr}(p) \propto (p_c - p)^{-\nu_{\parallel}} \quad \text{with} \quad \nu_{\parallel} = 1.733 , \quad (3.49)$$

The space-like correlation length (width) diverges as

$$x_{corr}(p) \propto (p_c - p)^{-\nu_{\perp}} \quad \text{with} \quad \nu_{\perp} = 1.097 \quad (3.50)$$

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

$$\rho \propto (p - p_c)^{\beta} \quad (3.51)$$

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 table 2.1).

An exponent $\beta = 0.27$ for the infinite cluster means that for $p > p_c$ the chance that a site is alive within the branching three scales as $(p - p_c)^{0.27}$.

Thus if one reduces $p - p_c$ by a factor 16, then the chance that a site is alive becomes a factor ~ 2 smaller.

Notice, that directed percolation cannot be solved analytically. 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?

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 3.2.

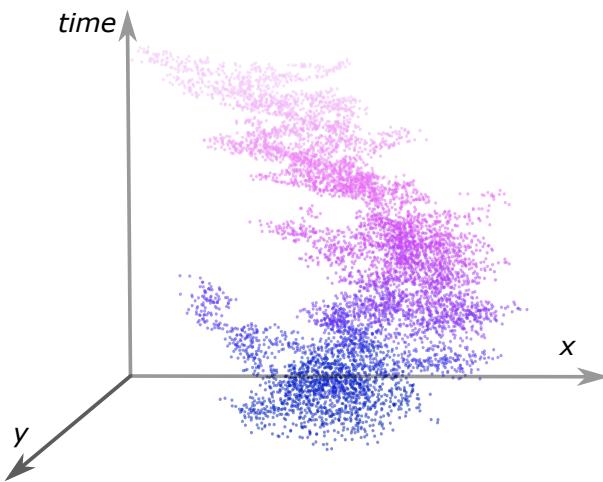


Figure 3.27: **Directed percolation in 2+1 dimensions.** From Anton G. Golles, master student at Copenhagen University.

Exponent	Mean-field	1-d DP	2-d DP
β	1	0.276	0.58
ν_\perp	1/2	1.097	0.73
ν_\parallel	1	1.734	1.2
δ	1/2	0.159	0.45

Table 3.2: **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 $\chi = 1/2$ (see later).

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. 3.26 that shows directed percolation on a 1-dimensional geometry, developing in time as activity spreads and/or dies out. Fig. 3.25 further examines directed percolation close to its critical point by highlighting a small sub part of the spreading process.

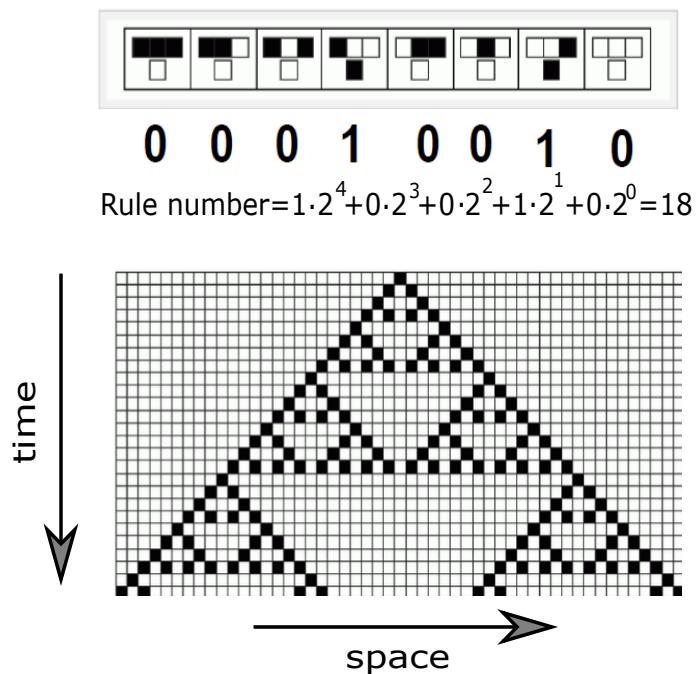


Figure 3.28: **Cellular automaton rules.** Rule 18 is an example of a cellular automaton where the updated site value depend on itself and two nearest site [26]. In rule 18 a site dies (becomes zero), except if just one of its neighbors are alive at the timestep before. The rule table is outlined in 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 resulting pattern is fractal, with space expanding proportional to time leaving recurrently larger patches of empty triangles behind. The rule can be made stochastic by assigning all sites dead except live sites for the 100 and 001 neighborhoods with probability p . p can then be fine-tuned to give the same large-scale properties as critical directed percolation in one dimension. The scaling exponents for this will then be spreading slower in space than the deterministic counterpart, reflecting the exponents of directed percolation.

In the context of directed percolation it is worthwhile mentioning Cellular automata [27, 28, 29, 26]. These are updating discrete rules acting on a lattice.

Typically one allows each lattice site to be either on or off (Boolean=binary). The update is discrete, with all sites being updated synchronously. At any timestep, each cell is updated according to itself and its nearest neighbors. An example of an update rule on a 1-D lattice is shown in Fig. 3.28. Many other rules exist, and for nearest neighbor interactions in 1-D, these rules can be labeled by one of 256 numbers [29]. However, in higher dimensions, there will be many more rules, including Conway's famous Game of Life [28]. The connection to directed percolation comes about when we add a probabilistic component to the cellular automata.

In particular, 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. 3.28). Notice that many other cellular automata could be considered. With nearest neighbors in one dimension (three sites input), there are $2^3 = 8$ inputs that each should be defined as one of 2 outputs, making $2^8 = 256$ deterministic cellular automata. In 2-dimension there are 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 be given in the formalism of Fig. 2.15?

Questions:

3.10) 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 $1/s^\tau$ that is relatively broad $\tau < 2$.

3.11) 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 $t = e^{-\nu_{\parallel}}$ is $\rho \propto e^\beta = t^{-\beta/\nu_{\parallel}}$. 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.

3.12) What rule number would standard directed percolation correspond to? (see Fig. 3.28)

Qlesson: Reminder that Directed percolation is a stochastic version of a cellular automaton. One could also make other rules stochastic.

3.13) Consider directed percolation, and convince yourself about the following scaling arguments for some relevant dimensions. The scaling of the mass m of the infinite cluster up to a correlation length $\ell_{\parallel} \sim e^{-\nu_{\parallel}}$: $m \propto e^{\beta - \nu_{\parallel} - \nu_{\perp}} \sim \ell^{(\nu_{\parallel} + \nu_{\perp} - \beta)/\nu_{\parallel}}$. Thus the dimension counted with a length measured longitudinally is $1 - \beta/\nu_{\parallel} + \chi$.

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 quantify the many scale-free phenomena in our surrounding world. Scale-free is the same as stating that quantities are power law distributed:

$$\frac{P(s \cdot k)}{P(k)} = \frac{P(s)}{P(1)} \Rightarrow \log(P(x)) = a \cdot \log(x) \Rightarrow P(x) = x^a$$

- This chapter presented explored scaling behavior as $p \rightarrow p_c$ in analogy with $T \rightarrow T_c$ for the Ising model. The focus was here on the power law distribution of cluster sizes, and how its cutoff changed as $p \rightarrow p_c$, exemplified for the Bethe lattice.
- Fractal dimensions characterize scale-free spatial structures. It is defined in terms of the scaling of the number of boxes with the size of each box:

$$D = \lim_{\epsilon \rightarrow 0} \left(\frac{\log(N(\epsilon))}{\log(1/\epsilon)} \right) .$$

Supplementary reading:

Christensen, Kim, and Nicholas R. Moloney. *Complexity and criticality*. Vol. 1. World Scientific Publishing Company, 2005.

Stauffer, Dietrich, and Amnon 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 4

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

4.1 Introduction

We have seen a lot of power laws, and examples of scaling in nature. However, we have not come up with some unifying understanding of how they arise. That is, if they comes from being close to a critical point in some sort of percolation or Ising mode, then what would secure that nature was close to such a point. This chapter is devoted to one such type of dynamics, the topic of self organization towards a critical point. We will see that such mechanism typically demand some separation of timescale in the dynamics, and some discrete underlying dynamics like a stick-slip motion in a earthquake fault.

Noticeably, each of these self organized critical models give different exponents dependent on details and dimensions. But some universality will emerge. Especially we will see that they in infinite dimension/random neighbor limit all tend to behave like a critical branching process. As the critical branching process, they will have some characteristic scaling behaviour that approaches $1/s^{3/2}$. Therefore we in the following first discuss critical branching and first return of random walkers. Subsequently we introduce 2 widely different types of Self Organized Critical (SOC) models and will see that they indeed give the $1/s^{3/2}$ scaling of the central avalanche distribution.



Figure 4.1: Avalanche. The central phenomena in self-organized critical phenomena are avalanches. These consist of causal sequences of toppling events that redistribute and relax local stress. The avalanches are separated by periods of very little or no activity.

4.2 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

$$x = \sum_{i=1}^{t/\tau} \eta(i) , \quad (4.1)$$

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 $t=0$)

$$\begin{aligned} \langle x^2 \rangle &= \left\langle \sum_{i=1}^{t/\tau} \eta(i) \sum_{j=1}^{t/\tau} \eta(j) \right\rangle \\ &= \sum_{i,j=1}^{t/\tau} \langle \eta(i) \eta(j) \rangle \\ &= \sum_{i=1}^{t/\tau} \langle \eta(i) \eta(i) \rangle \\ &= \sum_{i=1}^{t/\tau} \ell^2 = (t/\tau) \ell^2 = (\ell^2/\tau) t = 2 \cdot D \cdot t , \end{aligned} \quad (4.2)$$

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 4.2 where the walkers are walking up or down along the x -axis. Importantly the dimensions of the diffusion constant

$$[D] = \frac{\text{meter}^2}{\text{second}}, \quad (4.3)$$

reflecting that it results from multiplying a velocity with the distance travelled between the times when direction of movement can be changed.

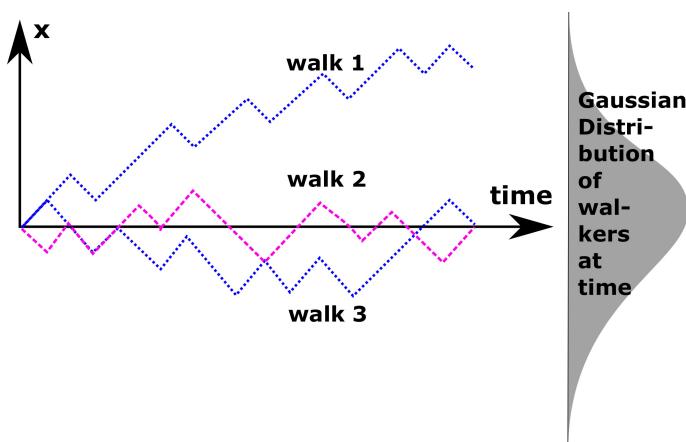


Figure 4.2: **Trajectories of random walkers.** Random walkers starting at position $x = 0$ at time $t = 0$ and ending at some position x . The final distribution of many walkers will be Gaussian in space, with width $\sigma \propto \sqrt{t}$.

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 across position x during the time τ . The current of particles across position x is (see Fig. 4.3):

$$\begin{aligned} J &= \frac{1}{2} \left(\rho(x - \ell/2) \frac{\ell}{\tau} - \rho(x + \ell/2) \frac{\ell}{\tau} \right) \\ &= \frac{\ell}{2\tau} \left((\rho(x) - \frac{\ell}{2} \cdot \frac{d\rho(x)}{dx}) - (\rho(x) + \frac{\ell}{2} \frac{d\rho(x)}{dx}) \right) \\ &= -D \cdot \frac{d\rho(x)}{dx} \end{aligned}$$

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).

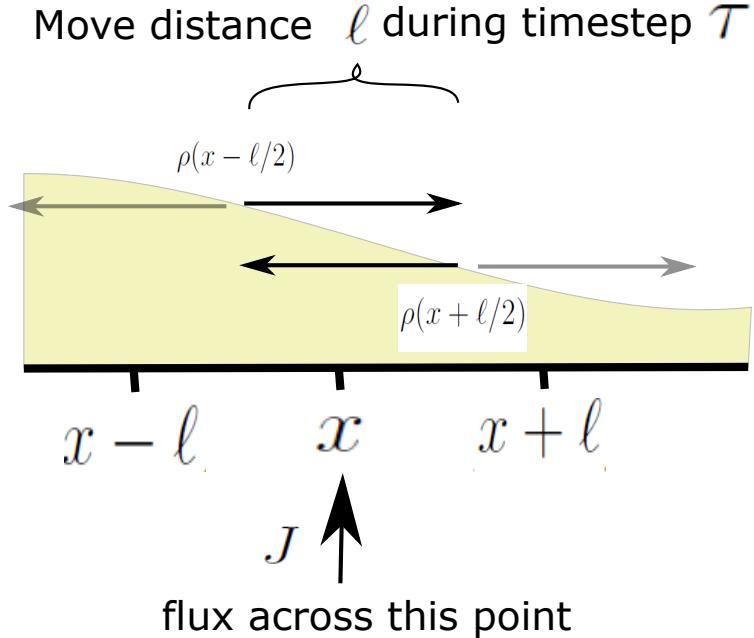


Figure 4.3: **Diffusion equation:** Flux(current) at point x gets contribution from distance of step length ℓ within time interval of randomization time τ .

The change in density, is subsequently given by the difference between what moves in and what moves out:

$$\frac{d\rho}{dt} = -\frac{dJ}{dx} = \frac{d}{dx} \left(D \frac{d\rho}{dx} \right) \quad (4.4)$$

A particle that starts at $x = 0$ at time $t = 0$, will at time t be found at position x with probability

$$\rho(x, t) = \frac{1}{\sqrt{4\pi t D}} \cdot \exp\left(-\frac{x^2}{4Dt}\right), \quad (4.5)$$

with root mean square displacement $\langle x^2 \rangle = 2Dt$ ¹.

¹The 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

$$\rho(x, t+1) - \rho(x, t) = \frac{1}{2}\rho(x-1, t) - \frac{1}{2}\rho(x, t) - \frac{1}{2}\rho(x, t) + \frac{1}{2}\rho(x+1, t) \quad (4.6)$$

where we at position x add and subtract contributions according to the exchange of particles with neighbor positions. Thus

$$\frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial x^2} \quad (4.7)$$

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)

The above solution to the diffusion equation can be validated by insertion of $\rho = F \cdot E$ with $F \propto t^{-1/2}$ and $E = \exp(-x^2/4Dt)$: Now

$$\begin{aligned} d\rho/dx &= -F \cdot \frac{x}{2Dt} \cdot E \\ d^2\rho/dx^2 &= -F \cdot \frac{1}{2Dt} \cdot E + F \cdot \left(\frac{x}{2Dt}\right)^2 \cdot E \end{aligned}$$

and

$$d\rho/dt = -\frac{1}{2t} \cdot F \cdot G + F \cdot \frac{x^2}{4Dt^2} \cdot E$$

which differs only by the factor D . The diffusion equation has the property that a Gaussian stays as a Gaussian, but with an ever increasing width.

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, see Fig. 4.4. In particular it should visit the starting point $x = 0$ a number of times proportional to \sqrt{t} .

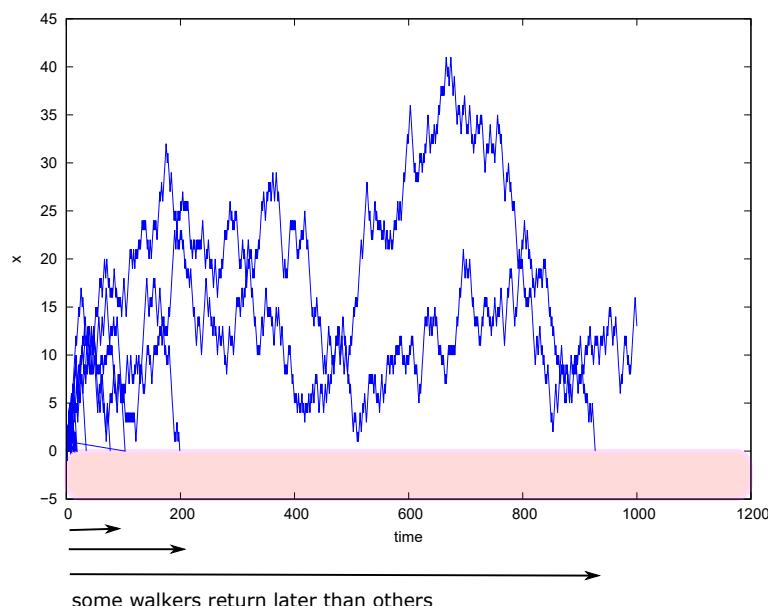


Figure 4.4: **First return of random walkers:** Simulation of 100 random walkers until they first return to starting value of zero.

We first consider a simple heuristic argument repeating the idea associated to the dimension formula in previous chapter (with $D_{num} = 1/2$, $D = 1$ and $D_{tot} = 1$). We first 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 length of each interval:

$$t = \sqrt{t} \int_0^t t' P_{first}(t') dt' \quad (4.8)$$

Here \sqrt{t} is the number of times that a random walker will cross zero during time t . This square root comes from our knowledge that the width of the walk scale as \sqrt{t} and the number of points is t , thus there have to be \sqrt{t} points along the x-axis. The above equation predicts $t = t^{1/2+2-\tau}$, or $P_{first}(t) = t^{-3/2}$. We will prove this more formally 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:

$$First_x(t) = \frac{x}{\sqrt{2\pi D} \cdot t^{3/2}} \cdot \exp\left(-\frac{x^2}{4Dt}\right). \quad (4.9)$$

We here prove this equation based on a derivation that was presented in [30].

Considering the cumulative probability

$$\mathcal{P}_x(t) = \int_0^t First_x(t') dt' \quad (4.10)$$

This is equal to the probability that the walk passed x before time t , i.e. that the max excursion of the walker is larger than x . (Notice we disregard the option of just hitting the position x , as we could consider x off grid).

The distribution of the “max” of the walk up to time t is related to the distribution of the end-point at time t :

- 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. 4.5. 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

$$\mathcal{P}_x(t) = P(\text{max excursion in } [0; t] > x) = 2 \cdot P(\text{end position is } > x)$$

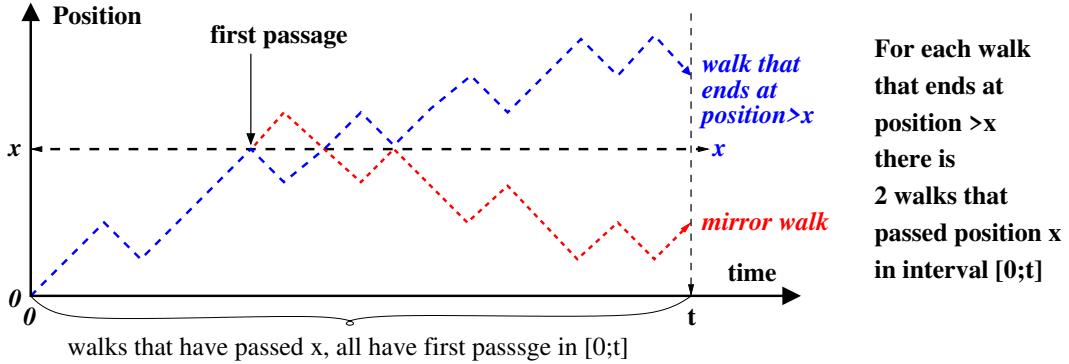


Figure 4.5: **First passage time problem.** Example of a random walker starting at position $x = x_0 = 0$ at time $t = 0$ and ending at some position larger than x . The figure illustrates that for each walk that ends at a position larger than x at time t , there are exactly two walks that have passed x at earlier times.

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 :

$$P(\text{end position is } >x) = \frac{1}{\sqrt{4\pi Dt}} \int_x^\infty dy e^{-y^2/4Dt}.$$

Thus the probability that the random walker exceeds x at some time before t is:

$$\mathcal{P}_x(t) = 2 \frac{1}{\sqrt{4\pi Dt}} \int_x^\infty dy \cdot e^{-y^2/4Dt} \quad (4.11)$$

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

$$\begin{aligned} First_x(t) &= \frac{d}{dt} \mathcal{P}_x(t) \\ &= 2 \frac{1}{\sqrt{4\pi D}} \frac{d}{dt} \left(\frac{1}{\sqrt{t}} \int_x^\infty dy e^{-y^2/4Dt} \right), \end{aligned} \quad (4.12)$$

which is easiest differentiated by bringing $1/\sqrt{t}$ under the integral and substituting. Thus we set a new variable $v = 2Dtx^2/y^2$:

$$dv = -\frac{4Dtx^2}{y^3} \cdot dy \Rightarrow \frac{dy}{\sqrt{t}} = -\frac{y^3 \cdot dv}{4Dx^2 t^{3/2}} = -\frac{1}{\sqrt{2Dx^2}} \cdot \frac{dv}{v^{3/2}} \propto -\frac{dv}{v^{3/2}}.$$

The minus sign means that the substitution ($y \rightarrow v$) 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 :

$$\text{First}_{x \sim 0}(t) \propto \frac{1}{t^{3/2}}, \quad (4.13)$$

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*.

4.3 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. 4.6. 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. 4.6) as a function of time.

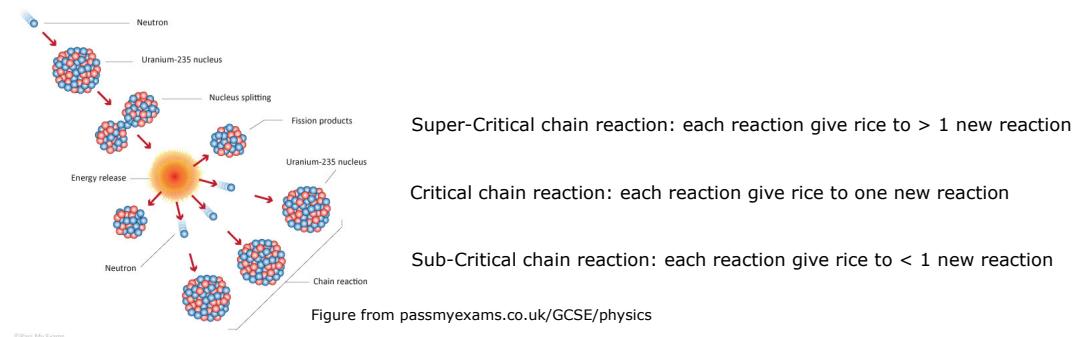


Figure 4.6: **Chain reaction as known from nuclear reactions.** Each fission leads to the emission of three neutrons, which in principle can lead to three new fission events. This process demands that there is a sufficient number of Uranium nuclei available to capture all three neutrons. If the amount 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 $12 \text{ fm} = 12 \cdot 10^{-15} \text{ m}$, A Uranium atom has a diameter of 3.2 \AA , giving a mean free path of about 10 cm .

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 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. 4.6). 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. 4.7. 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 can be obtained from partitioning the tree into two sub-trees at the root. Thus we start with one node, $s = 1$, which can then split into two branches, into one right branch with one additional node, or terminate. The basic assumption of a critical tree is that the probability to split into two branches equal the probability that it terminate and give zero branches. Criticality also means that the probability $p(j)$ for any one of the eventual branches to have j downstream nodes is exactly the same as the probability that the original node have j downstream nodes.

Summing over all partitions of a tree of size s ,

$$p_t(s) = \sum_{k=1}^{s-1} p_{t+1}(k) \cdot p_{t+1}(s-k),$$

where $P_t(s)$ is the probability to have a tree of size s after generation time t . The partitioning correspond to all possible sizes k of the left tree, and the additional requirement that the corresponding right tree should have the remaining size $s - k$. An invariant probability distribution must fulfill $p(s) = p_t(s)$ for all t . I.e. assuming that probability for any branch is distributed as the probability for full tree one obtain

$$p(s) = \sum_{k=1}^{s-1} p(k) \cdot p(s-k), \quad (4.14)$$

The top of Fig. 4.7 shows one such possible partition of a total tree. The recursion relation defined by Eq. 4.14 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 4.7),

$$p(s) \propto \frac{1}{s^{3/2}} \quad (4.15)$$

as it simply reflect the first return of a random walker.

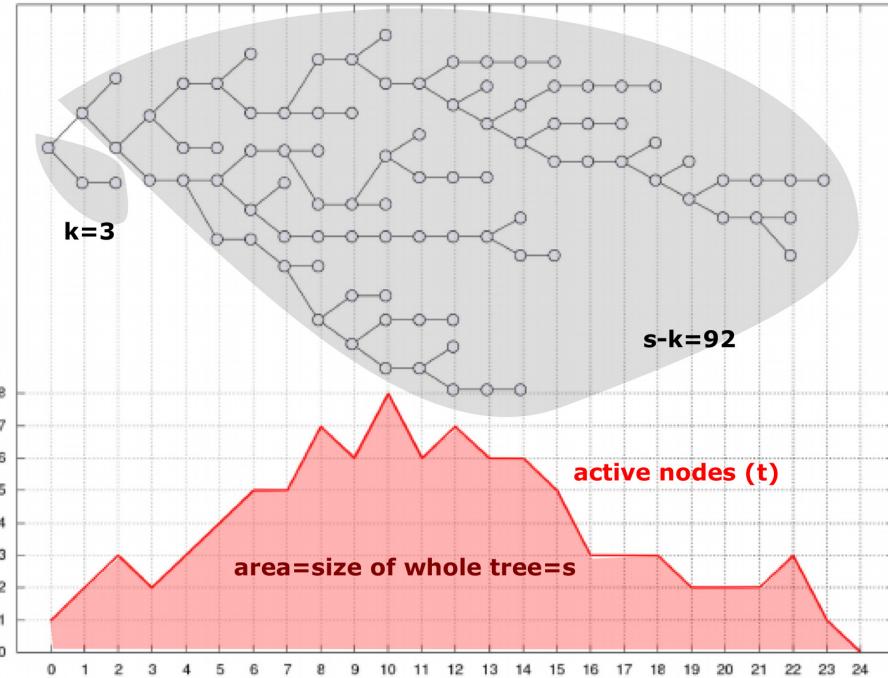


Figure 4.7: **Galton-Walton critical branching process and its relation with a random walk.** At each time update t (horizontal axis) there are several active nodes, and thus the changes tend to be larger when this number is large. Instead one may follow the process as a function of the number of active 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.

Mini 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. Assume mean free path of 10cm.

Noticeably, one may also consider the distribution of survival times for the critical branching trees in Fig. 4.7. 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. 4.7).

4.4 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. 1.2) 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. Some random field Ising models have recently been shown to be more robustly associated to mixture of large and small scale events [31]. However, in general 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. 4.8.

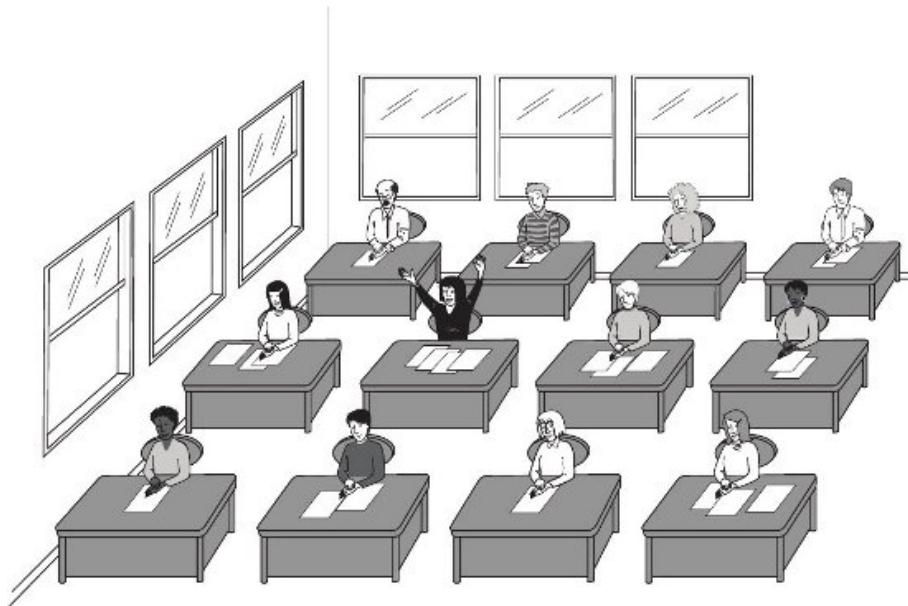


Figure 4.8: **Sandpile model cartoon.** The classical “sandpile model” of Bak, Tang and Wiesenfeld, here re-explained in an office version by Peter Grassberger.

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 $h_i = 0, 1, 2, 3, 4, 5, \dots$. All sites with $h_i = 4, 5, \dots$ 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. 4.8.

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?)

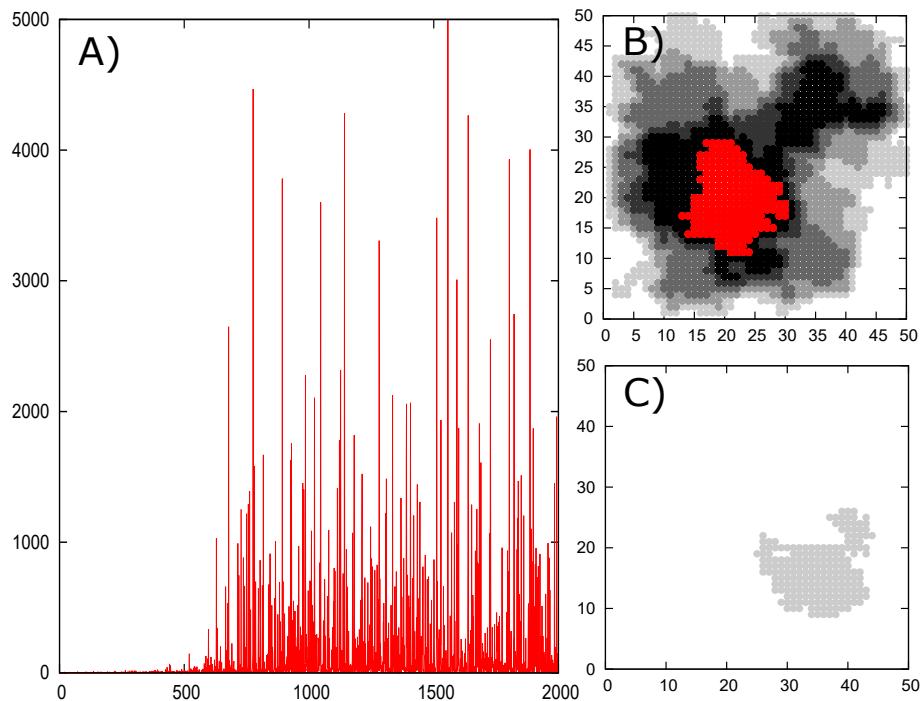


Figure 4.9: Avalanche dynamics in the sandpile model. Panel A shows the time series of avalanche sizes as one starts filling the sites in a 50×50 square lattice. After some time (a transient phase), a steady state distribution of avalanches is obtained. In panels B and C we show the spatial extent of 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).

The key observable is the sequence of topplings that take place when one new grain is added until the system is settled (see Fig. 4.9). This constitutes an avalanche, which is measured by the sum of all activity until all sites are below the threshold. That is its size s is equal to number of topplings caused by one

grain addition until the system is again at rest. Fig. 4.10 show the distribution of avalanche sizes simulated for 4 different systems sizes. The distributions are all fitted by

$$p(s) = \frac{1}{s^{1.33}} \cdot f\left(\frac{s}{L^{2.86}}\right) \quad (4.16)$$

where the cutoff function $f(x) \sim 1$ for $x < 1$ and decays rapidly for $x > 1$ (ref. K. Christensen and N.R. Maloney, 2005).

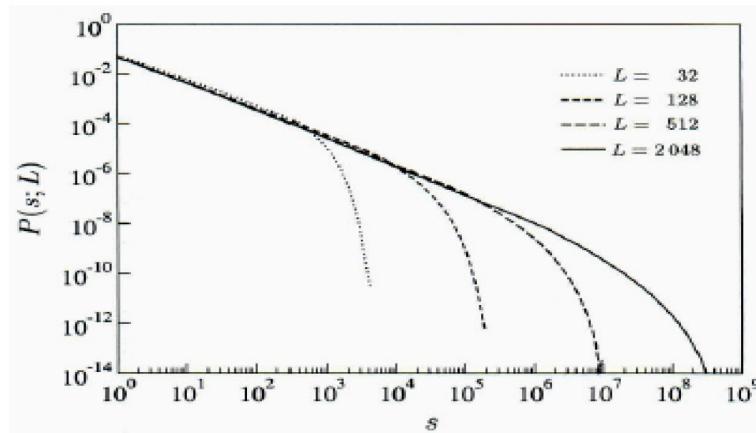


Figure 4.10: **Sandpile model simulation.** Avalanche distribution for different system sizes. As system gets larger the upper cut off moves to larger sizes, reflecting a cutoff associated to grains falling off the edge of the sandpile (out of the windows in the bureaucrat illustration).

In general the distribution of the sizes s of these avalanches turns out to be power law distributed,

$$P(s) \propto \frac{1}{s^\tau} \quad , \quad \text{with } 1 < \tau < 3/2 \quad , \quad (4.17)$$

with the exact value of τ dependent on the dimension. In two dimensions $\tau \sim 1.3$ whereas $\tau = 3/2$ in infinite dimension or a random neighbor or mean field version. Such a power law reflects that the system dynamics organize itself to an attractor around the critical point. 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. 4.11). And it does so without any fine tuning.

Fig. illustrate an experiment on avalanches in 1 dimensional rice piles, which much the same behaviour as above, except the unexplained large exponent of 2.04.

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?

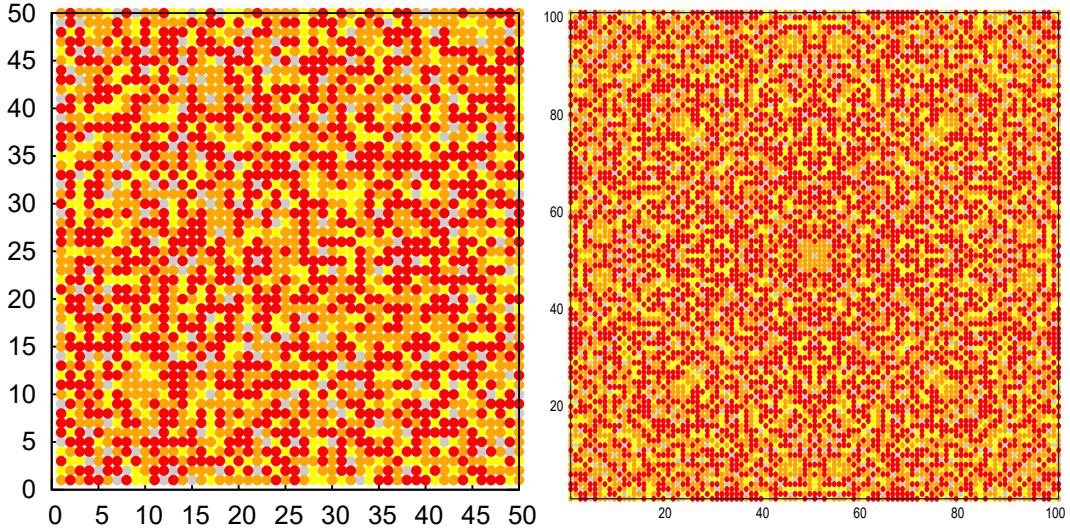


Figure 4.11: **Site configuration in an SOC model.** The model is driven by adding sand grains to random points (left panel), respectively always adding to point in the middle of the 101×101 square lattice used here (right panel). Sites colored in yellow mark $h = 1$, orange ones mark $h = 2$, red ones mark $h = 3$.

A scaling relation from following a grain of sand.

One can derive some scaling relations for the sandpile model, using that it has to balance is a transport in ad out of the system. Following Kadanoff we consider the avalanche size distribution in the self-organized critical state:

$$P(s) = \frac{1}{s^\tau} \cdot f\left(\frac{s}{L^D}\right) \sim \frac{1}{s^\tau} \cdot \exp\left(-\frac{s}{L^D}\right) \quad (4.18)$$

where the function f is ~ 1 for $s \ll L^D$, and where f decreases very fast when the avalanche extend to the system size, $s > L^D$. Here for simplicity we use an exponential distribution as an illustration of a possible cutoff behaviour. The function is often even more on-off like, for example

$$f(x) \sim \exp(-x^2) \text{ or even } \sim \Theta(1 - x)$$

where Θ is the step function $=1$ for $x < 1$ and $=0$ for $x > 1$. D is the dimension of the avalanche, i.e. how the number of topplings in the avalanche scales with its horizontal extension, $s \sim L^D$. 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

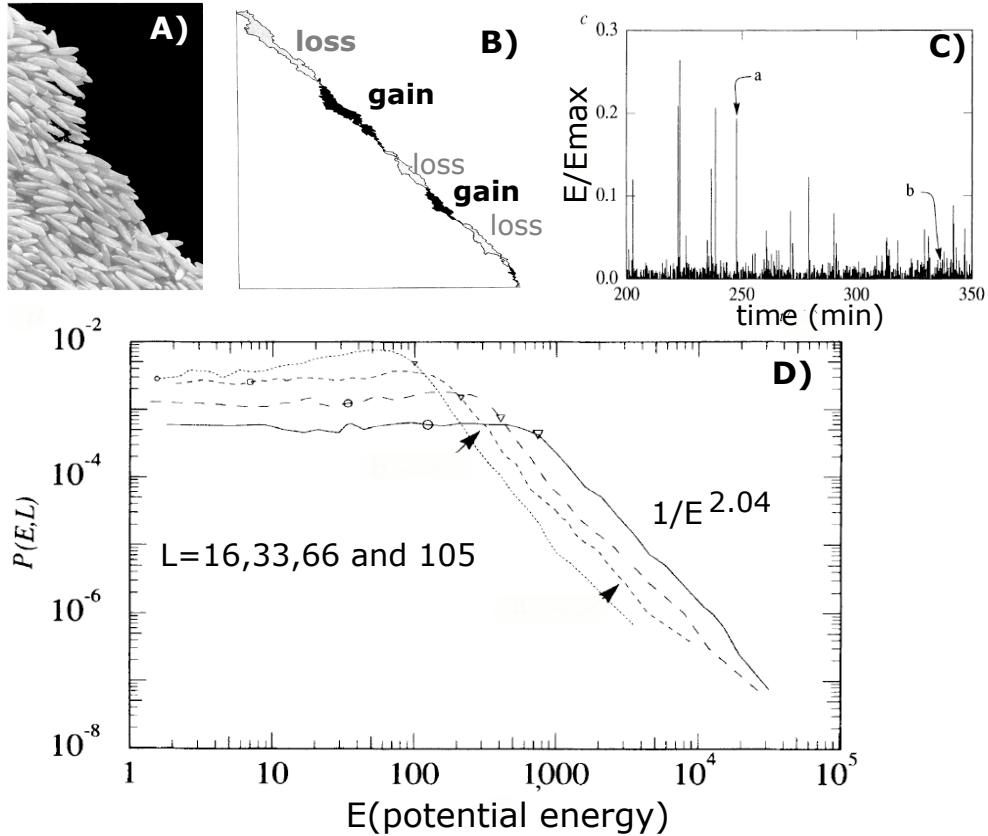


Figure 4.12: **Experiment on 1-dimensional rice pile.** Avalanches in rice from ref. [32], with measured avalanche exponent of 2.04. Panel A) show a small part of the rice pile. B) show whole rice pile with black and white areas marking where grains was moved to and from. C) Time sequence of avalanches. D) show the distribution of released (dissipated) potential energy in avalanches for 4 different system sizes L .

thus cannot contribute to a loss. However, sometimes avalanches reach the boundary and many grains will leave the system.

The steady state condition means that on average one avalanche has to provide sufficiently many topplings to transport one grain to the boundary and subsequently 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. See Fig. 4.13. Therefore, for $\tau < 2$, the steady state implies (inserting the cut-off function in the upper end of the integral)

$$L^2 \sim \int_0^\infty s \cdot P(s) ds \sim \int_0^{L^D} s^{1-\tau} ds = L^{D \cdot (2-\tau)} \quad (4.19)$$

and thus we get

$$2 = D \cdot (2 - \tau) \Rightarrow \tau = 2 - \frac{2}{D}. \quad (4.20)$$



Following a single sand grain as it moves as the current site it resides topples. Its random walk terminates at the boundary, typically a distance $L/2$ away from the initial site of addition.

Figure 4.13: **The random walk of a single grain.** Each toppling pushes each involved grain in a random direction, implying that it takes $\sim L^2$ topplings to move one grain to the boundary.

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. Thus the $D = 2.86$ from Fig. 4.10 would be consistent with $\tau = 1.30$ which is indistinguishable from the estimated exponent of 1.33 from that figure.

If $D = 2$, corresponding to flat avalanches then $\tau = 1$. In fact in the above sandpile model, then avalanches that starts at the boundary remain flat, i.e. no site can topple more than once during one avalanche. Further, as we shall see below, then τ changes (increases) when one consider a sandpile model that is only excited at the boundary.

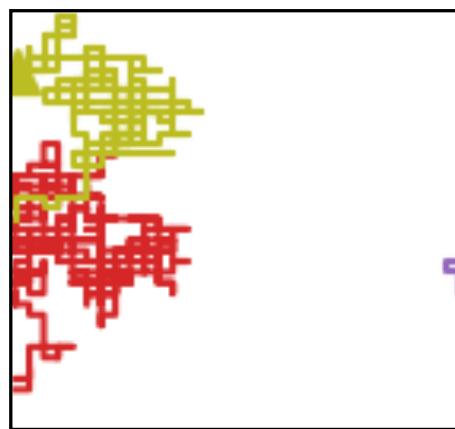


Figure 4.14: **The random walks of single grains started at boundary.** Notice that the walks are typically short, although we here depict some pretty long walks. When walks start inside the system they always have to be fairly large.

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:

$$\text{average exit time from boundary} = \int^{L^2} \frac{t \cdot dt}{t^{3/2}} \propto L \quad (4.21)$$

where the upper boundary is set by the time it takes to cross the system and exit on the other side. See Fig. 4.14. 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

$$\begin{aligned} L &\sim \int_0^\infty s \cdot P(s) ds \sim \int_0^{L^D} s^{1-\tau} ds = L^{D \cdot (2-\tau)} \\ \Rightarrow 1 &= D \cdot (2 - \tau) \Rightarrow \tau = 2 - \frac{1}{D}. \end{aligned} \quad (4.22)$$

which is a steeper exponent than when grains are added in the bulk. I.e. Large avalanches are less likely because of the increased loss of excitation at the boundary.

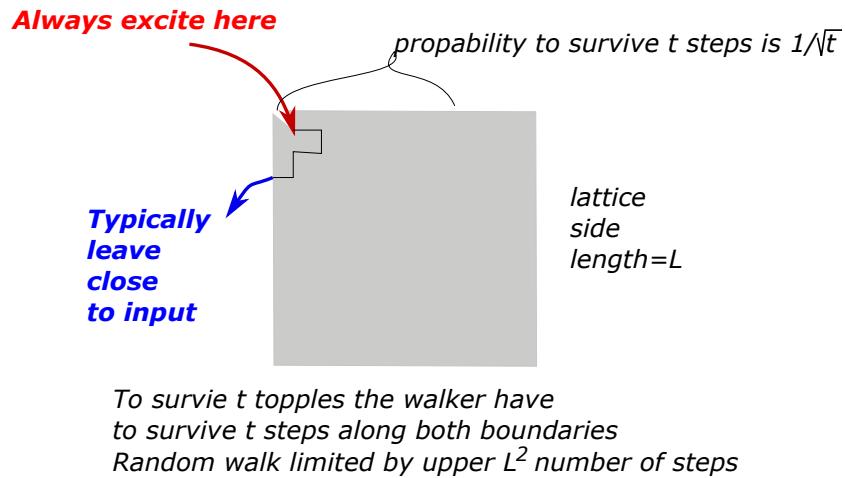


Figure 4.15: **The random walks of single grains started at corner.** Here the walks are mostly very short, but some will reach sizes of order L^2 . The resulting scaling of avalanches are $1/s^2$.

One could even consider the extreme case of exciting only in one corner of the 2-dimensional grid, see Fig. 4.15. In that case the grain can be lost along two boundaries. The probability that it passes one boundary after t steps is $\int_t^\infty dt/t^{3/2} = 1/\sqrt{t}$, and the probability that it is lost at times after t along any of the two boundaries is $(\int_t^\infty dt/t^{3/2})^2 = 1/t$. Differentiating this cumulative distribution implies that the first exit from a corner decays as $1/t^2$ with

$$\text{average exit time from corner} = \int^{L^2} \frac{t \cdot dt}{t^2} = \ln(L^2) \propto \ln(L)$$

The avalanches initiated only by excitation at the boundary gives:

$$\begin{aligned} \ln(L) &\sim \int_0^\infty s \cdot P(s) ds \sim \int_0^{L^D} s^{1-\tau} ds = L^{D \cdot (2-\tau)} \\ \Rightarrow 0 &\sim D \cdot (2 - \tau) \Rightarrow \tau \sim 2 . \end{aligned} \quad (4.23)$$

which shows that one indeed can get exponents up to 2 in SOC models, provided careful constraints in the way one introduce energy.

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. 4.8). 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. 4.13. Thus the avalanche size distribution is

$$P(s) \sim \frac{1}{s^{3/2}} \cdot \exp(-s/N) . \quad (4.24)$$

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. 4.16.

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 receive two grains and thus for certain topples at next step. Also by nature no site will topple

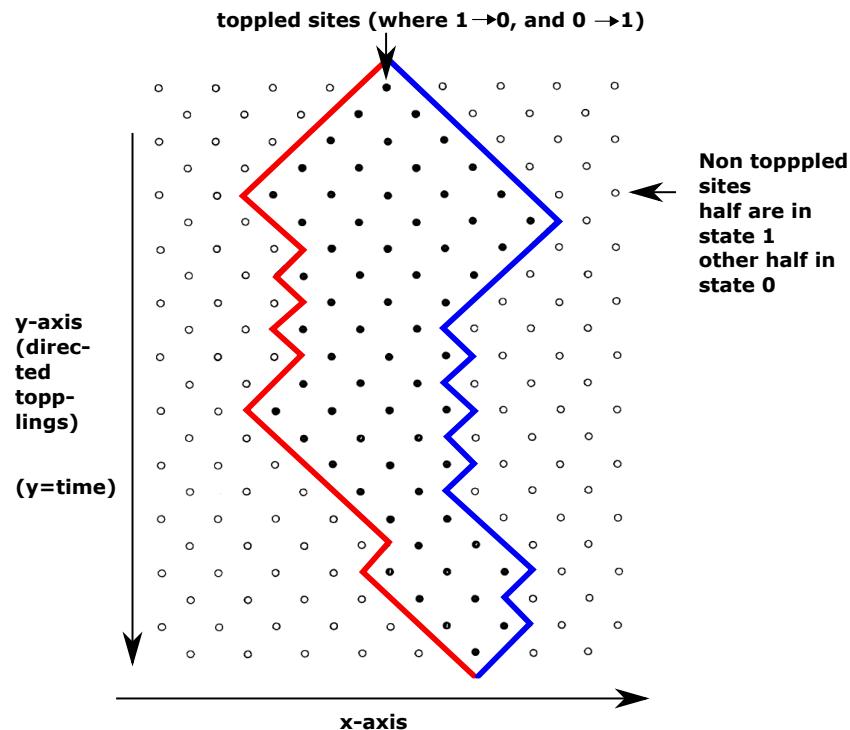


Figure 4.16: **Directed sandpile.** Each site can contain zero, one or more sand grains. If a site contains two or more grains it topples and delivers one grain to each of the two sites below it (downwards on the figure, see green arrows). The critical state contains zero or one with equal probability=1/2 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.

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. 4.16): 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 random walks is again a random walk, the avalanche duration will be power

law distributed as the first return of a random walk (eq. 4.13):

$$P_{duration}(t) \propto \frac{1}{t^{3/2}}, \quad (4.25)$$

and the chance for an avalanche to propagate more than 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 :

$$P_{size}(> s) = P_{duration}(t > s^{2/3}) = \frac{1}{s^{1/3}}, \quad (4.26)$$

yielding the avalanche size distribution $p(s) = dP/ds = 1/s^\tau$ with exponent $\tau = 4/3$.

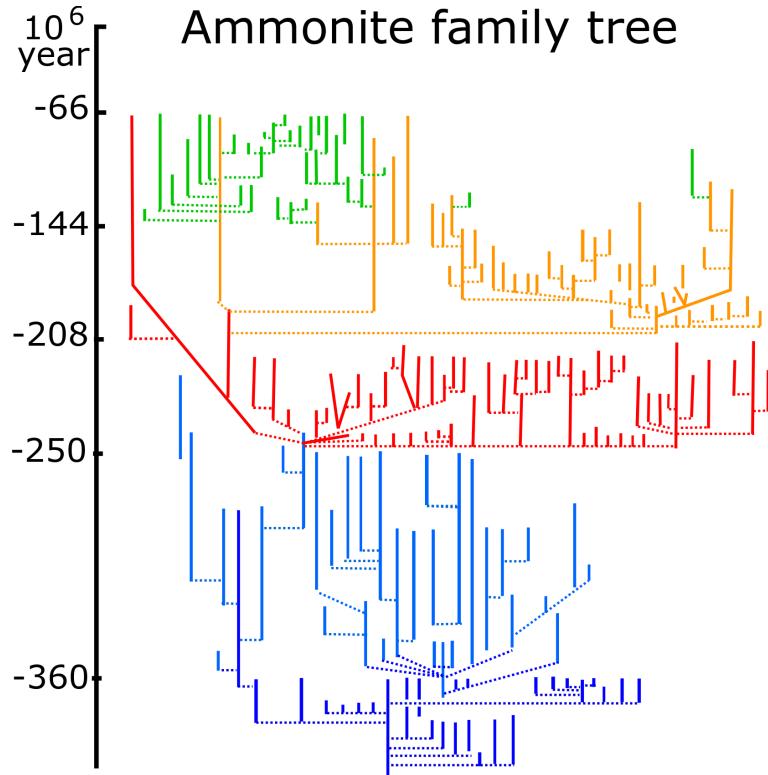


Figure 4.17: Life & history in term of the Ammonite family tree. Reproduced from Ref. [33]. Ammonites lived in water, and left a highly diverse fossil record with ~ 7000 species, from 400 to 66 Million years ago. Notice the intermittent dynamics with calm periods interrupted by coherent extinction/speciation events.

4.5 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 4.18. 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.

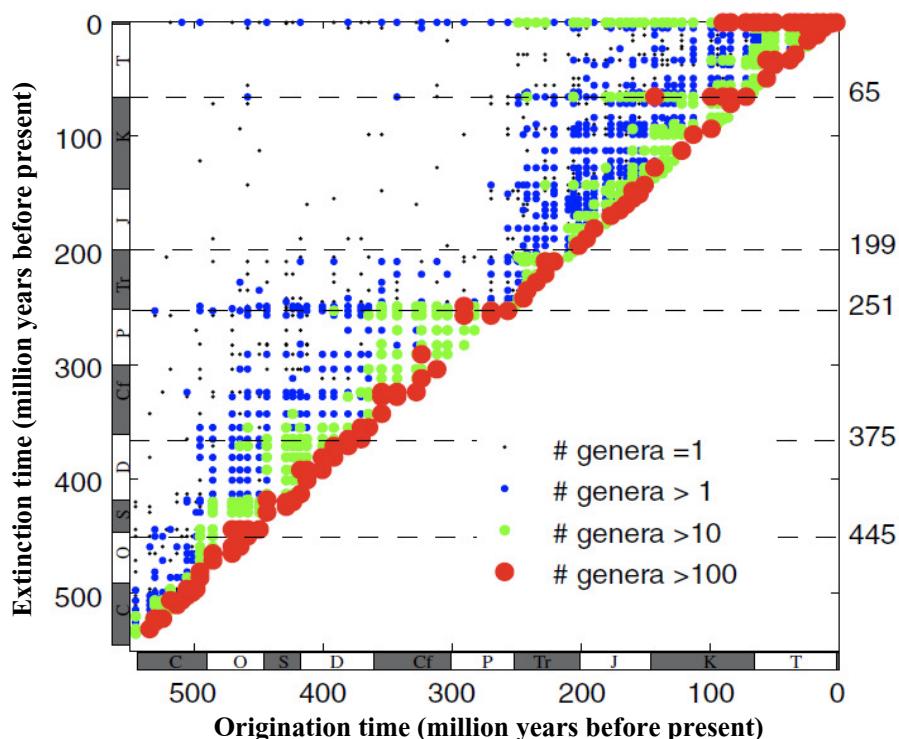


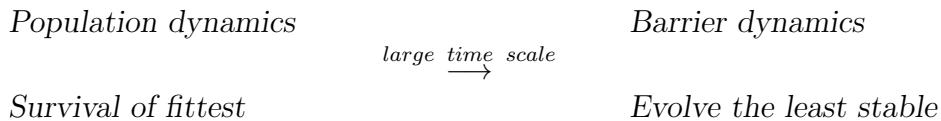
Figure 4.18: Origination and extinction. The graph shows existence periods of 35,000 genera in the Phanerozoic [34] as visualized by [35]. Every dot measures a number of genera, which each in turn is a group of closely related species. The vertical distance from a point to the diagonal measures the residence 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.

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. 4.19 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:

- 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 [36] 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. Of 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:



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 [37].

Mini Tutorial: Draw ten real numbers from a uniform distribution between zero 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 , $i = 1, 2, \dots, N$ are placed on a line, mimicking a one-dimensional model

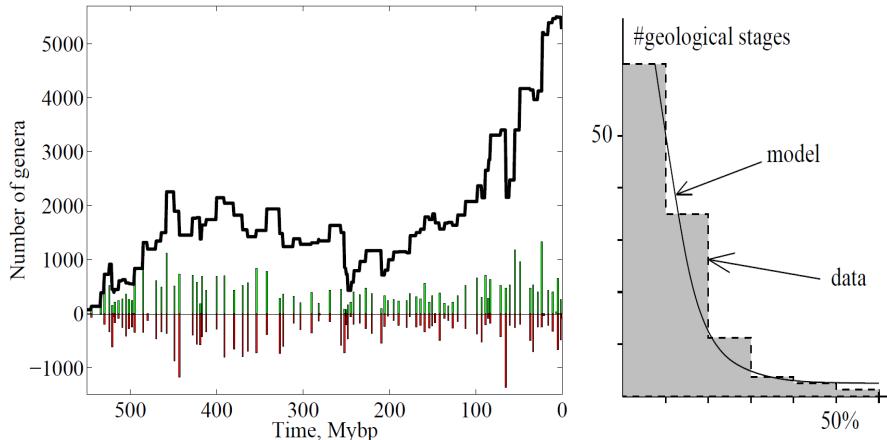


Figure 4.19: **Comparing the Bak-Sneppen model to data.** *Left:* Trajectory of events (corresponding to projecting matrix from last figure on a and y axis). *Right:* Histogram of family extinctions in the fossil record as recorded by Raup & Sepkoski [34]. The prediction, $p(s) \propto 1/s^{1.5}$ of the random neighbor version of the Bak-Sneppen model is marked as “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 [38, 39]:

- **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. This can be followed as in Fig. 4.20. For the infinite system size limit ($N \rightarrow \infty$) 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. 4.21. The right panel illustrates how the minimal B_i sites changes in the “species space” as 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. 4.21 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

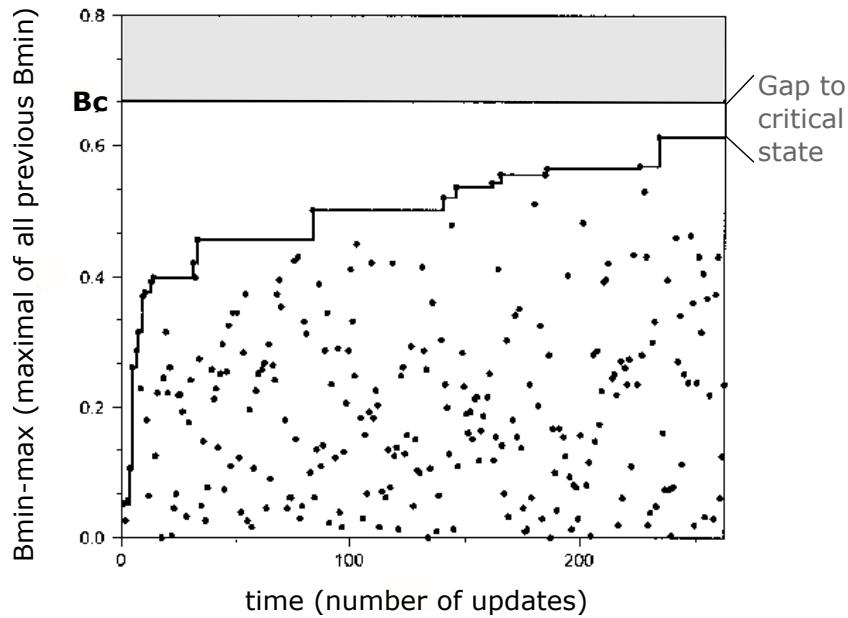


Figure 4.20: **Development to critical state takes time.** The dots show selected minimal B_i at subsequent times early in the evolution. At any time we calculate $\text{Gap} = \max_{t < t'}[B_{\min}(t')]$, a function that is increasing in steps. As long as $\text{Gap} < B_c$ we are in transient. When $\text{Gap} > B_c$ we are on the critical attractor.

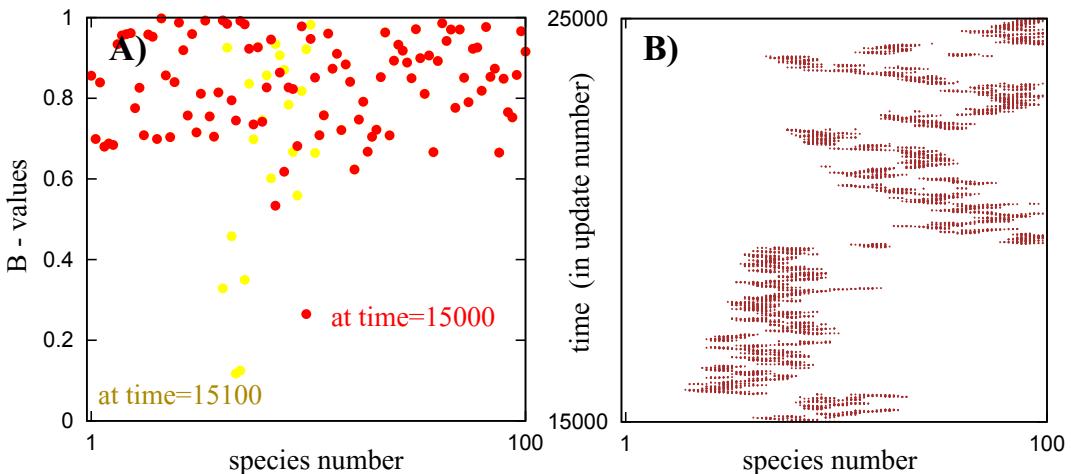


Figure 4.21: **Dynamics of the Bak-Sneppen model.** A) Example of distribution of B_i in the space spanning from 1 to 100. Overall the B 's take fairly random numbers, although some pattern emerge as result of the self organizing process. In particular, the $B > B_c = 0.6670$ are distributed randomly in space. In contrast, sites with $B_i < B_c$ are correlated and within a small region close to the current minimal B value. B) Space-time plot of the activity in self organized chritical state where “time” is counted by the number of updates. At each timestep the single site with the minimum B_i is highlighted by a small red dot.

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 [40, 41] to larger evolution and origination of new taxonomic groups[42, 43].

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 two 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 which in turn takes time to develop. With sufficient time the self-organization allows the evolving system to develop towards a dynamical “attractor”. A steady state where the system moves among states where the numbers on the lattice exhibit long distance and temporal correlations. This is seen by the fact that the distribution of “species” with smallest B_i moves in fractal way 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.

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 $\mu \rightarrow 0$ limit of the following local model [44]:

- **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 .

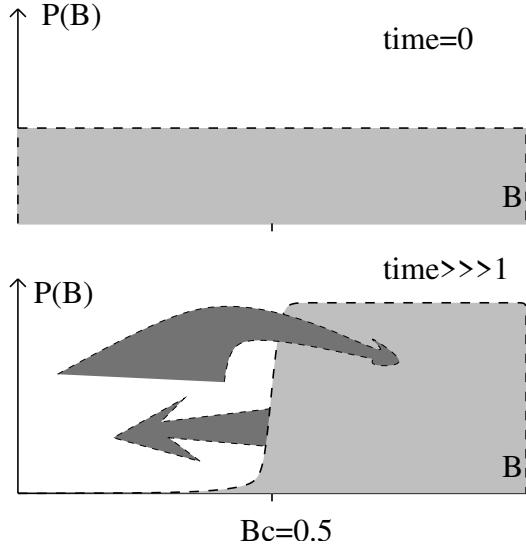


Figure 4.22: **Distribution of barriers/fitnesses in the BS model for the random neighbor version.** At each update one takes the minimum B_i and one random B_i value, and replaces each of them with a new random number within the interval $[0 : 1]$. The resulting distribution of B_i 's reaches a steady 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.

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. 4.23). Time scales that may be set by associating each step of the algorithm to a time interval

$$\Delta t \propto \frac{1}{\sum_i e^{-B_i/\mu}} \sim e^{B_{min}/\mu} \quad \text{for } B_c - B_{min} \gg \mu. \quad (4.27)$$

Here, the final approximation uses that the distribution of barriers below B_c is scarce.

For $\mu \rightarrow 0$ 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

$$P_0(s) \propto s^{-\tau} \quad (4.28)$$

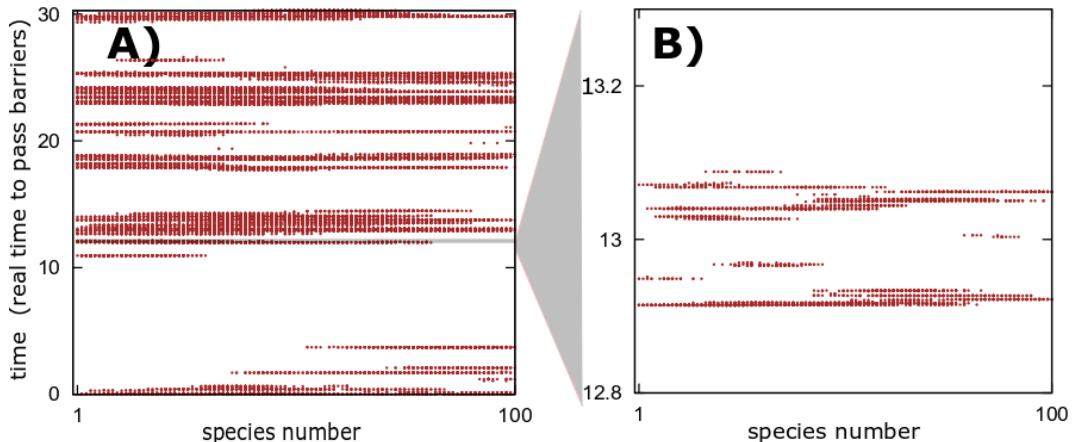


Figure 4.23: Space-time plot of the activity in the Bak-Sneppen evolution model. Each update is shown as a red mark. With the coarse time resolution of the plot, the avalanches appear as almost horizontal lines. The magnification on the right shows that there are avalanches within avalanches. The calculation was done at a mutation rate $\mu = 0.005$.

Here the exponent $\tau = 3/2$ for ecological networks with dimension $d \geq 4$ [45] which compare well with the histogram of extinction events shown in Fig. 4.19.

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 [46, 47]. Here, at each update one changes the site with minimal B_i as well as one other randomly selected B_i , see Fig. 4.22. Because of the absence of spatial correlations between the B_i 's this simpler model can be solved analytically [47].

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 $N \rightarrow \infty$ 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 were selected as the random neighbor, irrespective of their actual B value. Therefore the distribution of B is constant above B_c .

At each step of the dynamics, one B is always selected from below B_c and the other B above B_c . (Notice that 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 mean that one must remove and add the same number of

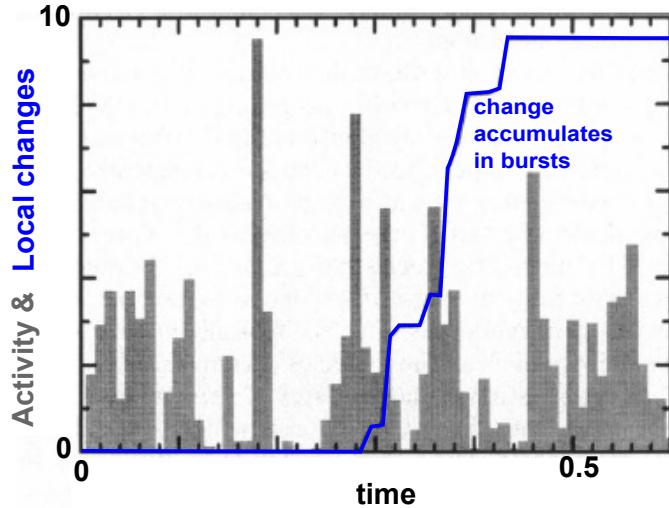


Figure 4.24: **Global activity in the Bak-Sneppen model.** Global activity (grey) plotted together with accumulated change at one position in the model ecosystem, using the finite mutation rate version [39, 44]. The figure illustrates that the individual species change most rapidly, when also the global ecosystem exhibits a lot of activity. Figure reproduced from [39].

species in the interval $[0, B_c]$:

$$-1 + 2B_c = 0 \Rightarrow B_c = \frac{1}{2} \quad (4.29)$$

as the chance to add each of the two species below B_c is in fact B_c . 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 be smaller than $1/2$. 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 selected 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 defined as the number of updates s before this random walk return to zero:

$$P(s) \propto s^{-3/2} \quad (4.30)$$

This again is the famous distribution of waiting times in the *Gamblers Ruin Problem*.

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

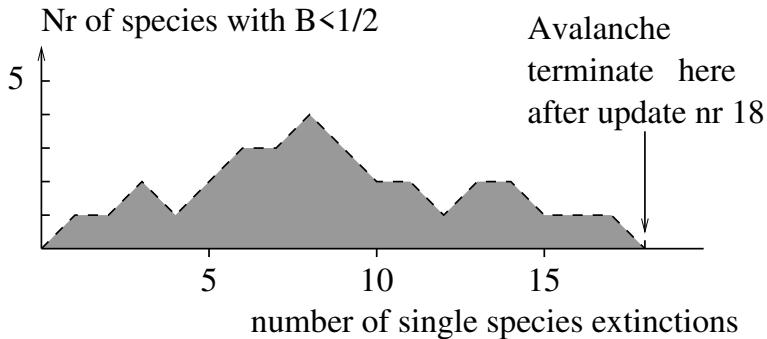


Figure 4.25: Evolution of the number of sites with $B < 1/2$ in the random neighbor version of the Bak-Sneppen model. At any timestep there is equal probability to increase or decrease the number of active species a , thus defining a random walk of this number. The first return of the random walk to 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.

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 [48], 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 [49].
- 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 4.24).

Questions:

- 4.1)** 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.
- 4.2)** 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.
- 4.3)** 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.
- 4.4)** 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.
- 4.5)** 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.
- 4.6)** 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.
- 4.7)** 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.

4.8) 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.

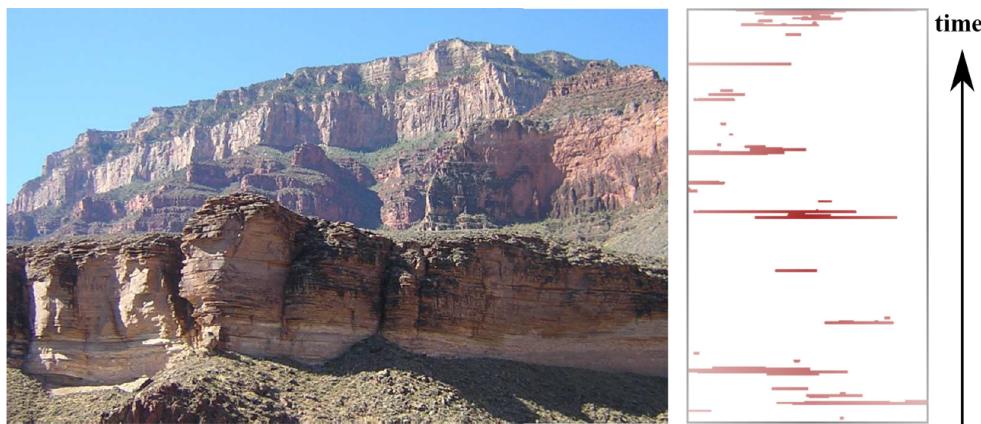


Figure 4.26: **Paleontology % evolution as SOC:** Separated timescales in the ecological and paleontological record show up in long periods with geological depositions of similar type of sediments, interrupted by sharp demarcations before a new geologic period starts. Left panel show an example of such a formed mountain landscape in south western USA. Right panel show a simulation of the activity in the co-evolution model with separation of time-scales (between local adaptation and real “inventions”). The BS model is here simulated for $\mu = 0.005$ and $N = 140$. Left picture with courtesy of S. Semsey.

Lessons:

- Self organized critical dynamics needs 1) time to organize and 2) separation of timescales such that avalanches are well separated from each other in time (fast dynamics within avalanche and longer times between avalanches).

- Random walkers, in particular the first return of random walkers is a recurrent theme in critical processes, from fine tuned cascade processes to self organized criticality. First return probability of random walkers are most simply derived from:

$$T = \sqrt{T} \int_0^T t P_{first}(t) dt \Rightarrow P_{first}(t) \propto \frac{1}{t^{3/2}} \quad (4.31)$$

- Exponents in the sandpile-model are connected by the observation that each grain in average must participate in a relaxation step $\langle s \rangle \sim L^2$ times to reach the boundary. This connects the exponent for avalanche sizes, τ , with the exponent D that sets the cutoff of avalanches with system size L , $s_{max} \sim L^D$:

$$L^2 \propto \int^{L^D} \frac{s}{s^{tau}} ds \Rightarrow \tau = 2 - \frac{2}{D} \quad (4.32)$$

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 5

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! ”

5.1 Introduction

5.1.1 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. 5.1 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. 5.1 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 [50].

Mini Tutorial: Why is it often useful to limit information spreading in living systems?

5.1.2 Basic concepts

Figures 5.2-5.8 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 Erdős-Rényi (ER) network [52]. 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 $L = p \cdot N(N-1)/2$. That is each of N nodes can connect to each of the remaining $N - 1$ nodes, with $1/2$ takes into account the double counts. The average degree (=connectivity), defined as number of neighbors

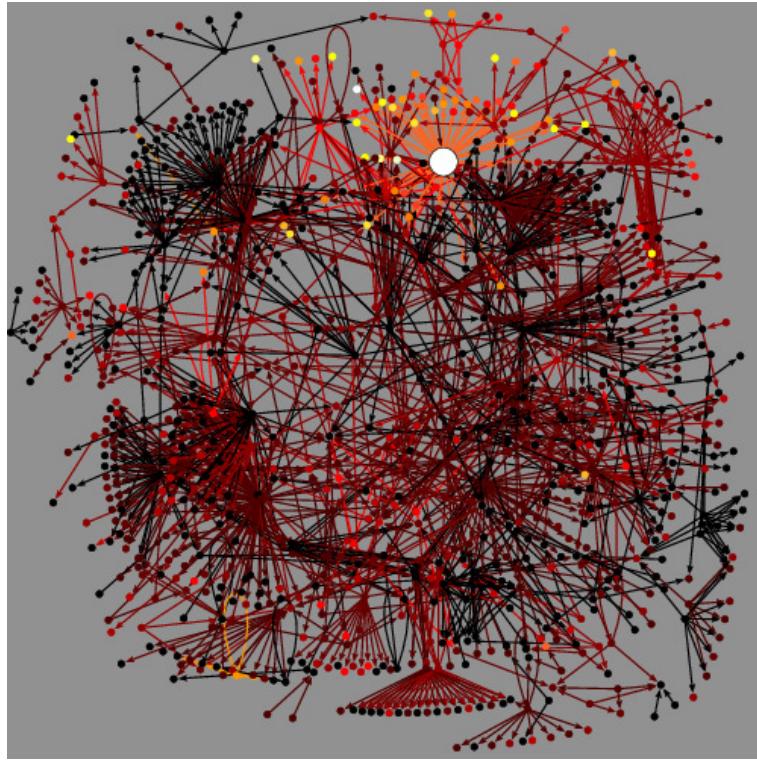


Figure 5.1: Genetic regulatory network in *Saccromyces cerevisiae*. Reproduced from [51]. Notice that the network consists of proteins that are all regulated through the cell nucleus, and does not reflect any “geographical” separation of the proteins. The highlighted nodes reflect genes that change expression in response to a particular external stimulus (amino acid starvation).

per node, is then

$$\langle k \rangle = \frac{2L}{N} = p(N - 1). \quad (5.1)$$

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

$$P(k) = \frac{(N - 1)!}{k!(N - 1 - k)!} \cdot p^k \cdot (1 - p)^{(N-1-k)}, \quad (5.2)$$

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. 5.3)

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \quad (5.3)$$

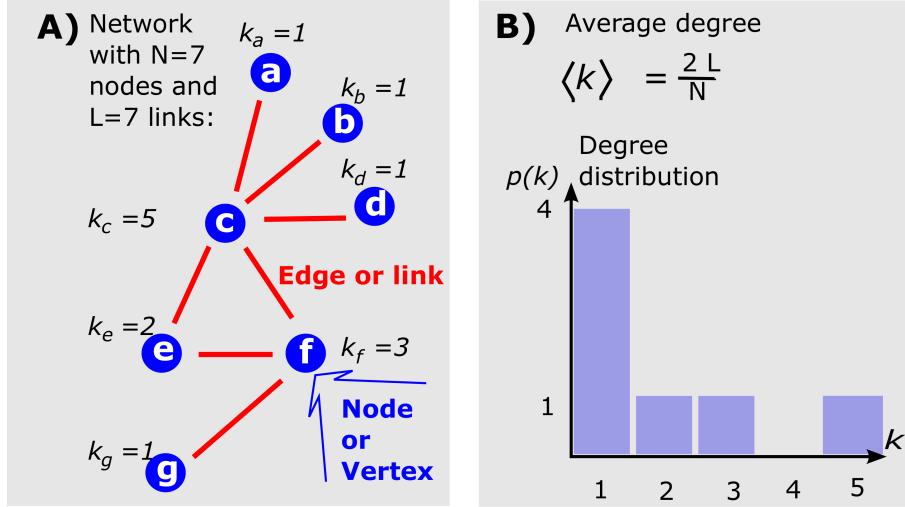


Figure 5.2: **Basic network definitions.** A) Network with nodes and links, and each node characterized by its degree, which is the number of links associated to the node. B) The degree distribution $n(k)$, $k = 1, 2, \dots$ for the network in panel A).

with an average degree

$$\langle k \rangle = p \cdot (N - 1) \quad (5.4)$$

and variance $\text{var} = \langle k^2 \rangle - \langle k \rangle^2 = \langle k \rangle$, 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 $\sigma = \sqrt{\text{var}} = \sqrt{\langle k \rangle}$ 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. 5.4A). The distance between two nodes in a network is defined as the minimal number of links that connects these nodes¹⁾. 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. 5.4B). Thus the neighbors of the neighbors in total

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.

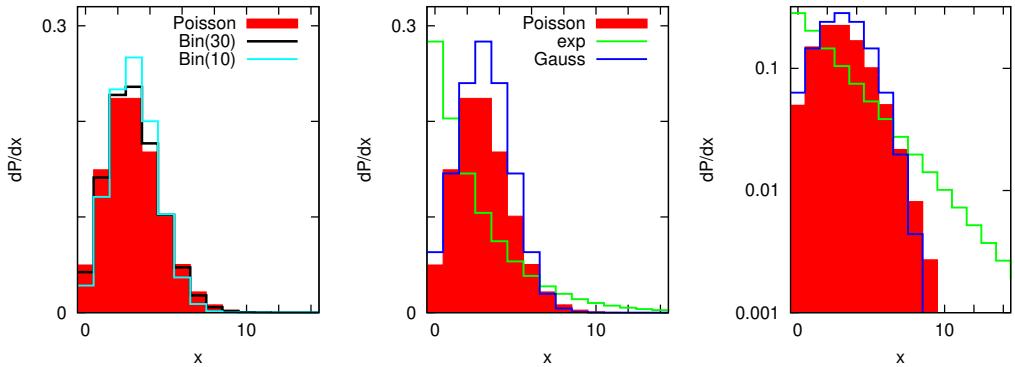


Figure 5.3: Comparing distributions. Left panel: Comparison between Poisson and Binomial distributions both with average 3, and Binomial sampled from 10, respectively 30 decisions. Right panels show comparison of Poisson with exponential and Gaussian distribution. The exponential has the same average 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 .

reaches $k \cdot (k - 1)$ nodes, provided that we ignore possible degenerate pathways. 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

$$\text{number(nodes within } d) = k \cdot \sum_{d'=1}^d (k - 1)^{(d'-1)} \sim k \cdot (k - 1)^{d-1}. \quad (5.5)$$

Therefore the number of visited nodes grows exponentially for any $k > 2$, see Fig. 5.4. 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

$$\left(\frac{\langle (k - 1)k \rangle}{\langle k \rangle} \right)^d = \left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right)^d \approx N, \quad (5.6)$$

which we will derive shortly through the "amplification factor" expression below. Importantly, eq. 5.6 takes into account that each subsequent node is selected by a probability proportional to its degree. This is because any node is at the end of a link: When one follow a link there is a relatively bigger chance to encounter a node with more links.

We now use the above considerations to estimate when a signal/epidemics that amplify and spread on each node will start interfering. For a completely random network 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

$$Diam \sim d \sim \frac{\log(N)}{\log\left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1\right)}. \quad (5.7)$$

The main lesson resulting from eq. 5.7 is that the diameter of a random network only grows very slowly with network size N .

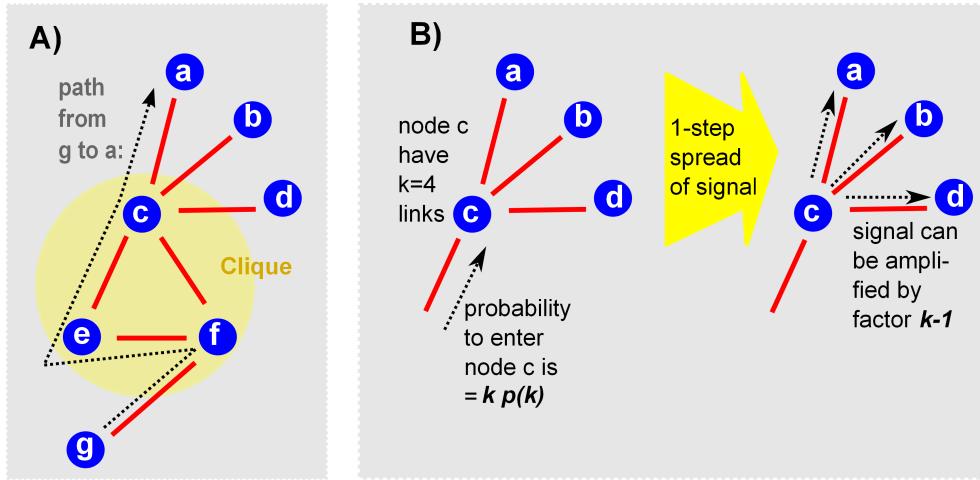


Figure 5.4: **Spreading process on a network.** A) Path from node g to node a in a network. Notice that the path shown is longer than the shortest path ($g-f-a$). B) Amplification of “virus like” signal as it enters and spreads across node c with connectivity $k = k_c = 4$. As such a signal subsequently 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?

5.1.3 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. 5.4B). 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 [53] is obtained by the

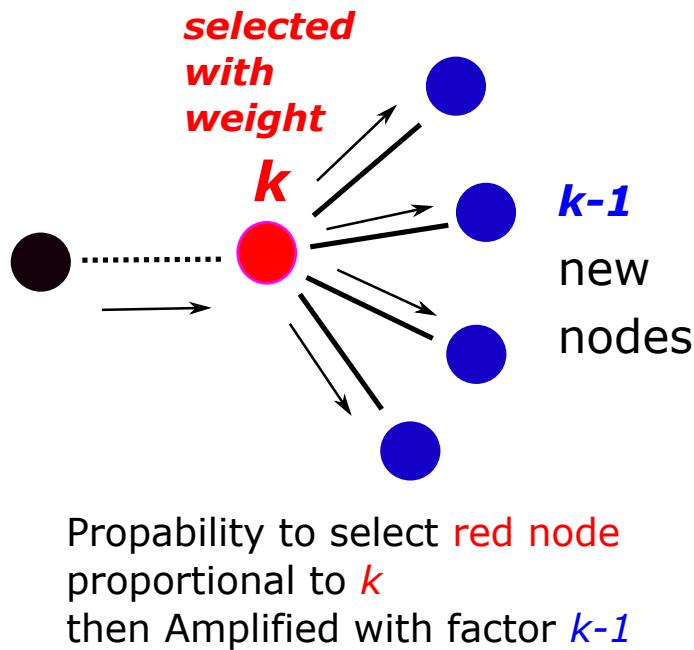


Figure 5.5: **Amplification factor.** Signal spreading across networks is followed by the amplification factor. The arrows are not referring to links, but are just indicating how an infinitely infectious virus would spread on the network.

weighted average

$$\mathcal{A} = \frac{\int k(k-1)n(k)dk}{\int kn(k)dk} = \frac{\langle k(k-1) \rangle}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle} - 1. \quad (5.8)$$

Obviously this equation assumes that signals can spread both ways across a link. This would be violated if network was directed, i.e. if signals only transmit one way along each link. Directed networks are not considered in these notes, but play an important role in for example biological regulation, or in hierarchical organizations.

Mini tutorial: Write the above expression for a directed network, using nodes in degree (k_{in}) and out degree (k_{out}). Consider situation where in and out degrees are uncorrelated.

Equation 5.8 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. 5.6.

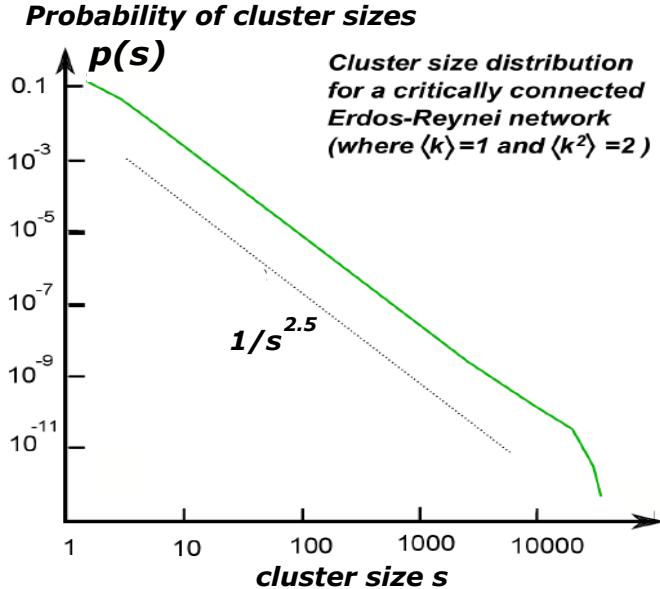


Figure 5.6: **Cluster size distribution:** Cluster size distribution in a critical ER network (where $\mathcal{A} = 1$), $P(s) \propto 1/s^{2.5}$ [54], implying that a random node will be in a cluster with size distribution $n \cdot P(n) \propto 1/s^{1.5}$. The 2.5 exponent recapitulates what we earlier found for critical percolation in the Bethe Lattice.

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. We will see that \mathcal{A} determine the fraction of nodes that need to be susceptible for a disease to propagate exponentially. This is in analogy to the number 2 for the Bethe lattice where each node has 3 neighbors (In general the 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* [55, 56, 57]. A clique is defined as 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 [57] if all its neighbors was connected to each other.

$$C_i = \frac{\text{number of pairs of neighbors to } i \text{ connected to each other}}{k_i \cdot (k_i - 1)/2} \quad (5.9)$$

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 Erdős-Rényi network the global clustering coefficient is

$$C = p \sim \frac{\langle k \rangle}{N} \quad (5.10)$$

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.

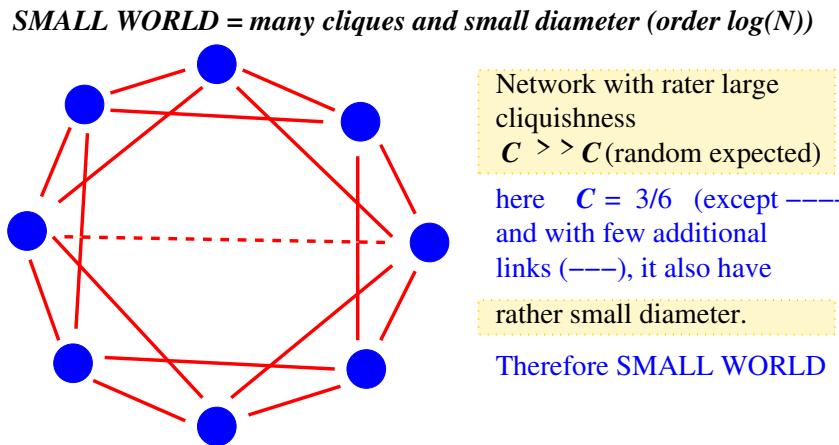


Figure 5.7: Classical small world network. Originally introduced by Watts and Strogatz [57]. See figure for explanations.

Mini Tutorial: Give a heuristic argument for why the numbers of 3-loops in a Erdos - Reynei network is independent of its size N . (number of attempts to make loops proportional to N , but each attempt only have some fixed probability to form a loop).

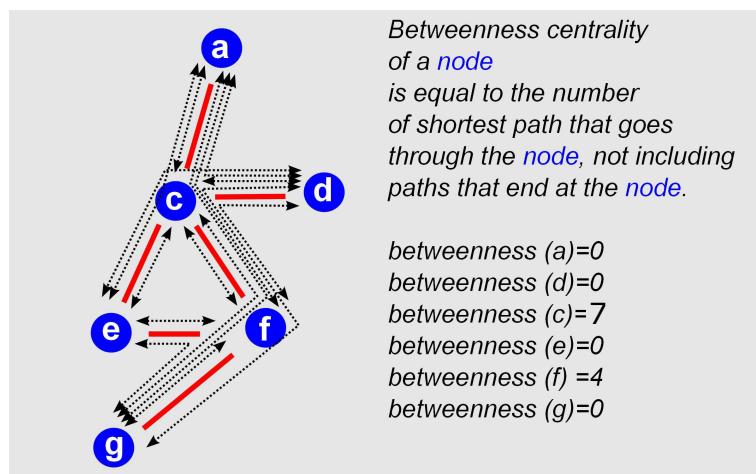


Figure 5.8: Illustrating the betweenness of a node. Betweenness measures the communication traffic that runs through the node, assuming that all pairs of nodes send messages to each other and always use the shortest path between them [58, 59].

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

Bi-directional links:
In network form:



In matrix form:

$$\mathbf{A} \mathbf{v}_1 = \begin{matrix} & \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{matrix} \\ \mathbf{A} & \end{matrix} \begin{matrix} \mathbf{v}_1 \\ \begin{matrix} 1 \\ 0 \\ 0 \\ 0 \end{matrix} \end{matrix} = \begin{matrix} & \begin{matrix} 0 \\ 1 \\ 0 \\ 0 \end{matrix} \\ \mathbf{v}_1 & \end{matrix}$$

$\begin{matrix} 1 & 2 & 3 \\ \text{to} & \text{to} & \text{to} \\ j & j & j \end{matrix}$

Directed links:
In network form:



In matrix form:

$$\mathbf{B} \mathbf{v}_1 = \begin{matrix} & \begin{matrix} 0 & 0 & 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix} \\ \mathbf{B} & \end{matrix} \begin{matrix} \mathbf{v}_1 \\ \begin{matrix} 1 \\ 0 \\ 0 \\ 0 \end{matrix} \end{matrix} = \begin{matrix} & \begin{matrix} 0 \\ 1 \\ 0 \\ 0 \end{matrix} \\ \mathbf{v}_1 & \end{matrix}$$

$\begin{matrix} 1 & 2 & 3 \\ \text{to} & \text{to} & \text{to} \\ j & j & j \end{matrix}$

Figure 5.9: **Matrix representation of a network.** A_{ij} mark links from node i to nodes $j = 1, j = 2, \dots$ to $j = N$. In right panel the network is directed reflected in an A_{ij} that is NOT symmetrical. In both cases the matrix formulation allows for simple computation of signal propagation on networks by using matrix multiplication.

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. 5.7.

Another noteworthy concept, is that of *betweenness centrality* of nodes, illustrated in Fig. 5.8. 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 [60, 61, 62].

Mini Tutorial: Draw or describe a network with a special node that has high betweenness and low connectivity.

5.1.4 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 $A_{ij} = 0$. This is illustrated in Fig. 5.9 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 . A directed networks will not be symmetric, see illustration in Fig. 5.12.

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 $v_j, j = 1, 2, \dots, N$. The dynamics of disease spreading is a process which allow copying to all

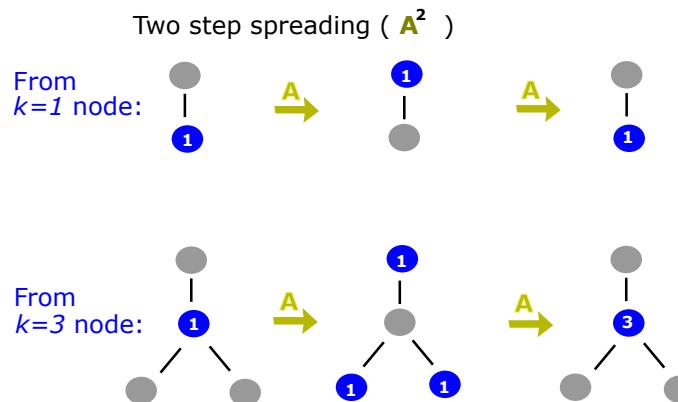


Figure 5.10: **Matrix multiplication and connectivity:** Applying the matrix twice in part means that each node gets contribution from itself through all its neighbours.

neighbor nodes, represented by the update:

$$v_i = \sum_j A_{ij} v_j, \quad (5.11)$$

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 , $t \rightarrow \infty$ is dominated by the eigenvector corresponding to the largest eigenvalue. The nodes with high representation in this eigenvector reflects the region in the network where self-amplification is strongest. In practice this will be network regions with high connectivity, and high number of closed loops.

Mini Tutorial: Why is high connectivity good for self amplification? Why is loops good for self amplification? (see Fig. 5.10)

Another interesting feature can be obtained by applying the matrix $\lim_{t \rightarrow \infty} \mathbf{A}^t$ 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 $\mathbf{A} \cdot \mathbf{A} = \mathbf{A}^2$. What does the trace of \mathbf{A}^2 represent for the network??

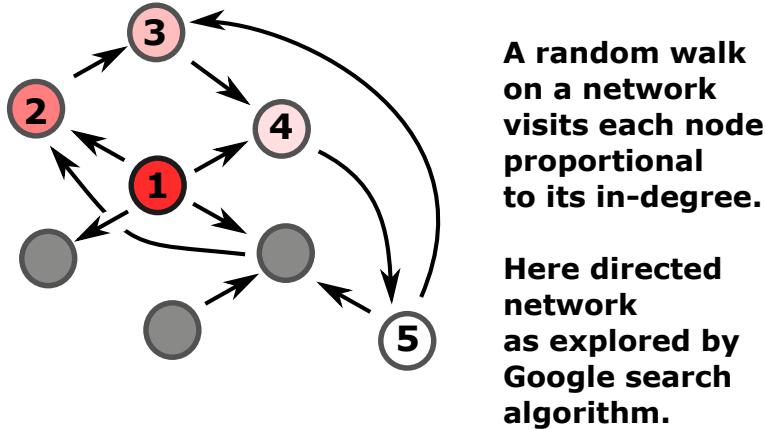


Figure 5.11: **Google search algorithm on directed network.** Take 5 random steps and then jump to another random node to start next 5 steps. The node is ranked with the number of times it is visited by this random walk. In text we discuss random walks on bi-directional networks (symmetric adjacency matrix).

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 [63, 64, 65, 66]. A modular network would be reflected in a nearly block diagonal matrix. The blocks along the diagonal will often be supplemented with a few non-zero entries at other places in the matrix.

Notice that the matrix representation opens for simple manipulations. The number of triangles in a non-directed network without self-links is

$$n(\Delta) = \frac{1}{6} \cdot \text{trace} (\mathbf{A}^3) \quad (5.12)$$

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 ($\text{trace}(\mathbf{A}) = 0$).

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 [64, 65, 67, 68, 66]. At each time the walker takes a random step along one of the (out) links from

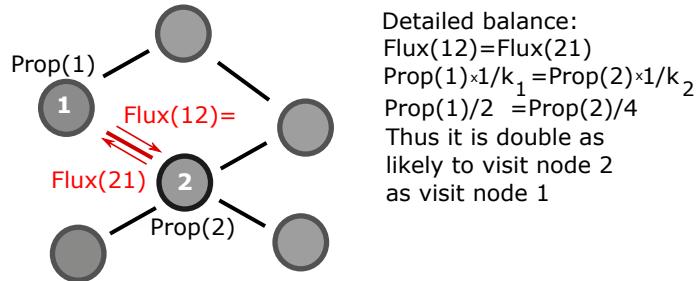


Figure 5.12: **Detailed balance for random walks on bi-directional networks** At steady state the probability/flux to go from node i to node j should be equal to the probability to go the opposite way. That is $\text{Flux}(i,j) = \text{Flux}(j,i)$

its current node see Fig. 5.11. The figure illustrate that 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 search engine ²⁾.

Now we turn to our usual bi-directional networks. Also here the random walk 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 of node i . Thus a node which has only one link will direct its random walkers that way. In contrast, for a node where there is 10 links the probability to walk away along a given link is only $1/10$. This is illustrated in Fig. 5.12 while Fig. 5.13 illustrate the transfer matrix for a corresponding bi-directional network.

Transfer matrix:

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{31} \end{bmatrix} = \begin{bmatrix} 0 & k_2^{-1} & k_3^{-1} \\ k_1^{-1} & 0 & 0 \\ k_1^{-1} & 0 & 0 \end{bmatrix} \quad \text{for the case where } A_{12}=A_{21}=1, A_{13}=A_{31}=1$$

only non-zero elements when $A_{ij}=1$

$$\mathbf{T} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & k_2^{-1} & k_3^{-1} \\ k_1^{-1} & 0 & 0 \\ k_1^{-1} & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ k_1^{-1} \\ k_1^{-1} \end{bmatrix}$$

corresponding to spreading $1/k$ to each neighbor

Figure 5.13: **Transfer matrix for a random walk on a network.**

If one considers a population of random walkers on a bi-directional network, the distribution of these walkers approaches a steady state in which the

²⁾Google originally used 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.

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 :

$$\frac{v_i^1}{k_i} = \frac{v_j^1}{k_j} = \text{constant}, \quad (5.13)$$

that is, the more connected, the more walkers will visit the node. This was also apparent from Fig. 5.12.

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. 5.14 and 5.15. Fig 5.14 shows a network constructed by using transmission of twitter messages (L. Weng et al, Scientific Reports 2012). In fact, random walkers can also be used to sample clusters in networks, because a walker typically tends to spend a lot of time in one cluster before it occasionally transits to another cluster. This is explored by using the eigenvectors associated to the largest eigenvalues (slowest relaxing modes) for \mathbf{T} in Ref. [64].

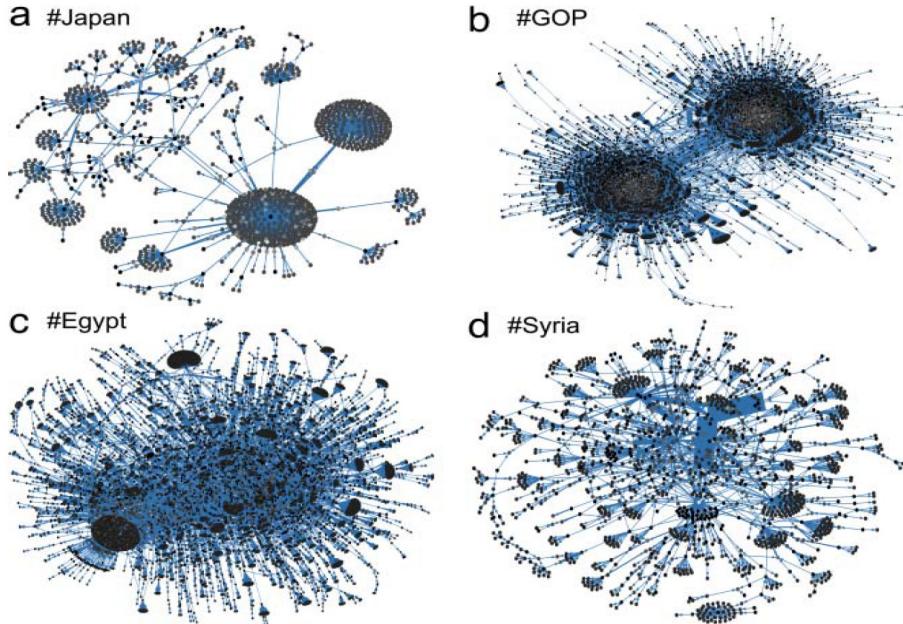


Figure 5.14: Networks constructed from links formed by retweets. A) is related to a Japanese earthquake in March 2011, B) Grand Old Party (GOP) is related to the republican party in USA, C),D) are related to the Arab Spring in 2011, focusing on Egypt and Syria, respectively.

Mini Tutorial: Inspect the CEO network on Fig. 5.15 Is the number of triangle loops larger or smaller than randomly expected?

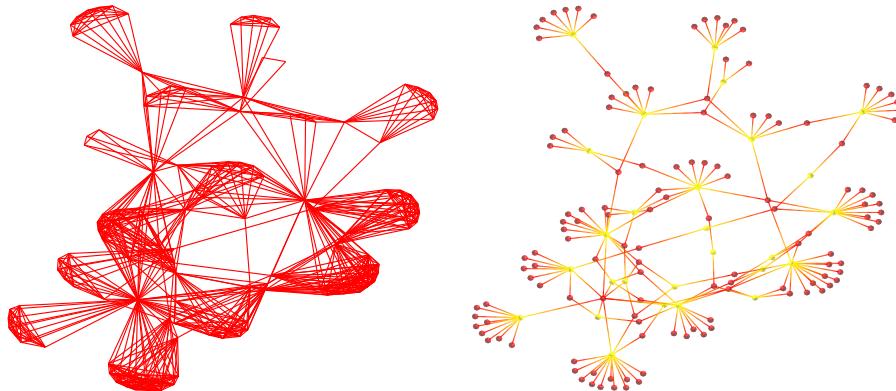


Figure 5.15: Professional social network. Small sub-section of the social network of company executives in USA. On left panel links are between the CEO's that sit on same board, the right panel shows the bipartite version of the network, where persons (red) and boards (yellow nodes) are separated.

5.1.5 “Scale free” networks

A common feature of many real networks is that their degree distribution is very broad [69, 70, 53], e.g., networks between genes in a cell (Fig. 5.1), between internet servers (Fig. 5.16) and between members of important social clubs (Fig. 5.15). 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,

$$n(k) \propto \frac{1}{k^\gamma}, \quad (5.14)$$

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.

Notably, such a scale-free degree distribution is far from the Poisson distribution from the previous sub-chapter (eq. 5.3). 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. 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 $\gamma = 1.8$?

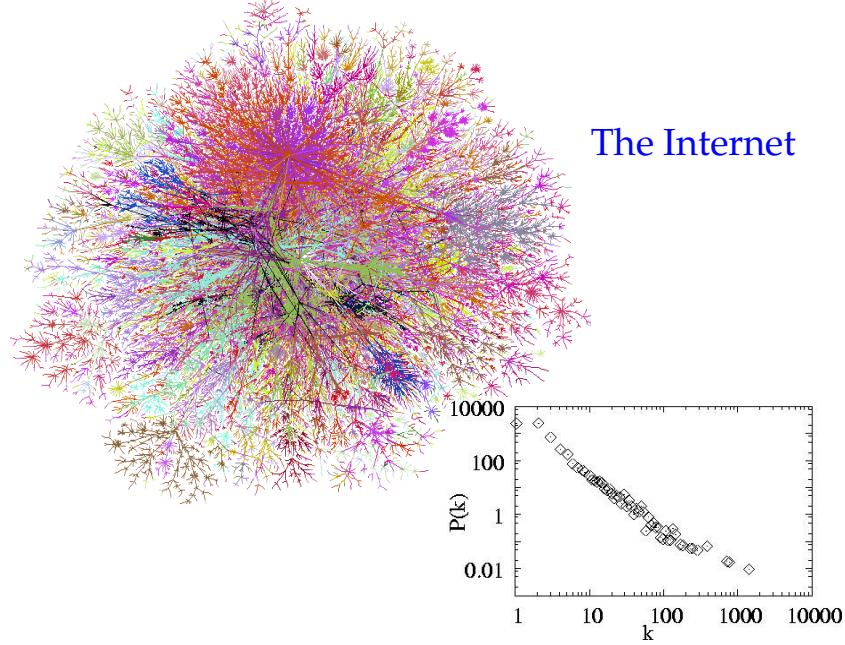


Figure 5.16: **Worldwide Internet:** Degree distribution in lower right corner. Notice the log-log scale, indicating the extremely broad range of degrees. Also notice that the distribution roughly fit a power law $P(k) \sim 1/k^{2.2 \pm 0.1}$

5.1.6 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. 5.8. 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 eq. 5.8,

$$\mathcal{A} = \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 = \frac{\int_1^N \frac{k^2 dk}{k^\gamma}}{\int_1^N \frac{k dk}{k^\gamma}} - 1 \sim N^{3-\gamma}, \quad (5.15)$$

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

³⁾In the literature [73] the derivation was first presented using the average connectivity of

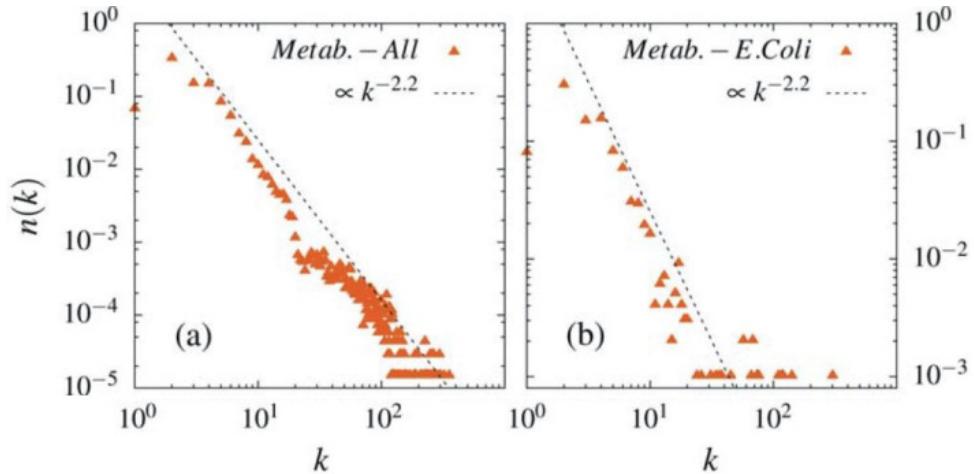


Figure 5.17: Degree distribution of metabolites in networks. Two metabolites are linked when one is entering the reaction (a substrate) and the other is a result (a product) of the same reaction. The left panel shows the average degree distribution for 107 organisms in the Ma-Zheng database [71]. The right panel shows the degree distribution from the metabolic network inside the bacteria *Escherichia coli*. Figure copied from ref. [72].

given potentially epidemic disease [74, 75, 76].

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 fraction, $1 - f$:

$$\mathcal{A} \rightarrow \mathcal{A}' = \mathcal{A} \cdot (1 - f) \quad (5.16)$$

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

$$(1 - f) > \frac{1}{\mathcal{A}} = \frac{1}{\langle k^2 \rangle / \langle k \rangle - 1}. \quad (5.17)$$

I.e. the percolation threshold for the network is $1 - f = 1/\mathcal{A}$: This is the fraction of nodes that need to be conducting to make an infinite cluster.

a nodes in the end of links

$$\kappa = \mathcal{A} + 1 = \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\text{Var}(k)}{\langle k \rangle} + \langle k \rangle$$

and one identifies the percolation threshold with this being = 2.

⁴Following a signal that enters into a node, see Fig. 5.4, each of its remaining $k - 1$ links have probability $(1 - f)$ to survive the pruning. Thus the local amplification $(1 - k) \rightarrow (1 - k) \cdot (1 - f)$ and the global amplification factor \mathcal{A} is reduced by the factor $(1 - f)$.

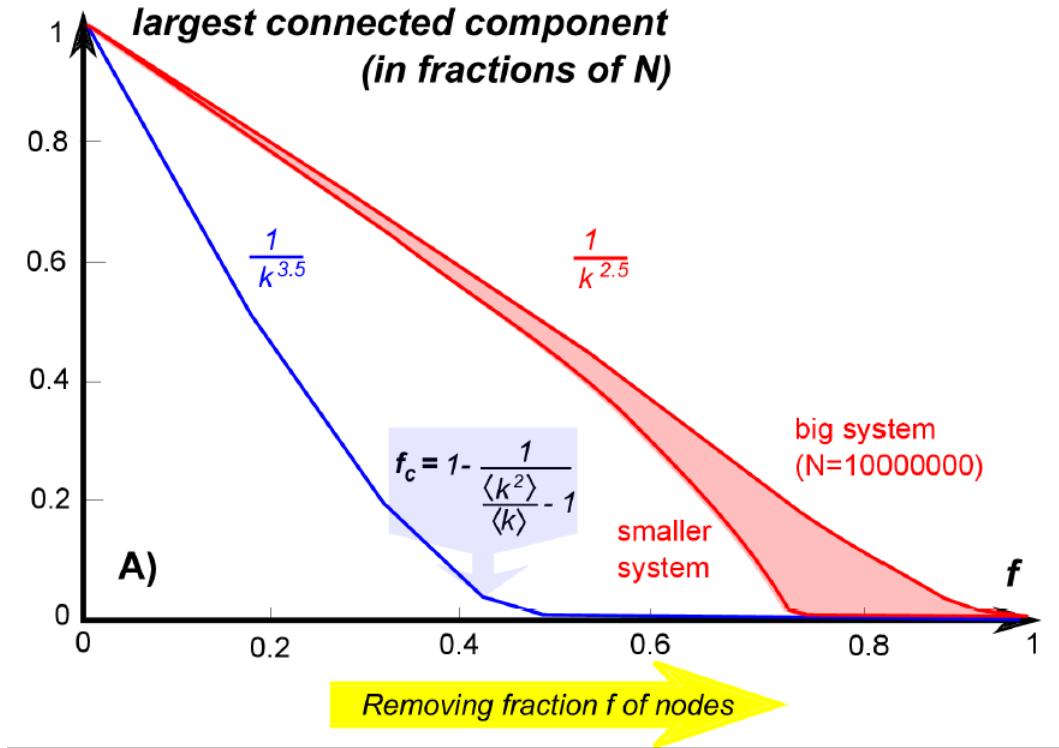


Figure 5.18: **Weight of the largest cluster.** A) The fraction of nodes that is in the largest cluster (F) as a function of the fraction of nodes removed[73]. The blue curve refers to a network with a relatively narrow degree distribution, $1/k^{3.5}$, which exhibits a critical threshold similar to 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.

Conversely, the critical fraction for vaccination against disease spreading [73] is

$$f_c = 1 - \frac{1}{\langle k^2 \rangle / \langle k \rangle - 1}. \quad (5.18)$$

This threshold is close to unity for scale free networks with degree exponent $\gamma < 3$, as is also seen in Fig. 5.18A. 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.

Mini Tutorial: Select one person (out of a total of N persons) to be in random end of a link. Why is the chance to not select a person of connectivity k equal to $1 - k/N \sim \exp(-k/N)$.

⁵After a fraction f is removed from an ER network, the fraction F in the largest cluster obeys $F = (1-f) \cdot (1 - e^{-\langle k \rangle \cdot F})$ [73]: That is, the probability that a given remaining node is 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. 5.18B.

One vaccination strategy in scale free networks is to ask random persons who their friends are, and then vaccinate these friends. Then if f nodes are vaccinated then a node with connectivity k will survive the vaccination process with probability $\exp(-f \cdot k)$. The degree distribution after vaccination is then changed:

$$n(k) \rightarrow n(k) \cdot \exp(-f \cdot k) \quad (5.19)$$

and the integral from eq. 5.15 change its upper limit to $k_{max} \sim 1/f$ and $\mathcal{A} \sim (1/f)^{3-\gamma}$. This is obviously much smaller than the amplification obtained with random vaccination.

Questions

5.1) 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).

5.2) 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.

5.3) 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 shortest path between nodes are short. But there is a lot of non-degenerate options when walking on them and it's easy to get lost.

5.4) Generate Erdös-Rényi networks with $p = 3/(N - 1)$ (3 neighbors per node on 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.

5.5) 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.

5.6) 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 algorithm.

5.7) 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 distributes the about 5+1 link per node randomly across the lattice).

Qlesson: There are much more triangles in the modular network.

5.8) 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.

Qlesson: There transition are sharper when removing links, because the largest component are less often depleted.

5.9) 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 [73]:

$$F = (1 - f) \cdot (1 - e^{-\langle k \rangle \cdot F}). \quad (5.20)$$

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 5.15.

5.10) One vaccination strategy is to vaccinate people at the end of links. By vaccinating a fraction f' of the nodes, one removes a fraction

$$f = f' \cdot \langle k^2 \rangle / \langle k \rangle \quad (5.21)$$

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.

5.2 Analyzing Network Topologies

5.2.1 Randomization: Constructing a proper null model

Mini Tutorial: How would you randomize a given network?

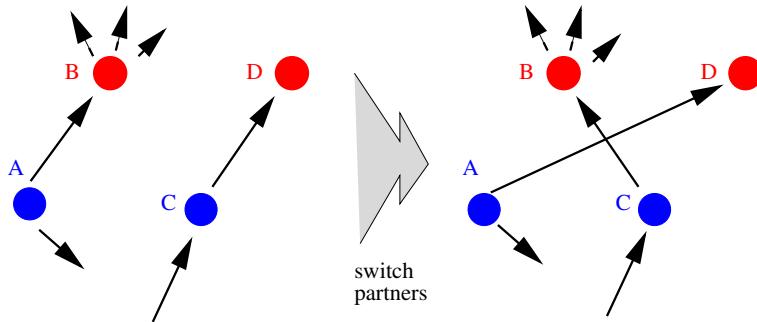


Figure 5.19: Network randomization. One step of the network randomization algorithm [77, 78]: A pair of directed links $A \rightarrow B$ and $C \rightarrow D$ switches connections in such a way that A becomes linked to D , while C becomes linked to B , provided that none of these resulting links already 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.

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 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 5.19. The idea is to swap links, pair by pair, multiple times until all nodes in the network are assigned new random links [77, 78]. At each swap there are 2 links that change, i.e. a probability of $2/L$ that any given link is reshuffled. For a system with L links, then after t link swaps, the probability that a given link is not changed is

$$\text{Fraction(unchanged links)} = (1 - \frac{2}{L})^t \approx e^{-2t/L}, \quad (5.22)$$

which becomes insignificant when the number of “swaps” t becomes substantially larger than L . More precisely, when fraction of non randomized links becomes less than $1/L$, meaning that less than 1 links is unchanged:

$$e^{-2t/L} \sim 1/L \Rightarrow t \sim \frac{1}{2} L \cdot \ln(L) \quad (5.23)$$

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 if one started with one big connected component one may keep this connected by only allowing “swaps” that maintain overall connectedness.

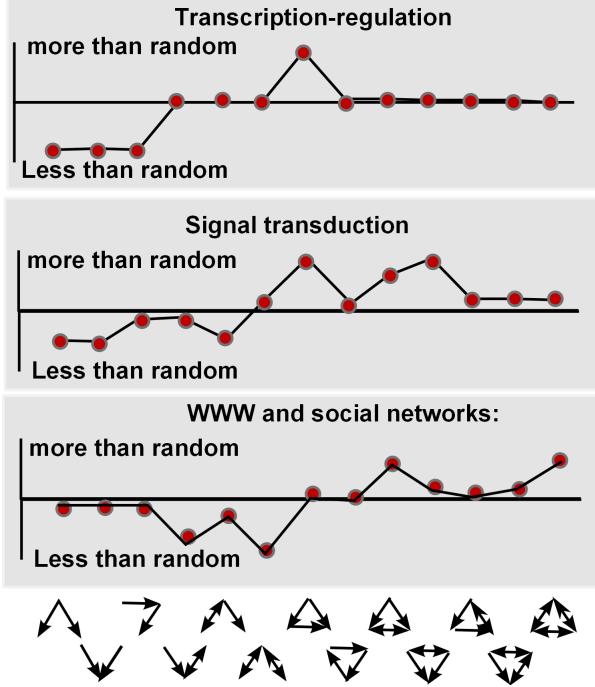


Figure 5.20: Network motifs with abundances. Super-families of motifs as suggested by statistical analysis of different types of networks by Milo et al. [79]. One sees that feed-forward is over-represented in transcription and signaling networks. In contrast, then social networks and the world wide web 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.

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

$$\text{Substance} = \text{Sub(pattern)} = \frac{N(Q)}{\langle N_{\text{random}}(Q) \rangle}, \quad (5.24)$$

whereas the the significance level of Q is quantified by its Z score:

$$\text{Significance} = Z_{\text{score}} = \frac{N(Q) - \langle N_{\text{random}}(Q) \rangle}{\sigma_{\text{random}}(Q)}, \quad (5.25)$$

where $N_{\text{random}}(Q)$ is the number of times the pattern occurs in the randomized network. Here

$$\sigma_{\text{random}}^2(Q) = \langle N_{\text{random}}(Q)^2 \rangle - \langle N_{\text{random}}(Q) \rangle^2 \quad (5.26)$$

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 patterns of control in various other networks, ref. [80, 81] found particularly frequent motifs in particular types of networks. Fig. 5.20 show their results in terms of over/under representation of motifs in different types of networks. For example for biological regulatory networks ref. [80, 81] report an excess of feed-forward motif and Ref. [82] suggested that these “feed-forward” acts as a noise filters. That is, they would only allow 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.

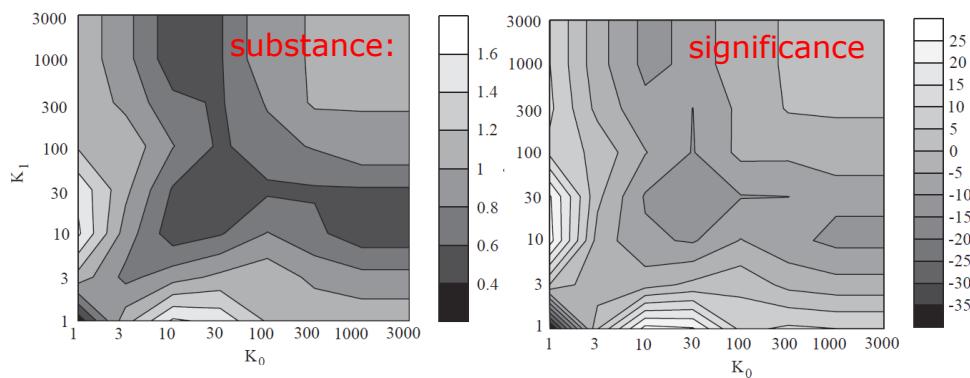


Figure 5.21: **Substance (Sub)** and **significance (Z)** of connections between nodes of different degree in the hardwired internet. The correlation [77] between internet servers is quantified in terms of probability to be connected in units of randomized expectation (left) and in terms of the Z-score (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 of 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. 5.21 shows relatively many links between nodes of similar degree (from ref. [77]). This is called an *assortative network*.

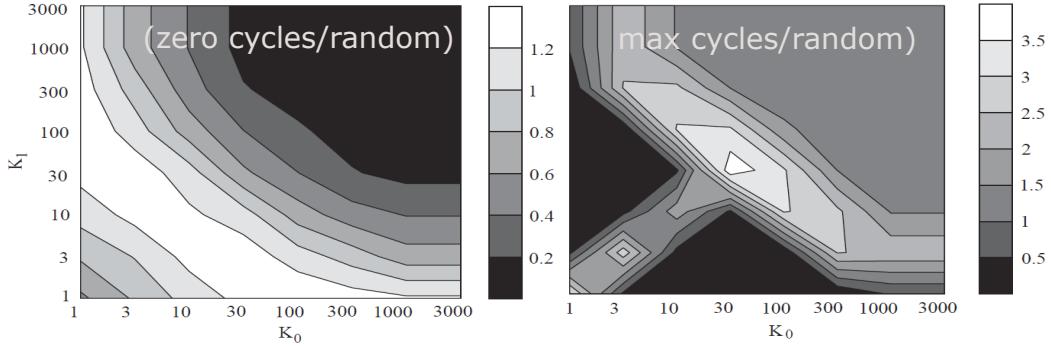


Figure 5.22: **Excess ratio Sub.** (eq. 5.24). Analyse of some artificial networks. In both cases the contour plots quantify correlations between connectivities of pairs of nodes that are directly linked to each other.

Left panel 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 show that zero-cycle network disfavour connections between highly connected nodes, compared to a randomized version that preserve the degree distribution.

Right panel we rewire the hardwired internet to maximize the number of triangles (obtaining a network with 59,144 cycles of length three). One then obtain a network where high degree nodes are more likely to be connected.

Fig. 5.22 elaborate further on the relation between triangle motifs (in non-directed networks the triangles correspond to most of the motifs from Fig. 5.20) 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 $Diff = N_{\text{after}}(\Delta) - N_{\text{before}}(\Delta)$. 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 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. 5.22, 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. 5.25), respectively dis-assortative (nodes of similar degree avoid each other).

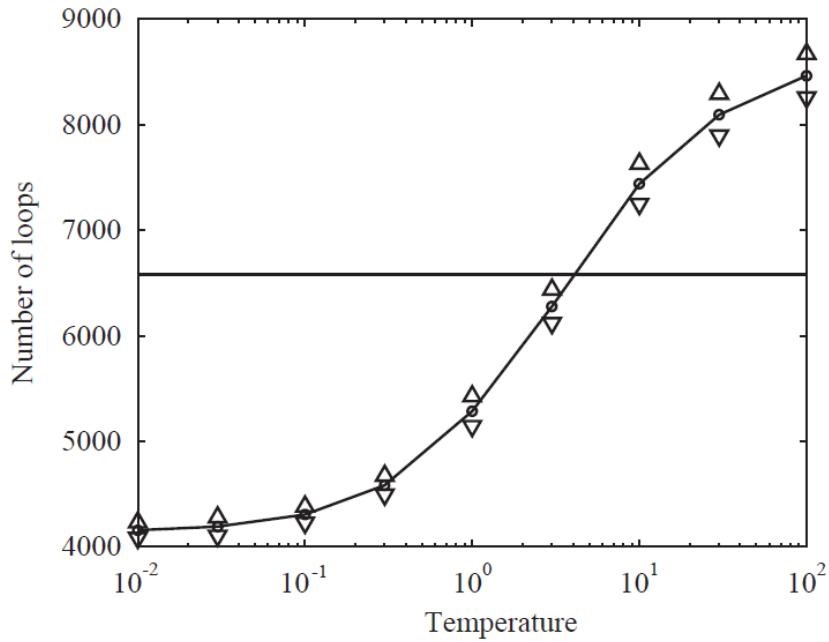


Figure 5.23: **Use of metropolis sampling of constrained networks:** Number of cycles of length three for rewired internet while fixing correlation profile (having 6,584 cycles of length three). The number of triangles depends on the temperature used. When temperature is zero, this corresponds to the 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.

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. 1.3), assigning an energy to the deviation of observed correlation profile:

$$H = \sum_{K_0, K_1} (N(K_0, K_1) - N_{\text{random}}(K_0, K_1))^2 / N(K_0, K_1), \quad (5.27)$$

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. Maintaining H small one attempt to keep the same number of connections between say nodes of connectivity $K_0 = 10$ and nodes of say connectivity $K_1 = 25$ as there was in the original network (i.e. $N_{\text{random}}(K_0, K_1) \sim N(K_0, K_1)$ for all pair of connectivities K_0 and K_1). At each update step one now attempts a link swapping, and calculates the change in H , $\Delta H = H(\text{new}) - H(\text{old})$.

- If $H(\text{new}) < H(\text{old})$, move is accepted.
- If $H(\text{new}) > H(\text{old})$: move is accepted with prob. $p = e^{-(H(\text{new}) - H(\text{old})) / T}$.

where “ T ” is some ad hoc assigned effective temperature for our sampling. Fig. 5.23 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.

5.2.2 Algorithm generating a synthetic scale-free network

In subsequent text we investigate topology of networks with a given fixed degree distribution. This is not entirely trivial to obtain: when the degree distribution is rather broad/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 is possible to find enough partners also for the high degree nodes.

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, for each node i one selects a random number $r_i \in]0, 1[$ and solve for K

$$\frac{\int_K^N dk/k^\gamma}{\int_1^N dk/k^\gamma} = r_i , \quad (5.28)$$

where $k_i = K$ 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 [83]. 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 ($(1 - 2/L)^t \sim e^{-2t/L} \sim 1/L$).

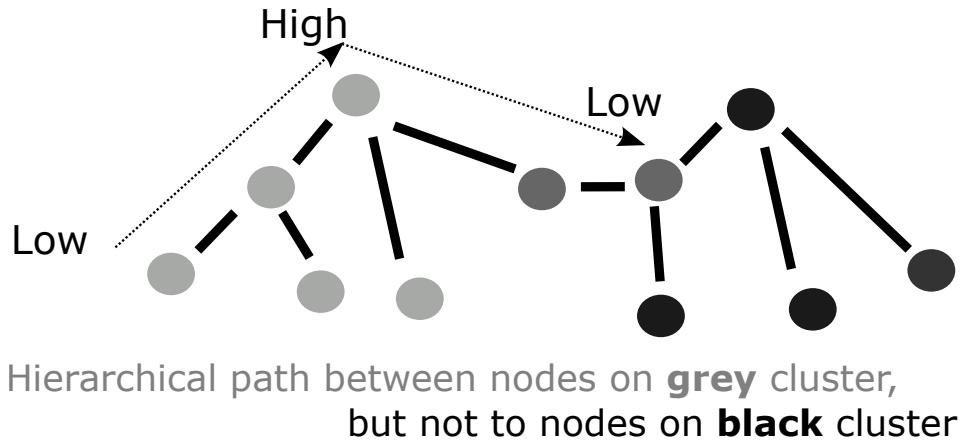


Figure 5.24: **Topological Hierarchy.** Illustration of hierarchy, and broken hierarchy. Hierarchical network is when one can go from low to top to low and then reach all nodes on a network (always along shortest paths). If one have to pass some valleys of low hierarchy then the hierarchy is not complete.

5.2.3 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 [83, 84], which assigns rank proportional to degree. That is, the shortest path first has go from low to high degree and then from high to low degree. This is illustrated in Fig. 5.24. The fraction of pairs in the network that is connected by such hierarchical paths is denoted as \mathcal{H} :

$$\mathcal{H} = \frac{\sum_{ij} (\text{shortest path between } i \text{ and } j \text{ hierarchical})}{N(N - 1)/2} \quad (5.29)$$

where the sums runs over all the $N(N - 1)/2$ pairs of nodes in the network of size N . $\mathcal{H} = 1$ if the shortest path between any pair of nodes follows a hierarchical path.

Fig. 5.25 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 often not directly connected to each other, and even the random network is anti-hierarchical.

The fact that exponents close to 2 are close to perfectly hierarchical is because it is very different for the very large hubs in these networks to connect to each other. For example imagine a network with a node i with 500 connections and a node j with 1000 connections, that in total have 10000 links and $\sum_x k_x = 20000$. Each of the links from node i have probability

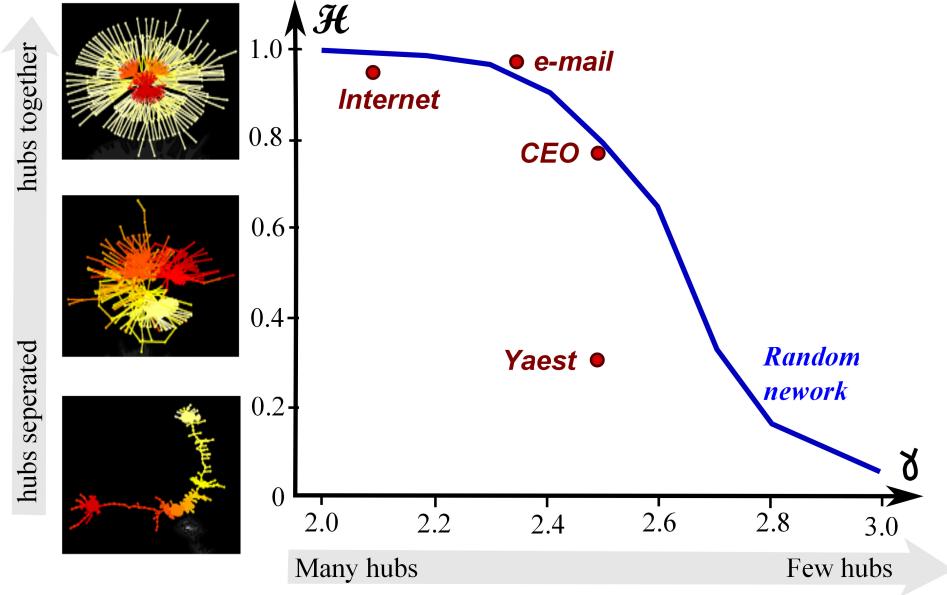


Figure 5.25: **Topological Hierarchy.** Reproduced from [83]. The left part of the figure illustrates maximally hierarchical (top) and anti-hierarchical (bottom) networks of size $N = 400$ with a $1/k^{2.5}$ degree distribution. The main figure shows how \mathcal{H} depends on the degree distribution 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.

$1000/20000=0.05$ to be connected to j . Thus that 500 links should not be connected to j would be $(1 - 0.05)^{500} \sim 10^{-11}$. Thus we are nearly certain that they are connected, and thereby that they merge to the same hierarchy with j on top.

The figure 5.25 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. 5.26):

$$\begin{aligned} P(K \text{ node is local top}) &= \left(1 - \frac{\int_K^N k^{1-\gamma} dk}{\int_1^N k^{1-\gamma} dk} \right)^K \\ &\sim (1 - K^{2-\gamma})^K \\ &\sim \exp(-K^{3-\gamma}), \end{aligned} \quad (5.30)$$

where we use that, when $\gamma > 2$, the integral is dominated by its lower boundary. Subsequently, we use that $1 - K^{2-\gamma} \sim \exp(-K^{2-\gamma})$.

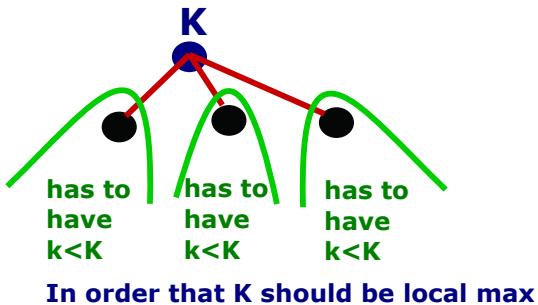


Figure 5.26: **Detecting a hierarchical network.** Calculation of the probability that a node with connectivity K has higher connectivity than all its neighbors. When this happens for nearly all highly-connected nodes, then the network will be characterized by highly-connected hubs that are not linked up to each other: A non-hierarchical network results.

Thereby $P(K \text{ node is local top})$ becomes large for $\gamma \sim 3$, even when K is rather small.

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 [83]. That is network with steep power laws is non hierarchical.

If one instead consider a network with $\gamma \sim 2$ 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 is then the only “top” in the system. and 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 hierarchical 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. 5.27).

In particular, in Fig. 5.28 we complement the methods to generate random networks (random one-mountain landscapes) [77] with preserved degree sequences, to generate *ridge landscapes*. In its simplest implementation, we assign a random rank $R_i \in [0, 1]$ to every node i in the network, and organize the nodes hierarchically based on their rank. That is one start with a random network and swaps pairs of links to minimize $Dist = ABS(R_i - R_j) + ABS(R_k - R_l)$ between the neighbours. One make many swaps and always accept swaps where $Dist$ decreases, and only allow increase with probability ε . This method creates non-random networks, distinguished by a separation

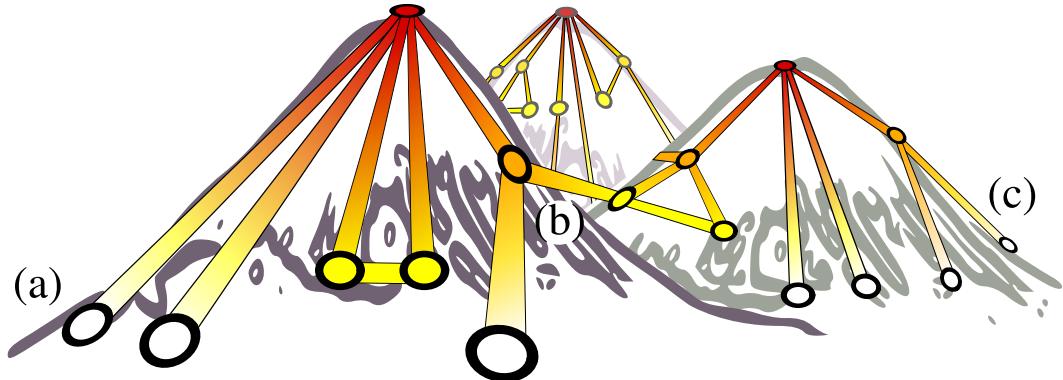


Figure 5.27: **A network as a degree landscape with mountains and valleys.** The "altitude" of a node proportional to its degree. A route over one mountain corresponds to making a connectivity-hierarchical path ((a) to (b)) while climbing over more than one mountain breaks the degree-hierarchical path ((a) to (c)).

of hubs (leftmost network). Alternatively one may assign each node a rank number $R_i = k_i/k_{max}$ equal to its connectivity, and thereby generate the highly centralized system in the rightmost panel.

Thus, Fig. 5.28 shows topologies that all originate from a random scale-free network (shown in Fig. 5.28e) with degree distribution $P(k) \propto k^{-2.5}$ and system size $N = 400$. The extreme networks, the perfect random-rank hierarchy in Fig. 5.28(c) and the perfect degree-rank hierarchy Fig. 5.28(g) ($\varepsilon = 0$), surround the networks with increasing error rate towards the random scale-free network with $\varepsilon = 1$ in the middle (Fig. 5.28(e)).

The intermediate networks are generated by a rewiring rule where links also sometimes, with probability $\varepsilon > 0$ accept swaps where *Dist* increases. 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. 5.28(d). Note that the color gradient indicates that the random-rank hierarchy remains nearly, and that the hubs ("mountain tops") mostly remain 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

- 5.11)** 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 $k = 1, 2, \dots, N$. Compute

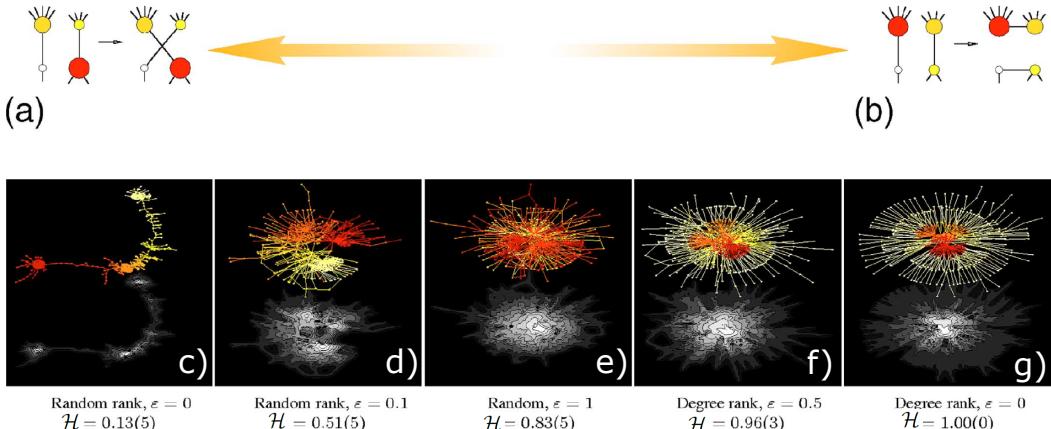


Figure 5.28: **Degree landscapes:** (c), via random landscapes (e), to peaked one-mountain landscapes based on degree (g). The links are swapped pairwise to connect high-ranked nodes and organize the nodes globally according to their rank (color coded from red for high rank, to white for low rank), with random swaps at different rates ε . The rank $R_i \in [0, 1]$ is set randomly to the nodes, as in the swap example in (a), in (c-d), and proportional to the degree of the nodes $R_i = k_i/k_{max}$, 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 $N = 400$. F mark the connectivity based hierarchy of the generated network.

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.

5.3 Models for Scale free networks

Mini Tutorial: Imagine that you jump from article to article, following random entries in the respective reference lists. At a random time you then make a random reference to one of the articles you visit. What is then the probability to reference an article as function of how often it was cited?

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

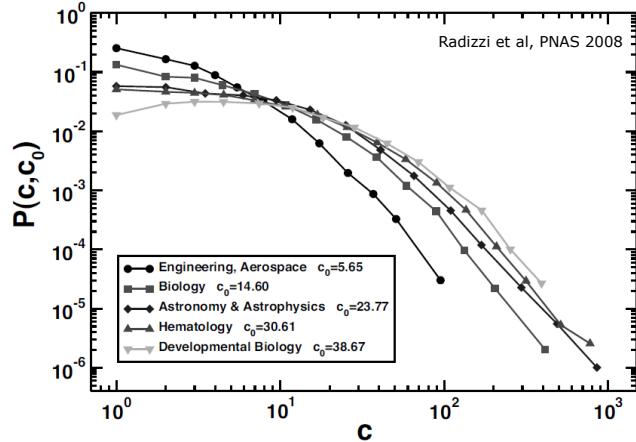


Figure 5.29: **Citations per article:** Notice that as the average of the field changes, so does the tail. The distribution is not really a power law but closer to a log-normal.

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.

5.3.1 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 , see Fig. 5.30. The evolution of n is given by (Bornholdt et al. (2001)):

$$\begin{aligned} n(k, t+1) - n(k, t) &= \frac{(k-1) \cdot n(k-1, t) - k \cdot n(k, t)}{\sum_{k'} k' n(k')} \quad \text{for } k > 1 \\ \Rightarrow \frac{dn(k, t)}{dt} &= -\frac{1}{\sum k' n(k')} \frac{d}{dk} (k \cdot n(k, t)) \quad \text{for } k > 1, \end{aligned} \quad (5.31)$$

because the probability to add a link to a specific node of connectivity k is $k / \sum k n(k)$. In the above equation the first term represents the addition of a

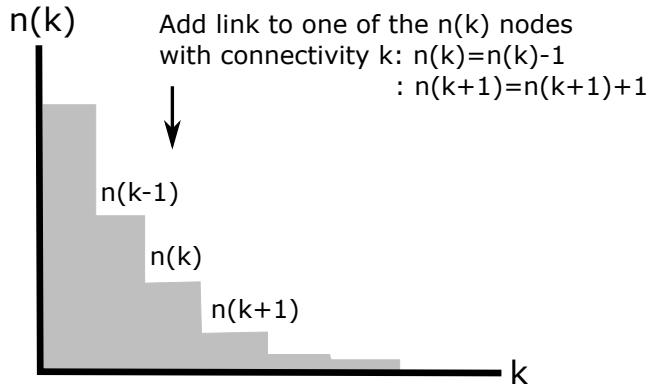


Figure 5.30: **Gain and loss for $n(k)$ with preferential attachment.** Notice that the probability to actually select a node includes its connectivity k

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:

$$\frac{dn}{dt} = -\frac{1}{2t} \frac{d(k \cdot n)}{dk} \quad (5.32)$$

To solve this equation we make the “ansatz”:

$$n(k, t) = f(t) \cdot N(k), \quad (5.33)$$

which gives:

$$2t \frac{f'(t)}{f(t)} = -\frac{1}{N} \frac{d(kN)}{dk}, \quad (5.34)$$

where $f(t) \propto t$ since $f(t) = 2t / \sum_k kN(k)$. Inserting $f'/f = 1/t$ in the above equation

$$-1 - \frac{dln(N)}{dln(k)} = 2. \quad (5.35)$$

From this one obtains $N \propto k^{-3}$:

$$n(k) \propto \frac{1}{k^3}. \quad (5.36)$$

The $1/k^3$ scaling remain unchanged if one added two links each attached to this new node. This is because both the nominator and denominator in eq. 5.31 would double, leaving scaling unchanged. If, however, one sometime add links without adding new nodes, see Fig. 5.31.

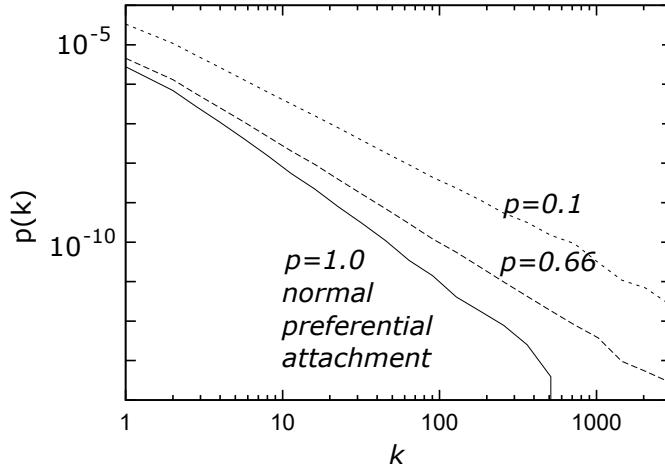


Figure 5.31: *Connectivity distribution in a preferential attachment model for three different values of p .* p is a parameter that specifies the fraction of times one adds a node with a link. At probability $1-p$ one instead adds links where each end of the link is attached to an existing node with 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 $p = 1$.

If such links are assigned to existing nodes preferentially in both ends they would contribute different in nominator and denominator, and change the scaling law $1/k^3$ to $1/k^\gamma$ with $\gamma \in [2; 3]$, see Fig. 5.31 In fact,

$$\gamma = \frac{2}{2-p} + 1$$

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 two already existing nodes. Thus we require an outside agent, that favors assignment of links between highly connected nodes, a requirement that goes against the idea of self-organization.

Tutorial: Prove the above equation for $\gamma(p)$ by revising eq. 5.31 appropriately.

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'économie 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

$$n(s) \sim 1/s^2; . \quad (5.37)$$

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

$$\langle s \rangle = \int_{\min}^{\max} \frac{s ds}{s^\tau} \quad (5.38)$$

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 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.

Preferential growth with random elimination (N=1000):

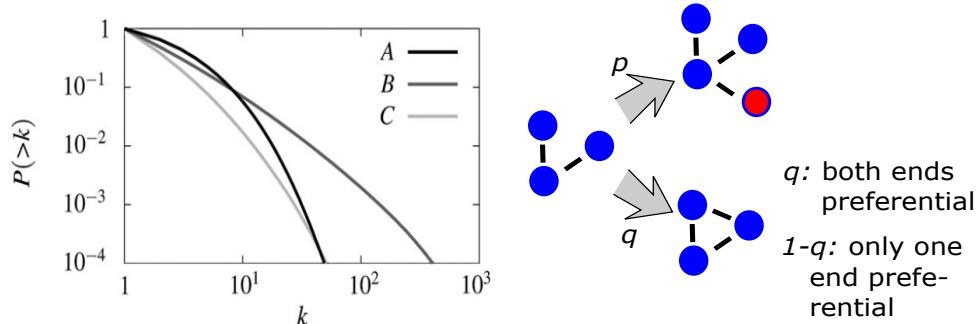


Figure 5.32: Preferential growth with random elimination of nodes. Nodes are randomly eliminated, when the total node number exceeds 1,000 (from :Gronlund et al., *Physica Scripta* 71, 680 (2005)). A: $p=0.4$, $q=0$, B) $p=0.4$, $q=1$ and C) $p=0.8$ and $q=1$ where p is how often one adds a new node 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.

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 an 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, if we abandon the ongoing growth and consider some sort of steady state with growth and removal, then the resulting network will not be scale free.

Mini Tutorial: Argue why rich-get-richer dynamics could be considered for sizes of families of related species on ecological scale (Youle's law).

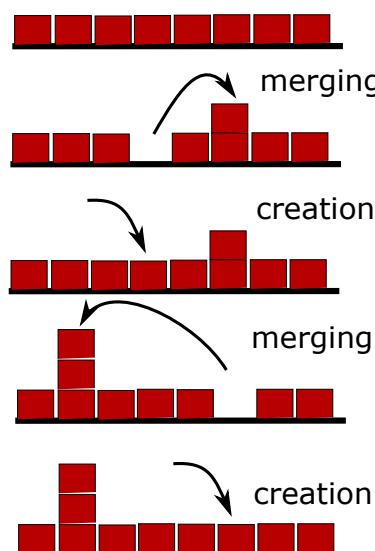


Figure 5.33: **Merging creation between a bunch of numbers** Two steps of the process.

5.3.2 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) and illustrated in Fig. 5.33:

- Consider a set of N numbers, s_i , $i = 1, 2, \dots, N$. 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 $s_i(\text{new})$: $s_i(\text{new}) = s_i + s_j$. Then reset the number at position j to unity: $s_j = 1$.

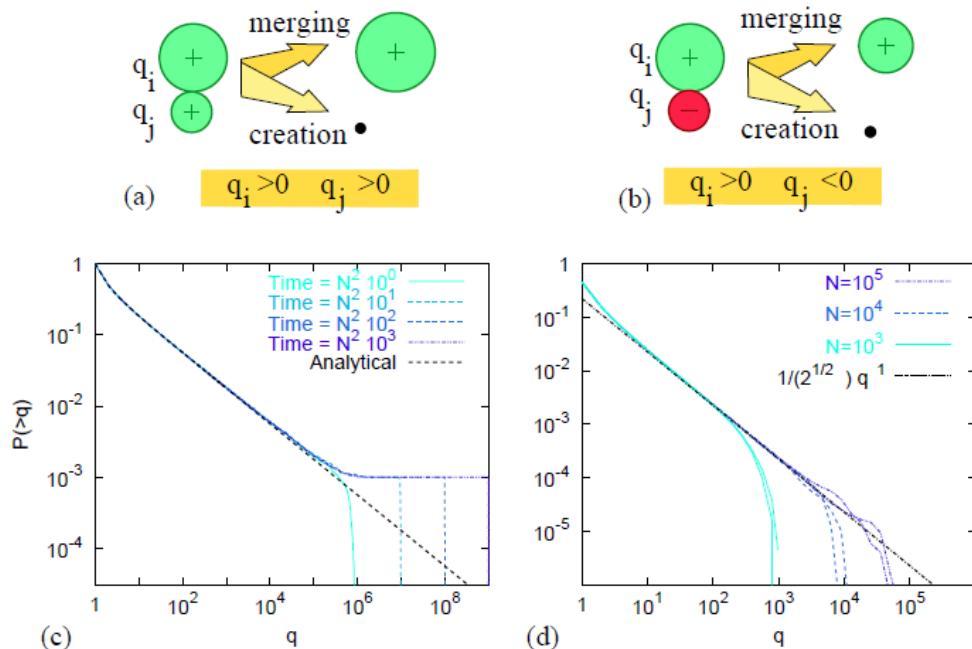


Figure 5.34: **Basic merging and creation process.** (a,c) also implemented while allowing different signs (b,d). Panels c,d show cumulative plots at different times for an $N = 1,000$ system. From Minnhagen et al. *Physica A* 340, 725 (2004).

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 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

$$\begin{aligned} p(s, t+1) - p(s, t) &= 0 \\ \Rightarrow \sum_{u=1}^{s-1} p(s-u) \cdot p(u) &= 2 \cdot p(s) . \end{aligned} \quad (5.39)$$

The first term represents the sum of all combinations of merging that result in a size-\$s\$ cluster. The second (loss) term $2 p(s)$ comes from selecting two numbers, which each could be of size \$s\$ with probability $p(s)$.

When a number of size \$s\$ is selected, it will be merged and then become larger than \$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 accumulated mass will grow as the largest cluster grows. This then separates from the power law that govern the remaining population of clusters and become a “sun like” system slowly sweeping up random other clusters securing that the sum of masses of everyone except the largest remain roughly constant.

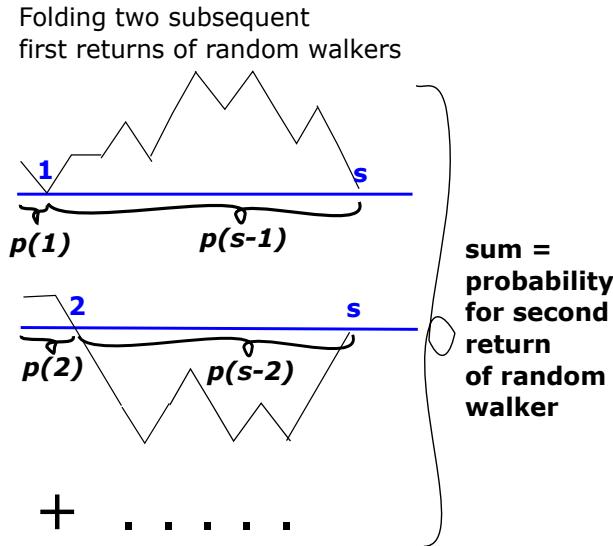


Figure 5.35: **Illustration of random walks.** First return distribution $p(s)$ where s is the time axis. The y-axis is not specified, and not related to the merging problem. The second return distribution of a random walker must fulfil $p_{second}(s) = \sum_{u=1}^{s-1} p(u)p(s-u)$. Further, we make 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, illustrating the connection to branching trees). First returns of random walkers scale as $p(s) = s^{-3/2}$.

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

$$p(s) \propto \frac{1}{s^{3/2}} \quad (5.40)$$

$(\sum_u u^{-3/2}(s-u)^{-3/2} \approx 5.22 \cdot s^{-3/2}$, here numerically simulated). 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. 5.39 is fulfilled for the first return of random walkers (now along an imaginary time variable s , and the return is to the starting point zero).

The above point is illustrated in Fig. 5.35 with the terms in the above sum, one that return at time “k” and a subsequent one that return at “time s. The thin line mark the random walk for each of the corresponding contributions to the second return of the random walker. 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 $s - 1$. Thus the explanation amounts to assuming that that second returns scale as first returns for random walkers. Notice that the basic equation also is found in equation for critical branching processes.

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. 5.34. As already mentioned, then the model also generates one very large cluster 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. In this revised model one at each step

- Select two numbers s_i and s_j of which one should be larger or equal 1. These two number are merged but with removal of 1 from the aggregate: $(s_i, s_j) \rightarrow (s_i + s_j - 1, r)$. The injection r now is some random injection number placed at position r with mean $\langle r \rangle = 1$.

Thus we here allow vacuum numbers (of size 0). This model amounts to a ”mass conserving” version of the standard merging model. It provides a scaling that is markedly steeper,

$$p(s) \propto 1/s^{5/2} \quad (5.41)$$

This was shown by solving the 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, average mass conservation 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 with 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 its neighbors and merge them into one node.

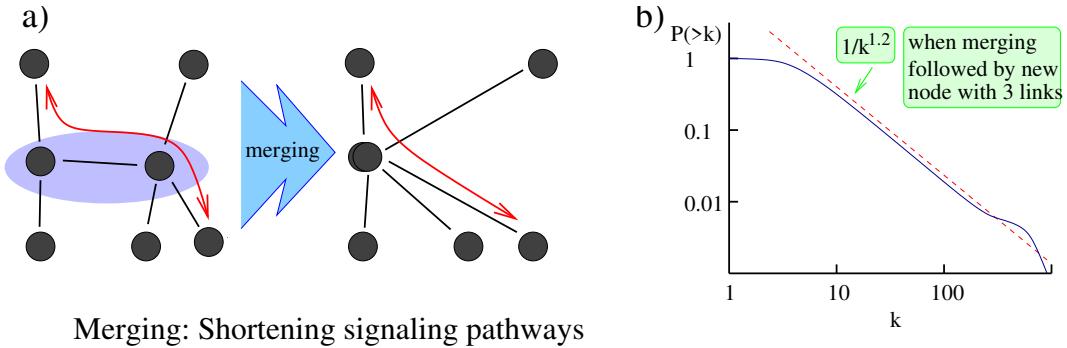


Figure 5.36: *Merging and creation model of Kim et al. (2003).* In addition to the merging step shown, a steady state network demands addition of a node for each merging. After a transient this evolutionary algorithm generates networks with scale-free degree distributions, as illustrated in right panel. 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.

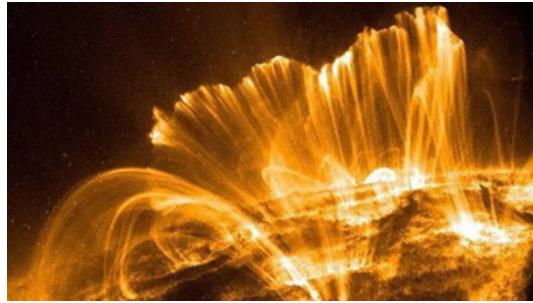


Figure 5.37: **Solar corona with network of magnetic field lines.**

- Adds one new node to the network and link it to a few randomly selected nodes.

The model is illustrated in Fig. 5.36a). 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. 5.36b). On another note, considering companies that may consider to merge with others. The justification for merging two companies could be efficiency and to shorten communication pathways. Creation, on the other hand then reflect introduction of new companies and their few start up relations.

In contrast to the preferential attachment model, 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.

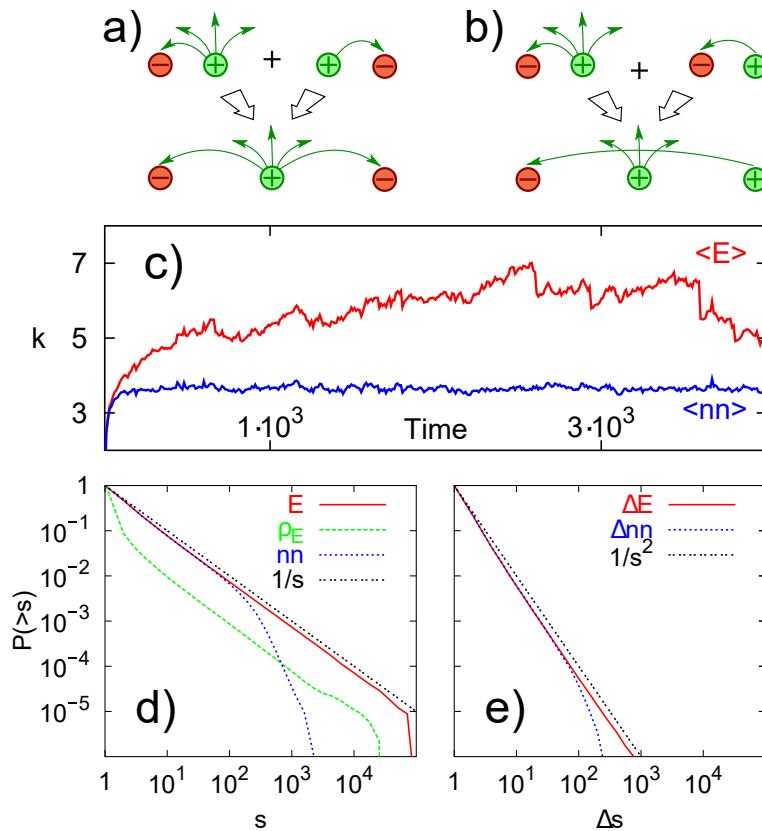


Figure 5.38: Bi-partite merging/annihilation model a) and b) illustrate merging events. Positive vertices (donor nodes) have outgoing links and negative (acceptor nodes) have incoming links. c) The dynamics of the average number of links associated to each node, $\langle E \rangle$, (upper curve) and the average number of neighbors $\langle nn \rangle$ (lower curve), d) The cumulative probability distributions total number of links 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 are all 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 $\tau = 2\gamma - 1 = 3$.

Merging in bi-partite Networks:

There is also a merging-creation model that have some potential relevance to solar flare dynamics [85]. Solar flares are 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 multiple magnetic north poles or south poles on the surface of the sun, see Fig. 5.37. These field lines may merge and grow or annihilate each other dependent on directions.

The above “mess” in the solar atmosphere suggest a simple model with donor ($q > 0$) and acceptor ($q < 0$) nodes to be connected via directed links, see Fig. 5.38(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 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. 5.38(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. 5.38(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

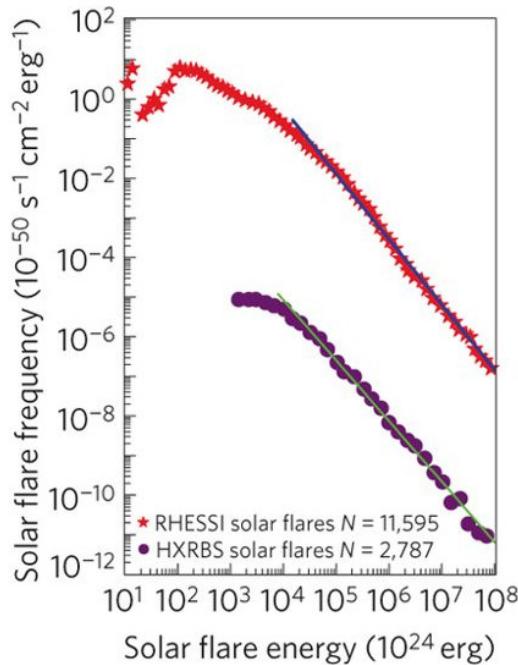


Figure 5.39: **Energy distribution in solar flares.** Reproduced from [15], suggesting that $P(E) \propto 1/E^{1.65}$

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:

5.12) 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 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.

5.13) 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.

5.14) 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 5.31 grows with 2 but the nominator with less. Thereby $\gamma = 1 + 2/(2 - p)$ that approaches 2 when $1 - p \sim 1$.

5.15) 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 , $i = 1, 2, \dots, n$, (with $n = 1,000$) which are updated according to

$$k_i, k_j \rightarrow k_i = k_i + k_j \text{ and } k_j = 1. \quad (5.42)$$

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.

5.16) Simulate the merging/creation model in terms of a set of integer numbers k_i , $i = 1, 2, \dots, n = 1000$, which are updated according to

$$k_i, k_j \rightarrow k_i = k_i + k_j - 1, \text{ and } k_j = \delta, \quad (5.43)$$

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.

5.17) Consider the random merging break up model of $i = 1, \dots, 1000$ numbers, each larger or equal to zero:

$$k_i + k_j \rightarrow k'_i + k'_j \quad (5.44)$$

Thus numbers just merge and then split up in whatever fraction between 0 and $k_i + k_j$. Set all $k_i = 1$ at time zero. Convince yourself that the change obey detailed balance. Plot time averaged distribution of numbers (see also fig 5.40). **Qlesson:** This is a model that preserve detailed balance. Thus the result is like an equilibrium distribution with a characteristic size.

Lessons:

- Networks are both about connecting large systems with heterogeneous components, but also about keeping individual nodes locally protected from information overflow.

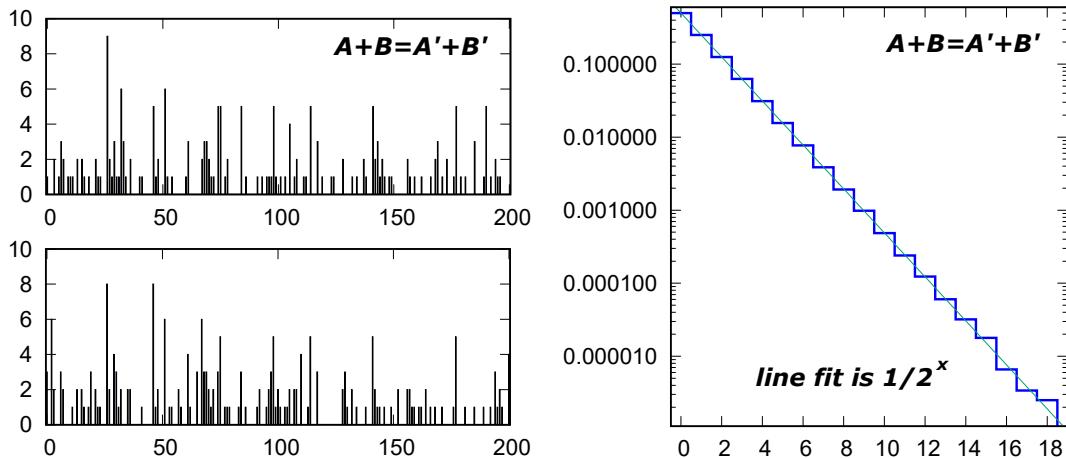


Figure 5.40: *Merging and fragmentation with detailed balance. the rules could resemble the exchange of energy between molecules that collide. I.e. a conserved quantity that is randomly distributed at each collision.*

- Real networks are often scale free $n(k) \propto \frac{1}{k^\gamma}$ with $\gamma \in [2, 2.5]$. These are amazingly “efficient” in transmitting diseases because of a large amplification factor

$$\mathcal{A} = \frac{\int k^2 n(k) dk}{\int k n(k) dk} - 1,$$

This quantity can also be used to estimate the percolation threshold when vaccinating/eliminating a fraction f of the nodes in the network; $\mathcal{A} \rightarrow \mathcal{A} \cdot (1 - f)$.

- Real Networks are formed by some dynamic process, taking place on much longer time scales than the dynamics on the network. This chapter suggested two possible dynamics that generate power laws: “Rich gets richer” and perpetual *merging* with noisy 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.

5.4 Appendix: Generating function for the merging process

Here we present a direct solution for the simple merging problem using Generating functions (proof by Ruijie Wu). The problem is equivalent to the earlier suggested critical branching process from the SOC chapter. 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

$$G(x) = \sum_{i=1}^{\infty} p_i x^i \quad (5.45)$$

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

$$\begin{aligned} G^2(x) &= \sum_{i=1}^{\infty} p_i x^i \sum_{j=i}^{\infty} p_j x^j \\ &= \sum_{k=2}^{\infty} \sum_{i=1}^{k-1} p_i p_{k-i} x^k \\ &= \sum_{k=2}^{\infty} 2p_k x^k, \end{aligned} \quad (5.46)$$

where we in the last equation use the basic equation defining p_k , i.e. the equation for critical branching (apart from a factor 2):

$$\sum p_i p_{k-i} = 2p_k \quad (5.47)$$

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 :

$$\begin{aligned} G^2(x) &= 2 \cdot (G(x) - p_1 \cdot x) \Rightarrow \\ G(x) &= 1 \pm \sqrt{1 - 2 \cdot p_1 \cdot x}. \end{aligned} \quad (5.48)$$

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 $G(1) = 1 \Rightarrow p_1 = 1/2$:

$$G(x) = 1 - \sqrt{1 - x}. \quad (5.49)$$

We can now Taylor expand this expression for the generating function, yielding

$$G(x) = G(0) + G'(0)x + \dots + \frac{1}{i!}G^{(i)}(0)x^i + \dots , \quad (5.50)$$

with the i 'th derivative equal to

$$G^{(i)}(x) = \frac{1 \cdot 3 \cdot 5 \cdot 7 \dots (2i-1)}{2^i} \cdot \frac{1}{(1-x)^{(2i+1)/2}} = \frac{(2i)!}{2^i \cdot i!} \cdot \frac{1}{(1-x)^{(2i+1)/2}},$$

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:

$$G(x) = \sum_{i=1}^{\infty} \frac{1}{4^i} \cdot \frac{(2i)!}{i! \cdot i!} \cdot \frac{1}{2i-1} \cdot x^i \quad (5.51)$$

Now we identify each order in x between this expression and the definition equation for the generating function $G(x)$:

$$p_n = \frac{1}{4^n} \frac{(2n)!}{n! \cdot n!} \frac{1}{2n-1} \quad (5.52)$$

Using Stirling's approximation $n! \sim \sqrt{2\pi n} \cdot (n/e)^n$:

$$p_n \propto \frac{1}{4^n} \frac{\sqrt{2n}}{n} \frac{(2n/e)^{2n}}{(n/e)^{2n}} \frac{1}{2n-1} \sim \frac{1}{\sqrt{n}(n-1)} \sim 1/n^{3/2}, \quad (5.53)$$

which is indeed the scaling guessed from assuming that the second return scale as the first returns of a random walker.

Chapter 6

Stochastic Dynamics of Discrete Systems

“The human brain is capable of a full range of behaviors and predisposed to none.”

Stephen Jay Gould

6.1 Introduction

In this chapter we attempt to understand aspects of our social/biological world while assuming that it is influenced by the randomness associated to a finite number of discrete entities. We will present models that incorporate discreteness in continuous equations, and discuss them in relation to their more detailed/realistic agent based versions.

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.

We will introduce agent-based models (ABMs) using a few simple examples, which have analogs in continuous differential equations. ABMs were originally introduced by Von Neumann [86], in an effort to deal with system properties of many relatively simple agents that repeatedly use specified rules of mutual engagements [86, 87, 88, 89, 90]. 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. 6.1.

ABMs are also suited for addressing self-organization and emergent phenomena, as one may e.g. view directed percolation or SOC as an agent based model. ABMs have also been used to study properties of living and, in par-



Figure 6.1: **Agent-based models can involve intrinsically different agents.** Each agent is defined in terms of its particular caricature of a strategy. Agents may also have identical basic characteristics, but then develop different properties as a function of "life experiences" during many updates of mutual interactions. In any case agent based models opens for a kind of "quenched disorder", that includes long term memory effects.

ticular, social systems, including segregation [87], traffic jams, evacuation behaviour [89], social insect organization [91], stock market dynamics, as well as dynamical pattern formation — with ABMs taking the form of a cellular automaton [86, 92, 93, 88].

6.2 Direct Stochastic simulation

Imagine one want to extend the simple equation

$$\frac{d\langle x \rangle}{dt} = \langle x \rangle$$

to include fluctuations that arrices because any changes in x only occur randomly and in step of +1. That means that one for small Δt can understand $x\Delta t$ as a probability that one add 1 unit in the time-step. With this interpretation one can rephrase dynamics in terms of stochastic changes:

- $x \rightarrow x + 1$ with probability $x\Delta t$

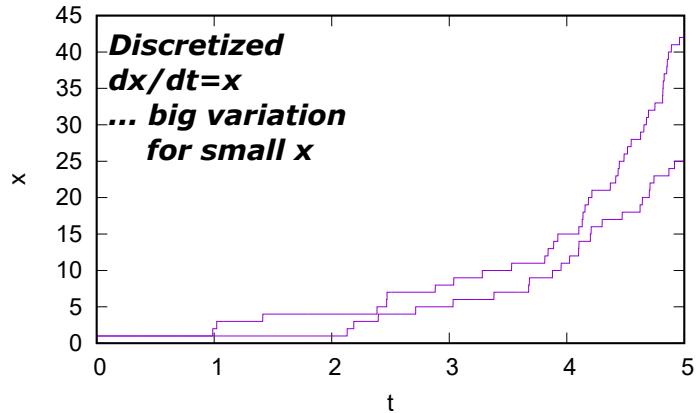


Figure 6.2: **Stochastic exponential growth.** Each of x agents may replicate with rate 1 independently of each other.

The dynamics of this over-critical branching process is seen in Fig. 6.2. One sees that there is big initial variation in the exponential growth.

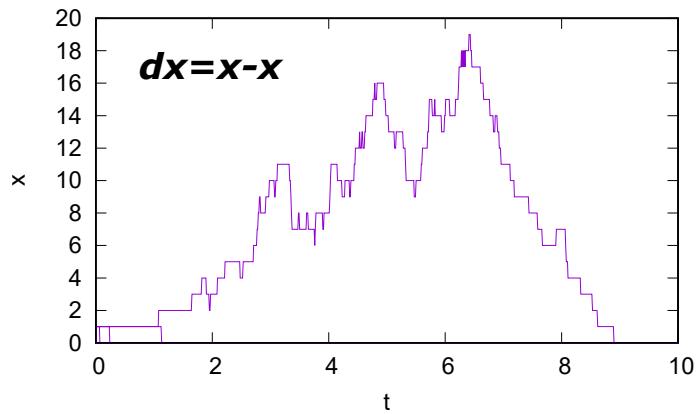


Figure 6.3: **Stochastic exponential growth.** Each of x agents may replicate with rate duplicate or die with equal probability, independently of each other. The figure show 4 trajectories, where 3 die very early.

Now let us consider the slightly more complicated dynamical system defined as

$$x \rightarrow x + 1 \quad \text{with rate } x \quad (6.1)$$

$$x \rightarrow x - 1 \quad \text{with rate } x \quad (6.2)$$

with x bounded to non negative numbers. This corresponds to an average equation

$$\frac{d\langle x \rangle}{dt} = \langle x \rangle - \langle x \rangle$$

where the two terms should be understood as a signal that there are two independent processes, one exponential growth in steps of 1 and another exponential decay in steps of 1. A stochastic equation for the above could also be written as

$$dx = \sqrt{2x} \cdot dR \quad \text{for } x \geq 0 \quad (6.3)$$

with $\langle dR^2 \rangle = dt$, meaning that R is a Brownian random walker with $\langle (R(t) - R(t'))^2 \rangle = t - t'$. More formally $dx = x(t + dt) - x(t) = \sqrt{2x(t)} \cdot (R(t + dt) - R(t))$. The factor 2 takes into account contribution to the variance by two independent processes each with variance x . This process and the Ito calculus is discussed more formally in the later course by Namiko Mitarai. Here we will just concentrate on simulating the basis process corresponding of two independent processes.

Simulation is started with one unit, $x = 1$ and terminates when $x = 0$. For small Δt one can rephrase dynamics in terms of stochastic changes:

- $x \rightarrow x + 1$ with probability $x\Delta t$
- $x \rightarrow x - 1$ with probability $x\Delta t$

The dynamics of this critical branching process is seen in Fig. 6.3.

Finally let us consider another more standard reaction like equation for mean development of a variable x

$$\frac{d\langle x \rangle}{dt} = Q - \frac{\langle x \rangle}{\tau}$$

with $Q > 0$. The stochastic dynamics is defined by the additional information that each term on the right hand side reflect one type of process and that any changes in x only occur in step of +1 or -1. That means that one for small Δt can rephrase

$$x(t + \Delta t) = x(t) + Q \cdot \Delta t - \frac{x}{\tau} \cdot \Delta t$$

in terms of two types of changes:

- $x \rightarrow x + 1$ with probability $Q\Delta t$
- $x \rightarrow x - 1$ with probability $x\Delta t/\tau$

and then just iterate the simulation forwards. Typical trajectory are shown in Fig. 6.4.

Importantly, in any of the above processes Δt should be so small that one can translate any of the above steps to probabilities, e.g. $Q\Delta t \ll 1$ and $x\Delta t/\tau \ll 1$. This can put some severe constraint on Δt , and one may often do faster simulation by focusing on the change only: Instead of simulating many time-steps where nothing happens one instead jump from change to change in the system with variable time steps. This is what we will describe in the following.

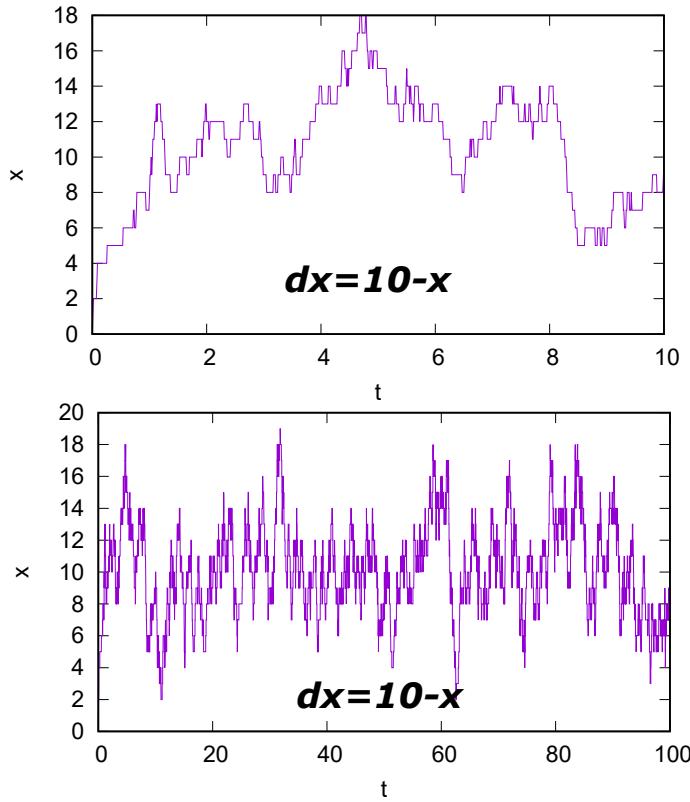


Figure 6.4: **Stochastic simulation.** x may be produced with rate 10 or decay with rate x . The two figures shows different time-windows.

6.2.1 Event based Stochastic Simulation

Gillespie (1977) developed a stochastic approach to chemical 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. Then the probability for the next event of any particular type is decaying exponentially with some rate set by this particular event type. Thus the chance that an event type i happens after time Δt is

$$P_{\text{next-event}}(\text{after } \Delta t) = \exp(-r_i \Delta t) \quad (6.4)$$

where r_i is the rate of event type i . This cumulative probability function varies between 1 and 0 as Δt varies between 0 and ∞ .

To select when an event of type i actually occurs, one selects a uniformly distributed random number $\text{ran} \in [0, 1]$ representing the probability that the event happens at a particular time increment Δt_i . This time interval Δt_i is found by solving $\exp(-r_i \Delta t_i) = \text{ran}_i$. Naturally, the Δt_i selected in this way will be exponentially distributed, reflecting our basic assumption on Poisson

random processes. Accordingly, if the current time is t , then the next event should be assigned to occur at $t + \Delta t_i$ with

$$\Delta t_i = -\frac{1}{r_i} \cdot \ln(ran_i). \quad (6.5)$$

Here, $1/r$ is the average time to the next event.

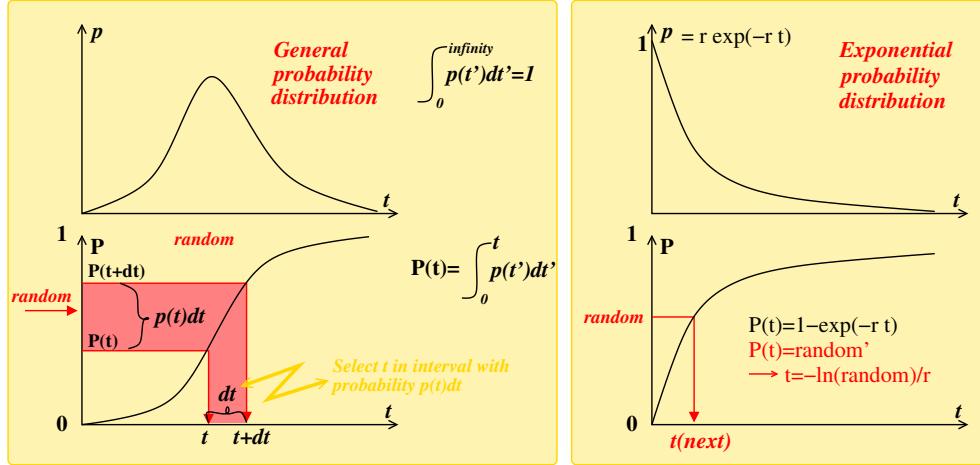


Figure 6.5: Event time selected from cumulative distribution:

Left panels: 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 $P(t) = \int_t^\infty p(t')dt' = \text{random}$. Thereby times between $[t, t+dt]$ are selected with probability $p(t)dt$. I.e. the part of $p(t)$ that has high values will be selected at higher likelihood because more random values corresponds to times where $P(t)$ has a large slope. Right panels: If $p(t)$ is an exponential, then $P(t)$ is also exponential and the selection of t is analytic.

In many cases there are several competing event types that can take place. Then one needs to make a list of times, one for each of the event types. And one then select the first of these times, do the update, and calculate an entire new list of event times. (one cannot use the old as the system most likely changed and rates have to be adjusted to latest event).

Consider, for example, events types where a variable increases, and a competing event type where the same variable decreases:

- $x \rightarrow x + 1$ with rate r_1 (that might be a function of x)
- $x \rightarrow x - 1$ with rate r_2 (that also might be a function of x).

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 happen to the stochastic development?

Now, as an example, consider the simple equation from earlier,

$$\frac{d\langle x \rangle}{dt} = Q - \frac{\langle x \rangle}{\tau}$$

where the creations and destruction process both takes place in units of 1. This 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 $x \rightarrow x + 1$, respectively $x \rightarrow x - 1$:

Q would then be the rate for the change $x \rightarrow x + 1$.

x/τ would be the rate for change of $x \rightarrow x - 1$. The stochastic update rule would be recursively performing the following after each event:

- Select time step for increase in x , $\Delta t_Q = -\ln(ran_1)/Q$ by choosing a random number $ran_1 \in [0, 1]$.
- Select time step for decrease in x , $\Delta t_x = -\ln(ran_2)\tau/x$ by choosing a random number $ran_2 \in [0, 1]$.
- If $\Delta t_Q < \Delta t_x$ then update $x = x + 1$ and set time counter $t = t + \Delta t_Q$.
- If $\Delta t_Q > \Delta t_x$ then update $x = x - 1$ and set time counter $t = t + \Delta t_x$.

The above changes can also have different step sizes 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/biology/sociology of the problem. If each term is changed in steps of Δ_i the simulation of a dynamics with $i = 1, 2, \dots$ 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.

The based simulations above entirely rely on the assumption of Poisson random events. In any other case there will be some sort of aging in the dynamic: The chance that an event happens next will depend on the time that has passed since last event. Such history dependence is for example seen for power law distributed event times, and in that case the whole concept of history independent update of times after each event would break down. Also aging distributions that are narrower than exponential are not allowed, although may go around this problem by representing history by intermediate states. Thus one sometimes can fit a narrower distribution by a sum, i.e. $t = t_1 + t_2$ where t_1 and t_2 are exponentially distributed.

Questions:

6.1) 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.

6.2) Use the Gillespie algorithm to simulate the stochastic equivalent of a variable x with a mean governed by $d\langle x \rangle / dt = 12 - \langle x \rangle$ 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 time-series of x in the two cases.

6.3 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

Competing states are part of society, where opinions spread through social contacts [94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107]. The most heavily studied systems of this type are the very simple “voter models” [97, 98],

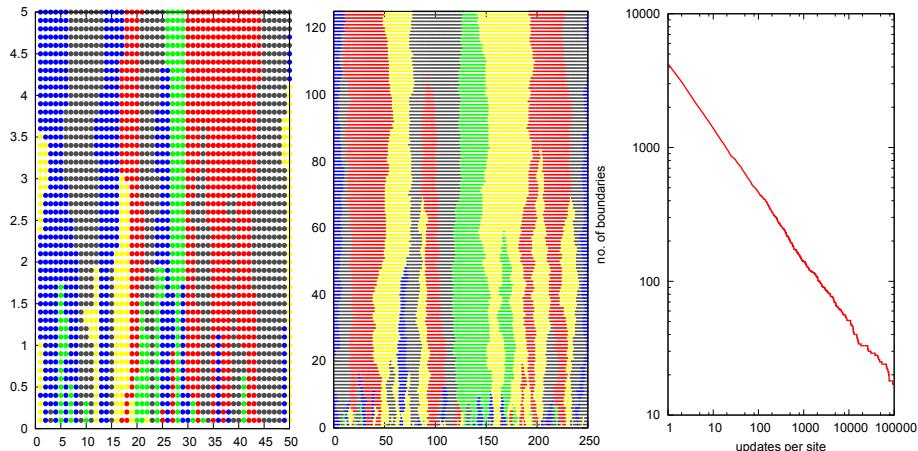


Figure 6.6: Dynamics of the Voter model. At each step one selects one site and then sets its state to the same as its neighbor. In the simulation above we assume five different states. The system always coarsens and the dynamics of boundaries perform random walks. Coarsening happens when two random walkers 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 $\sqrt{\text{time}}$ and the number of boundaries coarsens as $1/\sqrt{\text{time}}$, see rightmost panel for simulations of an $L=100.000$ system.

where agents take one of several opinions, and update these by repeatedly setting one agent to be equal to the state of a random of its neighbours. Thus one at time zero start with some opinions distributed among agents on a lattice. Each agent have one opinion.

- At each update step one select a random agent i . One then select one of its nearest neighbors and sets its opinion equal to that of agent i

Fig. 6.6 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. This is illustrated in the right panel of Fig. 6.6, where number of domains along a 1-d line decreases as $1/\sqrt{\text{time}}$.

Noticeable, 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 related models include the Axelrod model [94], where opinions are multidimensional (and agents only communicate to the extent that they at least share some opinions with each other). Also the Axelrod model coarsens, but versions of it often develops into a locked state consisting of non-communicating clusters perhaps resembling the polarized non-communicating sub-populations in American politics.

6.3.1 Voter model with cooperativity

We now consider a Voter like model with an intermediate state, and simulate this on a 2-d lattice. We will see a remarkable similarity to the Ising model, in spite of the fact that the model is not an equilibrium model. In the model each lattice point can take one of 3 values: -1, 0 or 1. These values influence each other. At each update step one

- Chose a site (x,y) . If the state s of this site is either +1 or -1 one randomly chose one of its 4 neighbor sites. The state of this neighbor site is adjusted towards s . Thus if $s = +1$ the neighbor change from 0 to 1 or from -1 to 0. Thus if $s = -1$ the neighbor change from 0 to -1 or from 1 to 0. If the two sites have same state value, or if $s = 0$ then nothing happens
- with probability β one select a random site and assign it a new random state, -1, 0 or 1 with equal probability.

This model reflects a cooperative voter model, where sites only change the full way from -1 to 1 by two subsequent exposures to a +1 site. The noise term β represent random changes, and is partly analogous to the temperature in the Ising model.

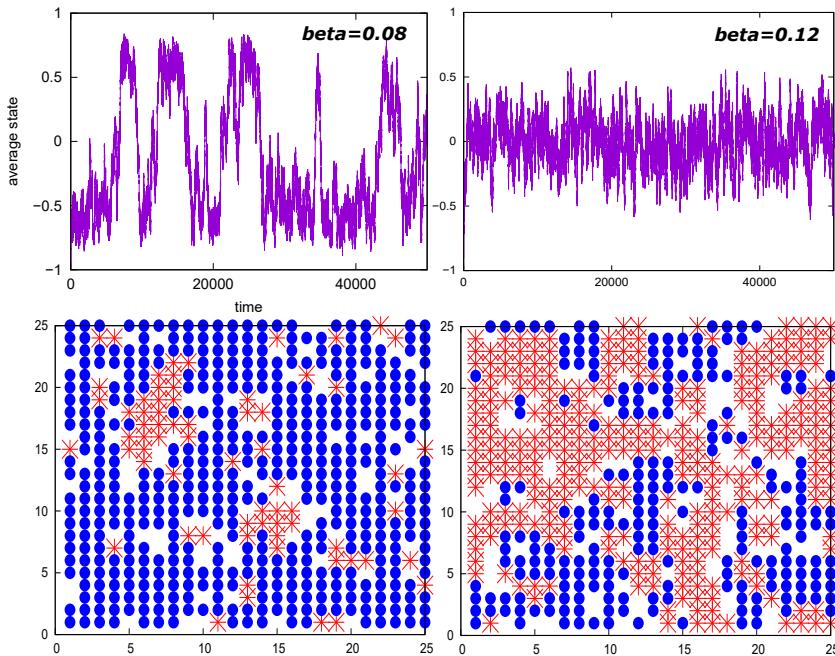


Figure 6.7: Social dynamics/non equilibrium variant of Ising model. The figure show simulations of the 2-dimensional version of the social recruitment model, highlighting its similarity with the Ising model.

The above model is not an equilibrium model. When a site have changed state, the agent that caused the state change is forgotten, and the site can move back without any concern about the previously active neighbor.

Fig. 6.7 show the typical dynamics for a lattice of linear dimension $L = 25$. Similarly to the Ising model, then for low β one see “bistable” behaviour, while larger β give a messy state with an average value of s that is close to zero.

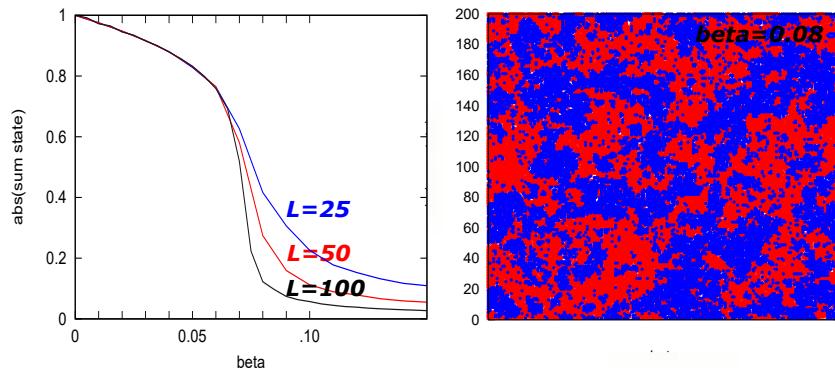


Figure 6.8: **Critical point of social dynamics model.** Left figure consider “magnetization” as function of β at $L = 25$ and $L = 50$. Right panel show $L=200$ system slightly above the critical value of β . At even higher β the correlation length gradually decreases.

Fig. 6.8 elaborate on the analogy with the Ising model by calculating the net “magnetization” as $Abs(\sum_{x,y} s_{x,y})/L^2$. One see a transition between order and disorder at around $\beta = 0.08$. Right panel show the typical state of the system at $\beta = 0.08$.

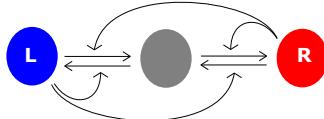
We now want to analyze the above model in terms of its mean field/random neighbor variant, see Fig top panel in Fig. 6.9. We first reformulate this 3 state model in terms of a simpler two-state model. These two states are denoted by L and R , and the basic idea is that two R ’s are needed to convert one $L \rightarrow R$. The two-state model is described in Fig. 6.9.

Mini tutorial: Can you mention any example of meta-stable systems in physics/your surroundings?

The agent based version of this model consider N individuals that each are in either L -state or R -state. A step of the model consist of selecting a direct conversion step with rate β or a recruitment attempt with rate 1.

- R** Next recruitment will occur at time increment $\Delta t_R = -\log(ran_1)$ with uniform $ran_1 \in [0, 1]$.
- D** Next direct conversion event will occur at time increment $\Delta t_D = -\frac{\log(ran_2)}{\beta}$ with uniform $ran_2 \in [0, 1]$.
- 3** If $\Delta t_R < \Delta t_D$ one randomly select two different agents. If these two agents are both in R -state, or both in L -state one select a third agent. This third agent is now set to the same state as the two agents.

3 state similar to 2-state+cooperativity



To go from minority state to majority
you need 2 unlikely
events in both cases

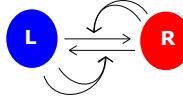


Figure 6.9: Models for bi-stability. The three-state model (upper panel) does not require explicit cooperativity [108]. The schematic of the two-state model (lower panel) indicates, that two representatives of R are needed to convert one L into an R [109, 110]. This process represents direct cooperativity. In any case the two models provides bistability where systems of many agents coherently “decide” a common state.

- 4 If $\Delta t_D < \Delta t_R$ one select one random agent and change the state of this agent to the opposing state.
- 5 Update time according to the selected event.

The model have one parameter β that characterize the extent random direct conversions occur independently on the overall system properties in general. The model assumes random contact, and thus ignores spatial correlations of the agent along for example the 2-dimensional grid explored earlier.

In terms of equations, the agent based model can be mathematically described through 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 $1 - r$. For each time step the r may decrease due to the events:

- Attempt a recruitment: There is a probability r to pick an R -agent, and a probability $(1 - r)^2$ that two different random sites are opposing this state. The probability that $r \rightarrow r - 1/N$ due to recruitment is $r(1 - r)^2$.
- With probability β one do a direct conversion attempt by selecting a random agent. This random agent has probability r to be in R -state.. Thus the probability for change $r \rightarrow r - 1/N$ is $\beta \cdot r$ due to direct transitions.

Similarly, r may increase due to symmetric probabilities for decrease in $1 - r$.

Adding up all recruitment and direct transitions per unit time, the average change r is governed by two recruitment processes and two noise processes:

$$\frac{dr}{dt} = (r^2 \cdot (1 - r) - r \cdot (1 - r)^2) - \beta \cdot r + \beta \cdot (1 - r) \quad (6.6)$$

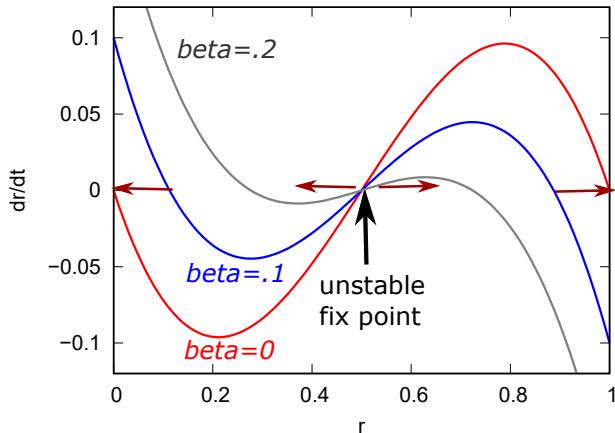


Figure 6.10: **Unstable fix point in cooperative voter model.** Figure show dr/dt for 2-state opinion dynamics model at different levels of random conversions. For $\beta > 1/4$ there is only one stable fixed point, while lower β leads to 3 stable fixed point.

This can be rewritten as

$$\begin{aligned} \frac{dr}{dt} &= r(1-r)(2r-1) + \beta \cdot (1-2r) \\ &= (r(1-r) - \beta)(2r-1) \end{aligned} \quad (6.7)$$

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 $r = 1/2$. Fig. 6.11 show the behaviour of the corresponding steady state order parameter, $\Rightarrow r = 0.5 \pm \sqrt{0.25 - \beta}$. Qualitatively this is similar to the behaviour of magnetization in the mean field Ising model.

Effect of Cooperativity: In case one only required one methylated site to make the recruitment, eq. 6.7 would be replaced by

$$\begin{aligned} \frac{dr}{dt} &= (r(1-r) - r(1-r)) + \beta(1-2r) + \xi \\ &= \beta \cdot (1-2r) + \xi, \end{aligned}$$

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.

1-dimensional version of the agent based model does not give bistability: In this case 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. 6.12. 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.

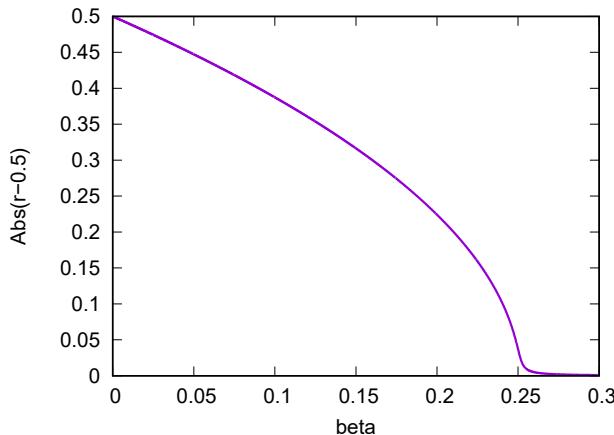


Figure 6.11: **Stable fix points.** Cooperative 2-state voter model, i.e. solution to $dr/dt = 0 \Rightarrow r \cdot (1 - r) + \beta = 0 \Rightarrow (r - 1/2)^2 = -\beta + 1/4 \Rightarrow r = 1/2 \pm \sqrt{1/4 - \beta}$.

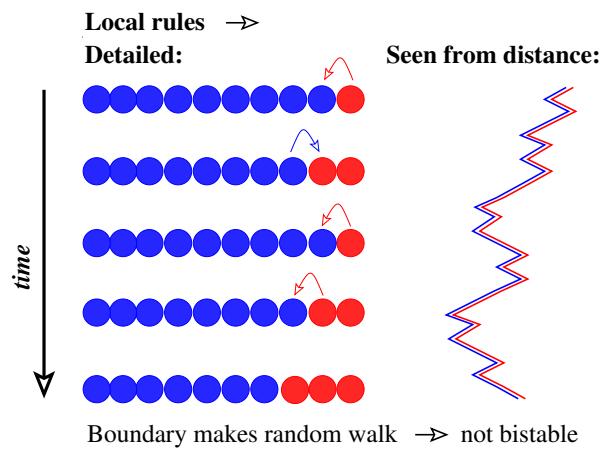


Figure 6.12: **Considerations for bi-stability in one dimension.** Local model in 1-d cannot give bi-stability, even when two neighbors in the same state are needed to convert a site. To see this, assume again a balanced model, where R and L are equally "strong". First, the system will coarsen into several 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 space for cooperativity.

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.

6.3.2 Information spreading on social scales

We now go to a different aspect of information spreading than simple opinions, namely spreading of news. This has much the same dynamics as spreading of diseases, as a disease in some way represent news on some biological scale, that ages as the host population gets immunized.

Information gathering allow 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 [111], 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 [106], 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 6.13 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 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 $v_{\text{word}} = 1 \text{ km/year}$ ($0.5\text{-}2 \text{ km/year}$). 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 stochastic at the center with a frequency f_{word} from the centre. Importantly, as words spread, they always retain their original birth time, assigned at their origination at the cultural center.

- 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.

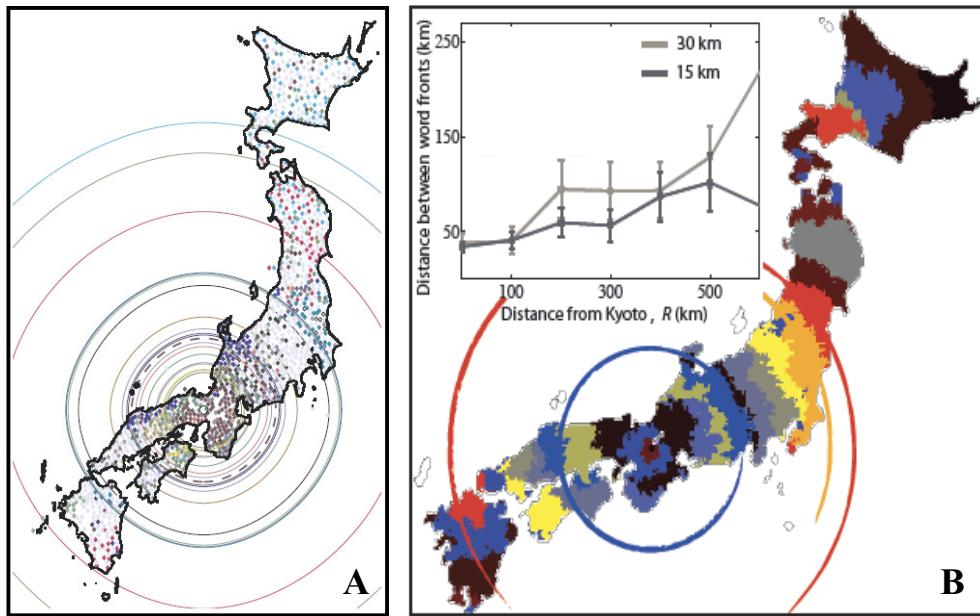


Figure 6.13: Swear word dynamics in Japan. The left panel shows the distribution of swear words as measured in the 1980's. The geographical distribution of concentric circles around Kyoto is the result of 600 years of history. The right panel shows a snapshot of a simulation of the spatial dynamics of 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 Panel B, insert this is quantified by the width of the respective word regions: As distance from Tokyo increases, the region of surviving words tends to increase. Figure reproduced from ref. [106].

- 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 older 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.

The used updating is called random stochastic update, and is easier to implement than synchronous updates of all sites at the same time. When one have performed a number of updates equal to number of sites in the system, they in average correspond to each other (therefore the division by N in first step, making f the parameter that quantifies number of new words in units of transfers per site). In Fig. 6.13 the 2-d lattice is constrained within the

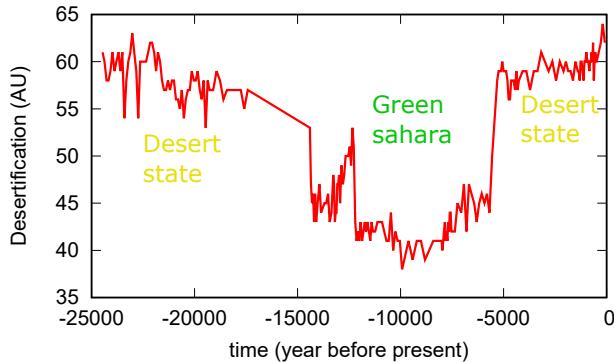


Figure 6.14: **Bistable vegetation model.** Switch to and from desert stage in Western Sahara, measured by dust in sediments off the coast. 6000 years ago the “green state” terminated.

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 6.13 shows the case where each “agent” represents a lattice “point” with 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. 6.13. 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 [112, 113].

Questions:

- 6.3)** Consider the 2-state recruitment model with cooperativity, system size $N=25$ and $\beta = 0.1$. Simulate the system with the Gillespie algorithm as outlined in text. Plot average order $\sum_{i,j} s_i s_j / N$ as function of β and compare with Ising model.
Qlesson: Quite similar.

- 6.4)** Implement a Gillespie simulation of the 2-state recruitment model as formulated in terms of the different processes in eq. 6.6. 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 the collective states increases.

Qlesson: Should correspond to the above agent based model with $N = 25$.

- 6.5)** Consider a 10-state agent based version of the cooperative model. I.e. take 100 agents, each can be in one of 10 states. At each timestep take two agents, and if same state then a third random agent is set to this state. Further, at each time-

step one with probability $\beta < 1$ select a random agent and set the state of this to one of the 10 possible states randomly. Simulate as function of β , and observe 1) multi-stability, and 2) possibility for a metastable state where all states are equally present.

[?] Qlesson: This is analogous to the Potts model in the same way as the normal 2 state model is analogous to the Ising model. See also discussion in [110]

6.6) Consider the metastable states of Sahara in Fig. 6.14, and develop a model the exhibit such bi-stability based on the assumption of 3 states of a local patch of land: Forrest (F), grass (G) or dessert (D) where the two extreme states favour conversion towards itself on neighbor patches of land. Simulate a 50 times 50 grid with recruitment dynamics as in the 3 state model outlined in top panel of Fig. 6.9. Adjust random conversion level β so you can see bi-stability. **Qlesson: Bi-stability is quite natural outcome of models where dominance is self amplifying. Tipping points are interesting. Notice analogy to Ising model.**

6.7) Simulate and visualize spreading of words 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.**

6.8) Lanchesters rules of military attrition.

Linear law: one soldier against one soldier at a time, implies equal about loss on each side independent of sizes.

Square law: Apply to armies with guns. Damage on an given soldier at each time-step now scale with number of soldiers of opposing army.

Simulate these two rules by stochastic attrition of individual soldiers, starting with 200 soldiers of army A and 100 soldiers of army B.

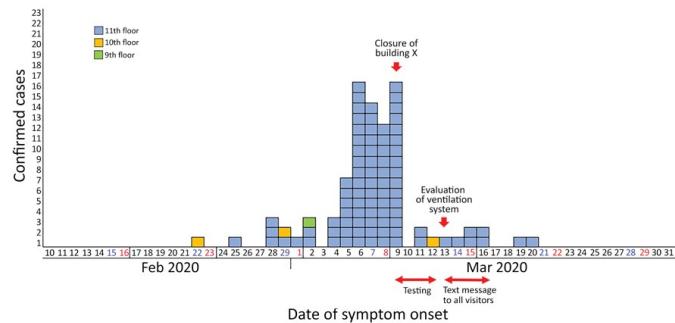


Figure 6.15: **Covid-19 epidemics in a Korean Call center** Notice the periodic feature of the outbreak, indicating that infections typically happens 5 days after previous infection (generation time is 5 days.)

Logic: We follow a sick from infection time and simulate how many days he infect. This number of days should be minimized by inserting tests and removal of positives.. Persons are infectious before test, but quarantined afetr first positive test.

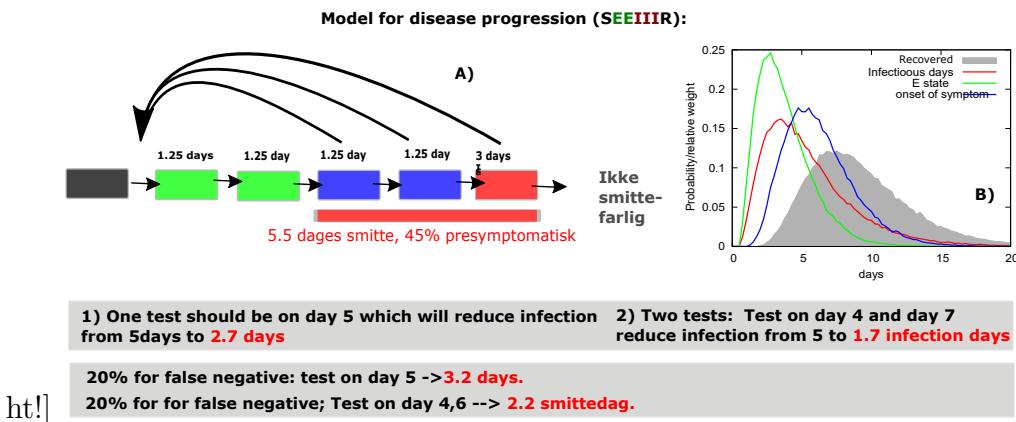


Figure 6.16: **Covid-19 disease progression** Agent based simulation of optimal test days given distribution of infectious period. Many agents are simulated and one calculate the impact of test days given that people are completely quarantined after test. These simulations was part of the Danish authorities design of testing strategies, given time of exposure to infected person.

6.4 Epidemics

6.4.1 SIR model

We now discuss spreading of an infectious disease, as illustrated on micro-scale in Fig. 6.15. It is a process involving individuals that meet and infect each other. However this fundamentally agent based process is often modeled by mean field approaches, where one forget all about the constraints of repeated contact networks.

Importantly Fig. 6.15 illustrate the progress of the disease spreading in generations, each lasting about 5 days and each characterized by about a factor 8 more persons. The 5 days discrete generation time comes about as a consequence of a delay between becoming infected and being able to infect. This so called exposed state takes about 2.5 days for Covid-19, while the subsequent infectious period takes about 5 days. These periods ar not exponentially distributed, by in fact substantially narrower.

Fig. 6.16 show expected progression of one Covid-19 infection, highlighting that it could be modeled as a sequence of states. The exposed state takes 2.5 days and is modelled as two exponential periods of 1.25 days duration (shown in green). The subsequent infectious period of 5.5 days is modeled in terms of the sum of 3 smaller exponentially distributed periods. By compartmentalize the disease into stages, one take into account that a person cannot infect just after he himself is infected. There is a delay. Each stage is modelled as an

exponential process, with a characteristic time τ :

$$P_1(t) = \frac{1}{\tau} \exp(-t/\tau)$$

Two subsequent processes then add up to a distribution

$$P_2(t) = \int^t P_1(t') \cdot P_1(t-t') dt' = \frac{t}{\tau^2} \exp(-t/\tau)$$

and 3 subsequent processes:

$$P_3(t) = \int^t P_2(t') \cdot P_1(t-t') dt' \propto \frac{t^2}{\tau^3} \exp(-t/\tau)$$

etc. for more progression with more steps (these distributions correspond to Gamma distributions and is used extensive in realistic disease progression models). In Fig. 6.16 we show how such progression can be converted to test strategies for reduction of an on-going Covid-19 epidemic. That is one can simply simulate a number of agent, infected on day 0, and then quantify whether a test on e.g. day 4 reduce time of infection. The rule is that the test only catch people in infectious period, and that persons are put in quarantine at time of positive test. Thus if test is too early, infected people are still in exposed state and do not test positive. If test is too late, many people will be finished with their infected state and test have no consequences. In between there is an optimum test day.

We now introduce the standard mean field approach often used in epidemic modeling. Here one simply forget space, and assumes that the interactions between types of agents simply are proportional to their respective densities:

$$\text{Frequency for an encounter between } A \text{ and } B \propto \rho_A \times \rho_B \quad (6.8)$$

where ρ_A and ρ_B are the densities of A , respectively B . In the mean field theory of the Ising model this corresponds to the assumption that any particular spin simply interact with the average state (density) of the other spins.

The simple approach of multiplying densities of course assumes that there is no substantial 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.

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 [114]. 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 [115, 116].

Epidemic models are similar to models for spreading information and ideas, i.e. our spreading of words in Japan. They just typically occur on a faster timescale, and with quite a time separation between subsequent epidemics. 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 transitions between S and I are described by the rate β per individual, while the rate of escape from the infected state is described by rate γ .

The SIR scenario can also be reproduced by a simple agent based model where each individual are in one of the 3 states. Most directly, the simulation is done in discrete time-steps of duration dt . At each timestep dt one goes through the entire list of agents. For each infected agent i one selects another agent j . If this other agent j is in S -state, it is infected (converted to I -state) with probability βdt . Furthermore, if the agent i is in the I -state it is converted to the R -state with probability γdt . When dt is so small that $\beta dt \ll 1$ and $\gamma dt \ll 1$ this agent based simulation quite closely resembles a reasonable encounter dynamics between people. Further, it obviously ignore any prior correlations between people and can thus easily be simplified by simple differential equations.

The corresponding mass-action kinetics for a epidemic was suggested by Kermack, W. O. and McKendrick, A. G. in “A Contribution to the Mathematical Theory of Epidemics” Proc. Roy. Soc. Lond. A 115, 700-721, 1927):

$$\frac{dS}{dt} = -\beta \cdot S \cdot I \quad (6.9)$$

$$\frac{dI}{dt} = \beta \cdot S \cdot I - \gamma I \quad (6.10)$$

$$\frac{dR}{dt} = \gamma \cdot I \quad (6.11)$$

Again the parameter γ decides for how long individuals are infectious (time period of I -state $\sim 1/\gamma$) and the parameter β/γ subsequently determines how many individuals each infected person can infect. A central parameter is the so called R_0 -factor:

$$R_0 = \frac{\beta}{\gamma}, \quad (6.12)$$

which is central in thinking about how widespread the disease spreads before herd immunity sets in. R_0 is the number of infections that each infected individual cause at the beginning of an epidemic (when few individuals have been infected). R_0 is very large for measles ($\sim 10 \rightarrow 20$) and smallpox (~ 6), it is ~ 2 for Ebola, about 2 for pandemic influenza and about 1.4 for common influenza. For the original Covid-19 the R_0 was about 3, while its Δ variant had $R_0 \sim 5$.

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 6.14. 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

$$R_0 \cdot (1 - q) < 1 \rightarrow q > 1 - \frac{1}{R_0}. \quad (6.13)$$

Thus, for an R factor of 3 one needs to vaccinate more than 67% of the population. This critical threshold in fraction of cases $1 - 1/R_0$ also sets the point where $dI/dt = 0$, i.e. where dI/dt stops increasing. When this fraction of people have been infected there is simply not enough susceptible people left maintain growth in infections. This does not mean that the epidemic suddenly stops, but it does imply that the peak of the epidemic is passed.

As the epidemic unfolds, the amplification number tends to decrease because people get immunized (or die), leaving fewer susceptible individuals (The R in the SIR refers to removed from the susceptible population). Dividing the first with the last equation one obtain:

$$\begin{aligned} \frac{dS}{dR} &= -(\beta/\gamma) \cdot S = -R_0 \cdot S \\ \Rightarrow \frac{d\ln(S)}{dR} &= -R_0 \Rightarrow \ln(S(\infty)) - \ln(S(0)) \\ &= -R_0 \cdot (R(\infty) - R(0)) = -R_0 \cdot (1 - S(\infty)) \\ \Rightarrow S(\infty) &= e^{-R_0(1-S(\infty))} \end{aligned}$$

where we use that $S + I + R = 1$, $S(0) = 1$, $R(0) = 0$ and that $R(\infty) = 1 - S(\infty)$ since there are no infected individuals after the epidemics have died out. Thus, if an epidemic with $R_0 \gg 1$ 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 herd immunity that in principle could stop the epidemic when $S = 1/R_0$. Thus with proper mitigation at the time where infections peak one could stop the epidemics leaving a fraction of $1/R_0$ uninfected, while a epidemic that runs its full course only leaves $\exp(-R_0)$ uninfected. For classical Covid-19 with $R_0 = 3$ the difference is large, from 0.33 of population untouched to a fraction of only 0.05.

Mini Tutorial: How would you stop the epidemic already at the herd immunity level.

The SIR model also predict a certain growth rate of the initial stages of the epidemic:

$$\frac{dI}{dt} = (\beta S - \gamma) \cdot I = (\beta - \gamma) \cdot I = \lambda \cdot I \quad (6.14)$$

with growth rate $\lambda = \beta - \gamma = \gamma \cdot (R_0 - 1)$ as long as few people are infected $S \sim 1$. This relation between rate constants in the SIR model and the early growth rate in an epidemics is of course modified when taken into account the one never are in I-state just after infection. Instead one have a progression through an intermediate E state, $S \rightarrow E \rightarrow I \rightarrow R$. This then reduces the growth rate because people spent a time τ in a passive state before they can contribute to new infections. For the original SARS-CoV-2 the length of the E-state was about 2.5 days, the length of the I state was about 5 days and the observed growth rate λ was between 10% per day in Japan to about 35% per day in Spain.

6.4.2 Agent based model with superspreading

Multiple studies of COVID-19 measured heterogeneity of infectivity among infected individuals, finding that 1% to 20% of infected people cause about 80% of new infections [117, 118, 119]. In the spirit of stochastic dynamics, such superspreading easily explain the larger variation in epidemic trajectories there is between localities in a given small country 6.17. This means Considerable evidence indicates that superspreaders are important in the spread of COVID-19, see [120] from which the following sub-chapter is copied.

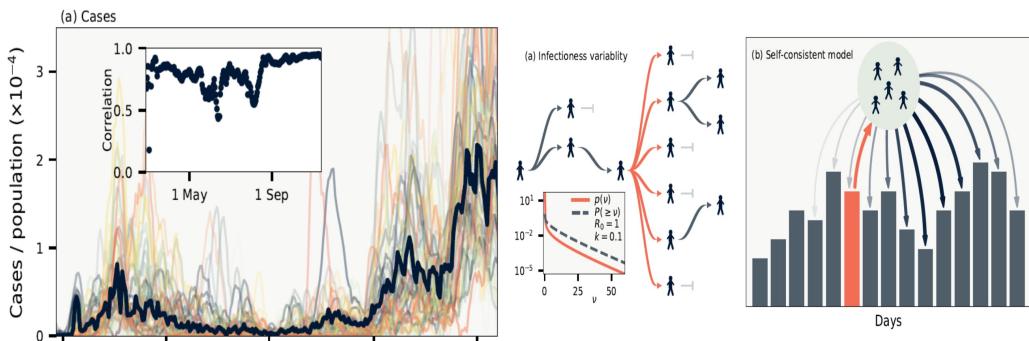


Figure 6.17: **Covid-19 in Danish municipalities** Covid cases in Denmark with epidemic trajectories in different municipalities. Notice individual outbreaks, probably mediated by superspreaders. The right panels models the observed pattern given infection times and degree of person to person variation in infectivity. The data was reproduced by 10% being responsible for 80% of infections. Figure from [119].

Superspreading is not a phenomenon which is particular to SARS-CoV-2, but has been observed in connection with several other pathogens, including coronaviruses such as SARS [121, 122] and MERS [123], as well as in diseases such as measles [124] and Ebola virus disease [125, 126]. Pandemic influenzas such as the 1918 Spanish flu, on the other hand, are believed to be far more “democratic” [127].

Given the observed evidence that superspreaders/superspreading events are important in COVID-19 transmission, models should not rely on a single pa-

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.

Agent based models are very well-suited to include the role of superspreaders. That is they naturally can incorporate the individual specific ability to infect. 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. One difference between SIR models and agent based models are illustrated in Fig. 6.18, where we follow an individual visiting some social settings. For the well mixed SIR model they would effectively go to a new home every evening (timestep). In agent based model we can assign different people different social contacts. Thus we can incorporate both person specific social life and fixed differences in infection properties.

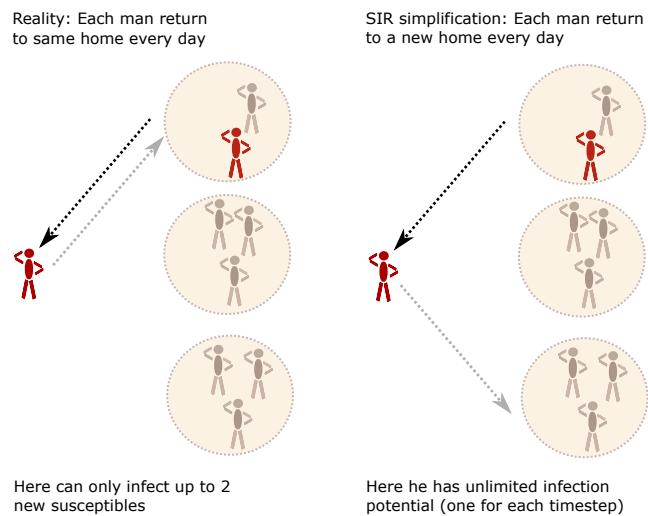


Figure 6.18: Mean field (SIR) versus its agent based version.

Mini Tutorial: Suggest ways that people could have large variation in number of people they infect? Which types of disease transmission would give a lot of such superspreading?

The heterogeneity of transmission is usually quantified using the Gamma distribution [124], see Fig. 6.19:

$$P_I(r) \propto r^{k-1} \cdot \exp(-kr/\mu) \quad (6.15)$$

with mean infectivity μ and dispersion parameter k . For $k = 1$ this is an exponential distribution. For $k > 1$ it becomes increasingly peaked around the mean values μ , with variance that decline with increasing k . Infection heterogeneity is normally parametrized by $k < 1$, a distribution that diverge for small infectivity r : Thus for $k \rightarrow 0$ then the Γ distribution becomes a

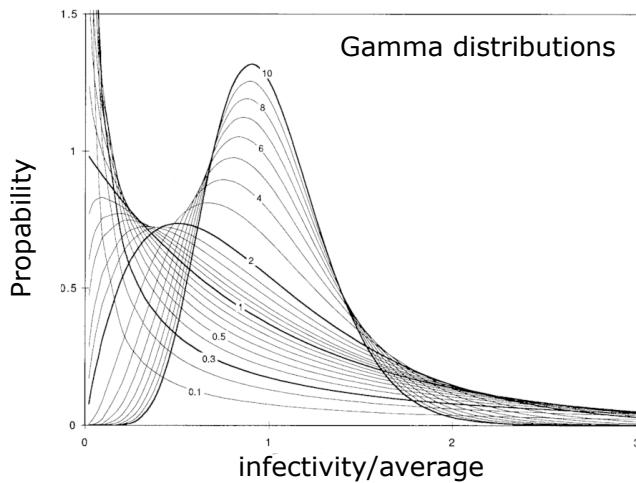


Figure 6.19: **Gamma distributions** Illustration of Gamma distribution as function of the dispersion parameter k . For $k < 1$ the distribution diverge at small s . For k larger than one it corresponds to the distribution one obtain by folding k identical exponential distributions.

power law with exponent -1 and cutoff μ/k , reflecting that most people will have $r \sim 0$ while most infections will be caused by individuals with large $r \cdot P_I(r) \sim r^k e^{-kr/\mu}$, i.e. by the few superspreaders with r between μ and μ/k .

Mini Tutorial: Sketch the Γ distribution for $k = 0.01$ and for $k = 2$

The *dispersion parameter* or *k value* determines the fraction of infectious individuals who account for the majority of infections (Fig. 6.20). Smaller k means greater heterogeneity – in fact, when k is small ($|k| \ll 1$) it approximates the fraction of infected individuals who give rise to 80% of infections. For COVID-19, which is believed to have a k value of perhaps 0.1 [118, 117, 128, 119], the most infectious 10% of individuals thus cause approximately 80% of infections.

The fundamental difference between a homogeneously spreading disease and a highly heterogeneous one is reflected in the infection networks they give rise to, as visualized in Fig. 6.20. When only a small fraction of individuals cause the bulk of infections, a reduction in social network connectivity amounts to decreasing the likelihood that a superspreader infects another superspreader and thus propagates the disease. Consequently, in a network characterized by superspreading (Fig. 6.20a), the outbreak can be stopped by cutting only a few select edges. Not so for the network in Fig. 6.20c.

We now present a model of superspreading phenomena which assumes that the driving force is a biological heterogeneity in infectiousness. We implement this as an agent-based model with contact networks, and are also able to capture much of the phenomenology in analytical formulae. In the model, N agents are placed as the nodes in a contact network. We investigate different

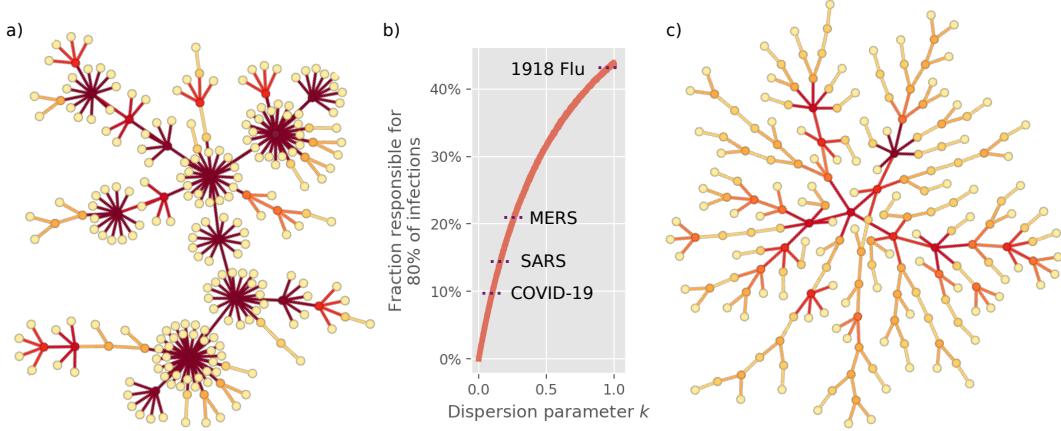


Figure 6.20: **The characteristics of superspreading.** **a)** Simulated infection network characterized by superspreading, with a dispersion parameter $k = 0.1$. Superspreaders appear as hubs, while most individuals are “dead ends” meaning that they do not transmit the disease. The epidemic mainly grows by spreading from one superspreader to the next. **b)** k provides a measure of superspreading, with lower k values corresponding to fraction of infected that cause 80% of infections (plot valid for large R_0 , for $R_0 \sim 1$ there is always many that does not cause any infections). **c)** Simulated infection network where all individuals have equal infectiousness, $k = \infty$).

types of network, but our base case is the Erdős-Renyi network, which is characterized by a Poisson degree distribution and an absence of clustering.

At initialization, the infectiousness of each individual is drawn from a Gamma distribution shown above. As such, it is an innate property of each individual. The possible states of each individual are **Susceptible**, **Exposed** (a latency time of 2.5 days), **Infected** (5 days) and **Recovered**. At each timestep, each individual randomly selects one of its contacts to interact with, meaning that only a subset of the network is active at any given time. While a link between an infectious and a susceptible individual is active, there is a constant probability of infection per unit of time, as determined by the individual infectiousness.

This basic setup also lends itself to analytic calculations. Consider a single infected person who has c contacts, who are all assumed susceptible. The infectiousness r of the individual is measured in number of persons – it can be thought of as the product of the infection rate and the duration of the infectious period. The r value is assigned each individual, drawn from a gamma distribution $P_I(r)$ with dispersion parameter k and mean μ . The distribution of the reproductive number R of an individual with a *known* infectiousness r and degree (i.e. connectivity) c is given by

$$P(R; r, c) = \binom{c}{R} (1 - e^{-r/c})^R (e^{-r/c})^{(c-R)} \quad (6.16)$$

where $\binom{c}{R} = c! / R!(c - R!)$. This equation express that a given person are in

contact with c people, of which R has to be infected and $c - R$ have to stay non infected. The exponential $e^{-r/c}$ counts the chance that a given contact is not infected in the allotted r attempts (corresponding to the individual infectivity number r).

The persons connectivity c represent the limits of time of contacts: If a person have c contacts he can only allocate a fraction $1/c$ to each of his contacts. Accordingly his total infection capacity r have to be divided among these contacts. When $r/c \gg 1$ the first term ($1 - e^{r/c} \sim 1$), and the person will be infected independent a change in r , or equivalent by a reduction in contact time t (infection number $r = \beta t$ can be modified by changing t but when t is high you are just going to infect anyway).

Mini Tutorial: Investigate the distribution of R in the limit of infinite c ($\sim (r/R)^R \exp(R - r)$)

Taking the variability in infectiousness into account, the overall distribution of R becomes:

$$P(R; c) = \int_{r=0}^{r=\infty} dr P_I(r) P(R; r, c). \quad (6.17)$$

In the limit of infinite connectivity, corresponding to a well-mixed population, this becomes a negative binomial distribution. That particular case has been studied in [124]. Given a contact network and a corresponding degree distribution $P_C(c)$ – for example a Poisson distribution in the case of an Erdős-Renyi network – the connectivities can be summed over to yield a distribution of individual reproductive numbers, $P(R) = \sum_c P_C(c) P(R; c)$.

As reflected in the equations above, the *actual* number of secondary infections depends not only on biological infectiousness. In Fig. 6.21a-b, we use this analytical framework to explore how the number of personal contacts affects the resultant distribution of infections. Without superspreading (6.21a), a reduction in the contact number has a very modest effect and the distributions overlap. When the heterogeneity is at a COVID-like level (6.21b), it is quite a different story. Here, a decrease in mean connectivity has a considerable effect, and mitigation suddenly looks feasible. Previously, another mitigation strategy which benefits from superspreading was suggested by [124], with the crucial difference that it relies on prior identification and targeting of superspreaders, in contrast to the broad reduction in mean connectivity explored here.

To quantify the sensitivity of the epidemic to social network size, we consider the basic reproductive number R_0 , meaning the average number of infections that each infected person causes in a situation where all contacts are still susceptible. In Fig. 6.21c, the R_0 is given as a function of the dispersion parameter k and the average contact number c . The epidemic is evidently much more sensitive to reductions in contact numbers when k is low, i.e. the transmission heterogeneity high.

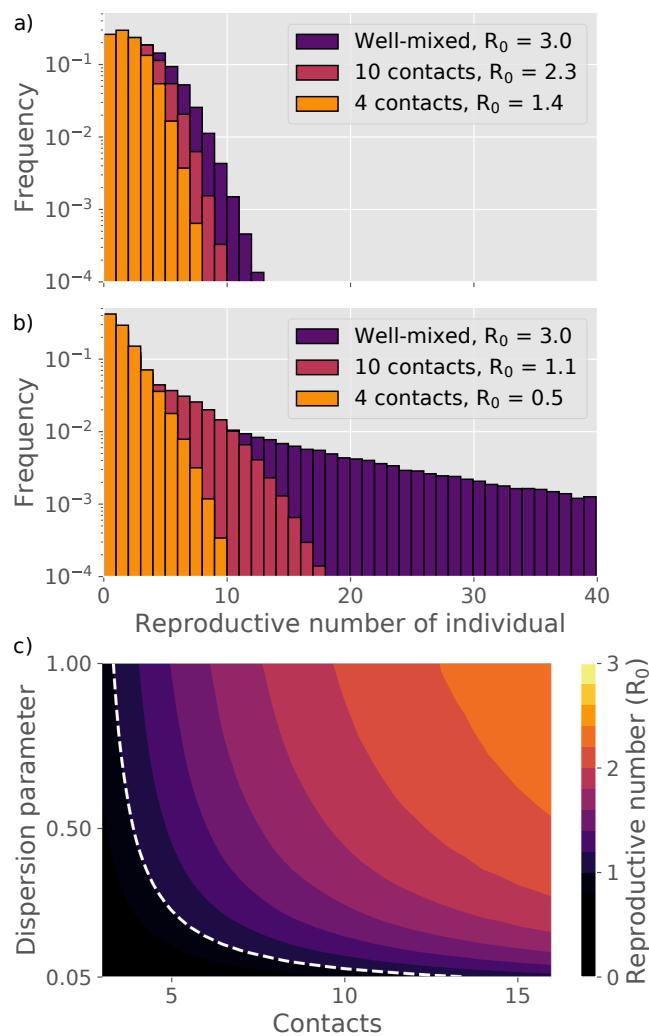


Figure 6.21: The reproductive number. Distributions of individual reproductive number R and value of R_0 for different dispersion parameters and number of social contacts during an infectious period. **a)** Distribution of R for a disease where all individuals have equal infectiousness. **b)** Distribution of R for a disease characterized by dispersion parameter $k = 0.1$. **c)** Basic reproductive number R_0 as a function of social connectivity and dispersion. The dashed line represents $R_0 = 1$. These calculations take into account the Poisson distributed contact number, and the fact that each infectious person will have one insusceptible person in their network (the individual from whom the infection originated), even when computing the basic reproductive number.

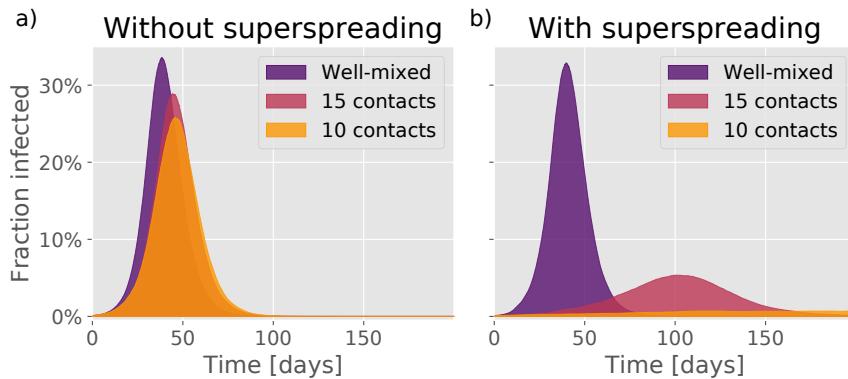


Figure 6.22: The epidemic trajectory of a heterogeneous disease is highly sensitive to mitigation. Epidemic trajectories as a function of the number of people that each person interacts with during an infectious period. **a)** Time evolution in the absence of any infection heterogeneity. **b)** Time evolution for a disease with dispersion parameter $k = 0.1$, roughly representative of COVID-19.

A mitigation in which the average number of contacts goes from being unrestricted, down to about 10, causes a reduction in R_0 which lowers both the *peak* and *total* number of persons infected during the course of the epidemic (the *attack rate*). The overall trajectory of a homogeneous disease is largely unaffected by social connectivity (Fig. 6.22a), whereas a heterogeneous epidemic is very sensitive (Fig. 6.22b). We find a particularly large sensitivity to a reduction of contact number from 15 down to 10 (Fig. 6.22b), indicating a critical threshold for disease spreading, in line with the threshold indicated by the dashed curve in Figure 6.21c.

Crucially, a reduction in contact *time* is not necessary when the disease is characterized by superspreading. What counts is rather a reduction in contact *diversity*, meaning the number of different persons with whom you come into contact during the time you are infectious [129]. This differs fundamentally from SIR models, where contact time and diversity are not differentiated between. In our treatment above, a reduction in the size of an individual's social circle is not accompanied by a reduction in contact time, since the same number of contact events is maintained, with each remaining person being contacted more often. Thus, a mildly infectious individual will not experience appreciable saturation by a reduction in contact diversity, whereas a superspreader will be highly limited by the resultant local saturation.

Mini Tutorial: Imagine that one have a superspreading disease where 10% infect 30 people and rest does not infect. Now chose between 2 types of mitigation: 1) give everybody face mask and reduce all infection rates by 50%. 2) Constrain people social network so they only meet at max 10 persons during the infectious period. What is most effective?

So far, our analysis has been based on the Erdős-Renyi network, which is largely devoid of clusters. This was chosen as a clean setting in which to probe

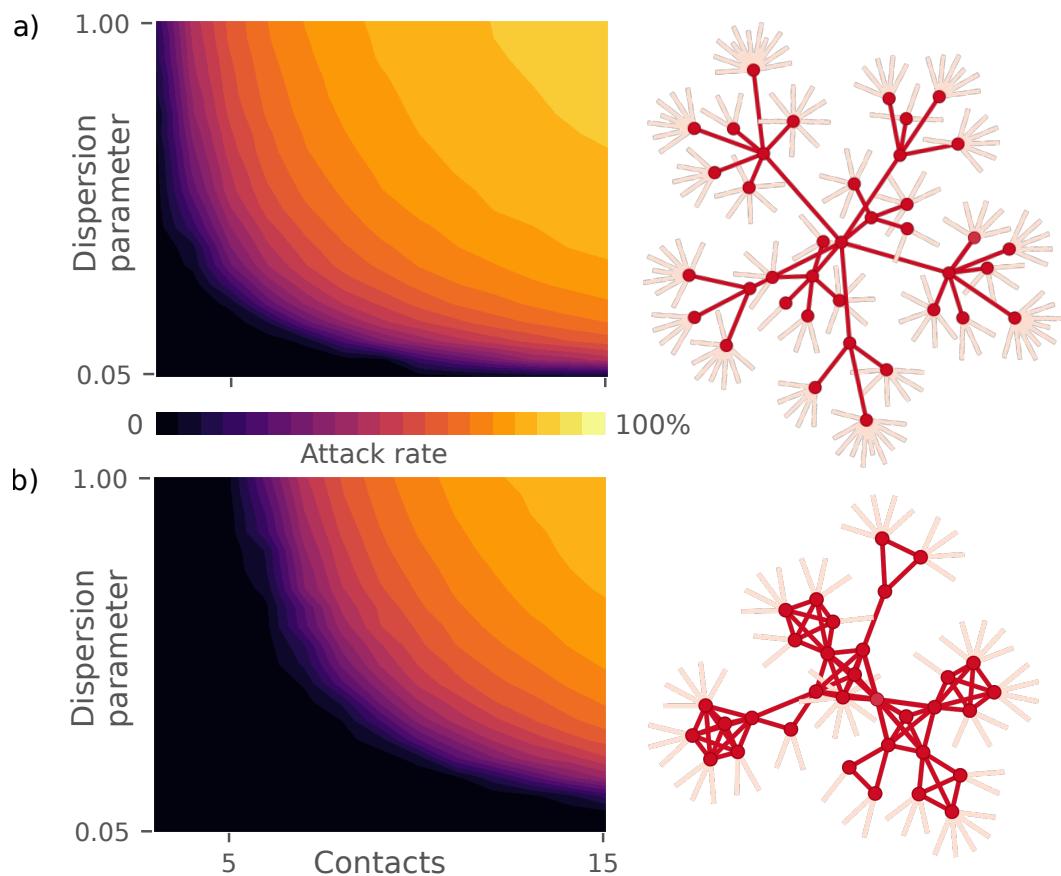


Figure 6.23: **Final attack rate** (total fraction of the population infected) as a function of network connectivity and transmission heterogeneity. **a)** investigate an Erdős-Renyi network, with the same degree distribution as in Fig. 6.21. **b)** explores a network where each person is assigned to two groups of people, leading to a highly clustered network. The black regions indicate conditions where the disease cannot spread in the population. On the right-hand side, small fragments of the networks in question are shown.

how social connectivity affects superspreading. However, any realistic social network will involve clusters of people who know each other [130, 131, 132] – after all, your colleagues know each other, as well as knowing you. It is thus natural to ask whether such *cliquishness* impacts superspreading. Fig. 6.24 compare a cluster-free network to one characterized by a high degree of clustering [133].

The attack rate of the disease is clearly lowered by clustering in general (Fig. 6.24), but the effect is especially significant when heterogeneity is high. The mechanism behind this is that of *local saturation*. If a superspreader infects a significant portion of his network, there is a risk that one of these individuals will turn out to be another superspreader. However, *if* there is clustering, a large part of this *second* superspreader's network will already have been exposed, and the second superspreader does comparatively little harm.

Its all a game of time and social constraints, heightened by the huge variation between people.

Regardless of the origin of superspreading, we emphasize the particular fragility of a disease in which a major part of infections are caused by the minority. If this is the case, the disease is vulnerable to mitigation by reducing the number of *different* people that an individual meets within an infectious period. The significance is clear; Everybody can still be socially active, but generally only with relatively few – on the order of ten persons. Importantly, our study further demonstrates that repeated contact with *interconnected* groups (such as at a workplace or in friend groups) is comparatively less damaging than repeated contacts to independent people.

Mini Tutorial: Above we assumed that people was heterogeneous due to variations in virus load. Suggest some other ways that people could have heterogeneous impact on disease progression during an epidemic.

Questions:

6.8) Simulate an agent based version of a SIR model with $\gamma = 0.2$ and $\beta = 0.6$ (like Covid-19), starting with $I=10$ infected among 10000 people (remaining 9990 in S-state). Plot epidemic trajectory and calculate the total number of people who have had the disease. **Qlesson:** Notice the noise, particularly in beginning of simulation. Eventually start several simulations with only one infected individual and notice large differences.

6.9) Simulate the SIR model with $\gamma = 0.2$ and $\beta = 0.6$ (like Covid-19), starting with $I = 0.0001$, $S = 1 - I$ and $R = 0$. Plot epidemic trajectory and calculate the total number of people who have had the disease. Now imagine an intervention that only occur when number of infected is maximal, i.d. when $dI/dt = 0$. At the time when this happens then reduce I to $I = 0.001$ and follow the disease after this reduction. How many people have in total had the disease after epidemic stops. **Qlesson:** Difference between this and previous simulation reflect the noise in an agent based

model. The difference in total attack rate is the difference between $\exp(-R_0)$ and $1/R_0$.

6.10) Obtain a general analytical expression for the mean and the spread of a Gamma distribution.

Qlesson: The coefficient of variation (spread/mean) is diverging as $k \rightarrow 0$

6.11) Construct an agent based model for a superspreadere driven epidemic where 10% of infected do all infections (i.e. not a gamma distribution but a simpler on-off model of infectiousness). 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 (superspreader infection rate of 3 infections per day, all other infected are does not transmit the disease). Consider a society with 10000 persons.

a) Assume first that persons contact each other randomly across the population, and simulate epidemic trajectory starting with 10 infected. Notice that often the epidemic stops in beginning, perhaps reflecting the slow start of the Covid-19 epidemic in Wuhan autumn 2019. Start then with 100 initially infected.

b) Assume instead that each person is embedded in an Erdos-Reynei network with average connectivity $c = 10$. You can just generate an ER network with 10000 nodes, where each node have a list of neighbors. Then propagate the disease keeping score on which nodes are in S, I or R state. Use same infection parameters as before. Calculate epidemic trajectory starting with 100 infected individuals. Simulate for different values of c and count final attack rate as function of c . **Qlesson:** Superspreadere like disease are vulnerable to mitigation efforts.

Lessons:

- Heterogeneity may be best treated by Agent based models. These for example allow models of a heterogeneous society, with different types of agents in fixed social networks.
- In case agents are similar and network constraints not important one may attempt the well-mixed approximation, where the probability for each reaction is simply the product of densities:

$$\text{Frequency for an encounter between } A \text{ and } B \propto \rho_A \times \rho_B,$$

where ρ_A and ρ_B are the densities of agent type A , respectively type B .

- Both well mixed models, and agent based models Models can be simulated in discrete updating steps that takes into account the randomness of the underlying discrete events. Th simulation can in both cases be done using event based simulations (Gillespie simulation) where each reaction i is assigned the time when it happens the next time:

$$t(\text{next}) = t(\text{now}) - \frac{\ln(\text{random}_i)}{\text{rate}_i},$$



Figure 6.24: **Heterogeneity in distribution matter:** During the peak of the Covid-19 epidemic maybe 100 million people were sick. Assuming that each person had about 10^{11} virus roaming his/hers body at any time, there was 10^{20} covid particles corresponding to about 10 kg of covid vira on the planet. One covid particle weight $10^{-15} g$ (and has a diameter of 100nm). The figure illustrate the different impact of distributing this bucket in a bath-tub, or homogeneously over the planet.

Here $rate_i$ is the rate of the corresponding process given the state of system. The $random_i$ is random number for each process, independently selected in the interval $[0, 1]$. 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.

Chapter 7

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 7.1: 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.

7.1 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.

The stock market refers to the collection of markets and exchanges where regular activities of buying, selling, and issuance of shares of publicly-held companies take place. (Investopedia)

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.

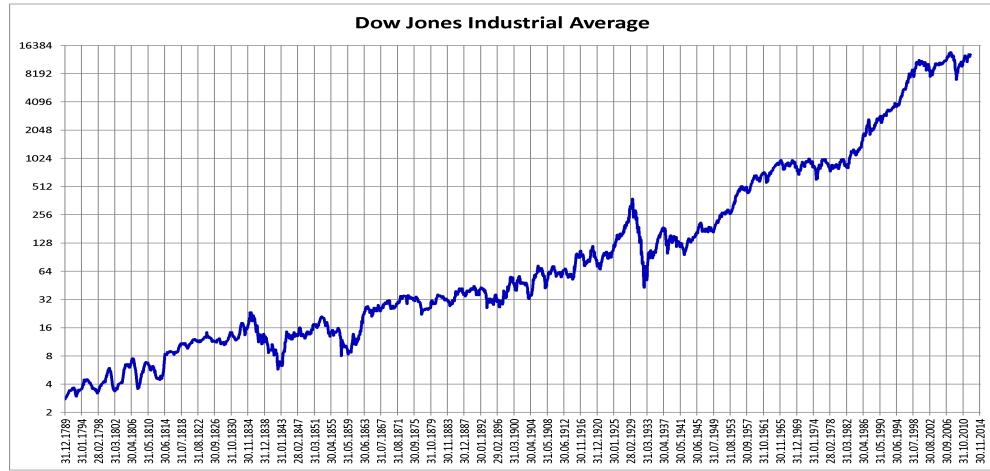


Figure 7.2: **Dow Jones:** An index following the average of the major shares in USA. The index increases with about a factor 4,000. For comparison, the US public debt changed from $\sim 10^8 \$$ in the period 1800-1850 to $\sim 5 \times 10^{12} \$$ in year 2000.

Fig. 7.2 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.

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, $\log(\text{price})$ follow a random walk. That is we in this chapter focus on

$$s(t) = \log(v(t)) \quad (7.1)$$

where $s(t)$ is the log of the price. It is the change in s that represent the Return

in fraction of what is invested:

$$\Delta s = s(t_2) - s(t_1) \sim \frac{v(t_2) - v(t_1)}{v(t_1)} \quad (7.2)$$

The random walk hypothesis was first put forward more than a century ago by Bachelier [134], and has been recently supported by analyzing price fluctuations $W(t)$ as function of time:

$$W^2(T) = \langle (\log v(t+T) - \log v(t))^2 \rangle_t = \langle (\Delta s(T))^2 \rangle, \quad (7.3)$$

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 $W(T) \propto T^{0.55 \rightarrow 0.65}$ 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!

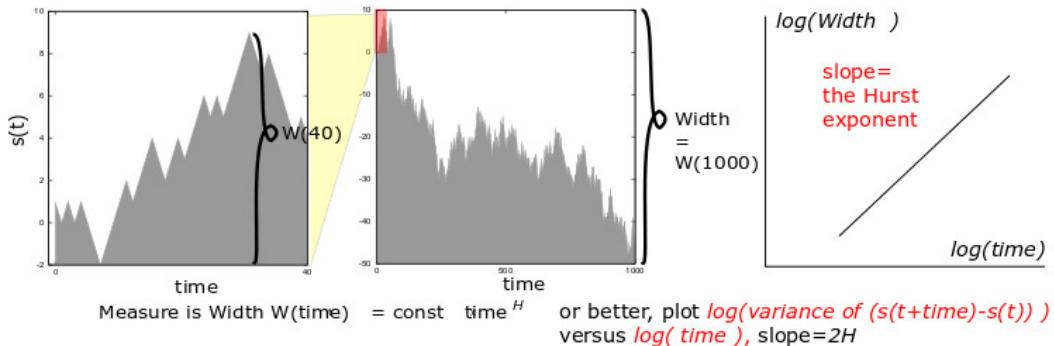


Figure 7.3: **Hurst exponent simplified.** The scaling between the spread in width when measured over different time intervals T . For economic time series the one consider the variable $s = \ln(v)$ and thus plot the spread in $\Delta s = s(t+T) - s(t)$ is a function of the time interval T .

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

$$\langle (\Delta s(T))^2 \rangle = \langle (s(t+T) - s(t))^2 \rangle_t \propto T^{2H} \quad (7.4)$$

where we again follow 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. 7.3. It can be proven (not here) that the Hurst exponents is related to the fractal dimension of the walk, $D = 2 - H$. For example a random walk with $H = 1/2$ have dimension $D = 3/2$.

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.

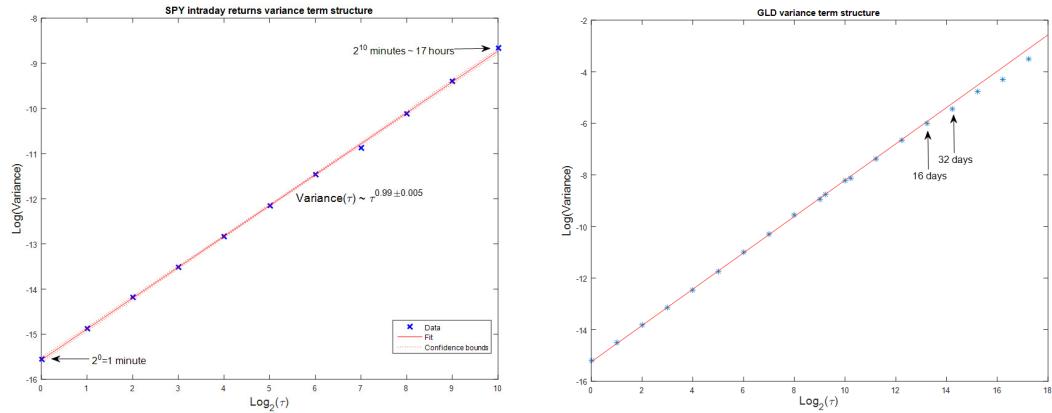


Figure 7.4: **Example of scaling of variance with time.** Shown on different timescales: The left panel focuses on the trading within one day; The right panel includes trading up to one year. In both cases the slope is close to one, corresponding to $H = 1/2$, as this is the Hurst exponent for a random walk walk ($\text{variance} \propto t$). Figure from a blog post by Ernie Chan (Quantitative Investment and trading Ideas (2017)).

The correlation between the past and future is related 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

$$C \equiv \frac{\langle (s(t) - s(t - T)) \cdot (s(t + T) - s(t)) \rangle_t}{\langle (\Delta s(T))^2 \rangle_t} = \frac{\langle -\Delta s(-T) \cdot \Delta s(T) \rangle_t}{\langle (\Delta s(T))^2 \rangle_t}, \quad (7.5)$$

a calculation that demands a few intermediate steps.

The denominator in eq. 7.5 is the variance over the time interval T , which can be re-expressed as:

$$\begin{aligned} f(T) &= \langle (\Delta s(T))^2 \rangle_t = \langle s^2(t + T) + s^2(t) - 2s(t)s(t + T) \rangle_t \\ &= 2(\langle s^2(t) \rangle - \langle s(t)s(t + T) \rangle_t) \propto T^{2H}, \end{aligned}$$

where we use the assumption that an average over all starting time points t makes $\langle s(t + T)^2 \rangle_t$ and $\langle s^2(t) \rangle_t$ equal.

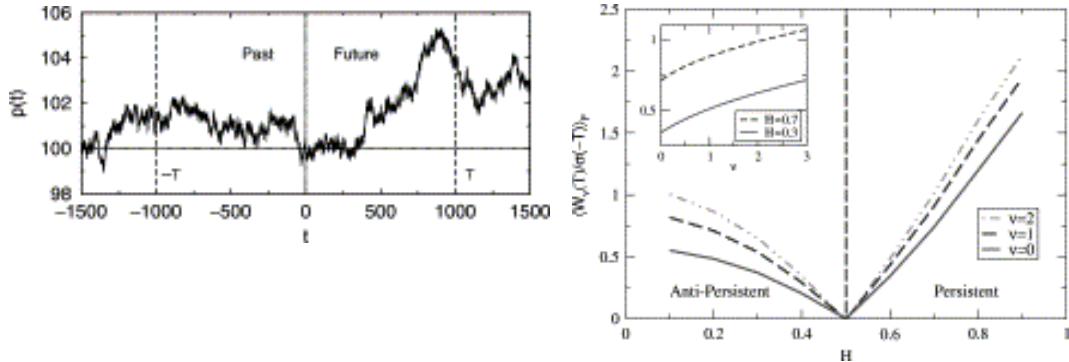


Figure 7.5: **Past → future.** Left shows an example of a time series with Hurst exponent $H = 0.40$, generated by a wavelet method (not pensum). Right panel examines the average return of investment as a function of H , where one buys according to trend [135]. The red curves shows the profit when one 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.

The numerator in eq. 7.5 can be re-expressed

$$\begin{aligned}
 \langle -\Delta s(-T)\Delta s(T) \rangle_t &= \langle -(s(t-T) - s(t))(s(t+T) - s(t)) \rangle_t \\
 &= -\langle s(t-T) \cdot s(t+T) \rangle_t + \langle s(t-T) \cdot s(t) \rangle_t \\
 &\quad + \langle s(t) \cdot s(t+T) \rangle_t - \langle s^2(t) \rangle_t \\
 &= -\langle s(t) \cdot s(t+2T) \rangle_t + \langle s^2(t) \rangle_t \\
 &\quad + 2\langle s(t) \cdot s(t+T) \rangle_t - 2\langle s^2(t) \rangle_t \\
 &= \frac{1}{2}f(2T) - f(T).
 \end{aligned} \tag{7.6}$$

Using that $f(T) = \text{const} \cdot T^{2H}$ one obtain [136, 135]

$$C = \frac{\langle -\Delta s(-T) \cdot \Delta s(T) \rangle_t}{\langle (\Delta s(T))^2 \rangle_t} = 2^{2H-1} - 1 \tag{7.7}$$

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 $\langle (\Delta s(T))^2 \rangle_t = 1$.

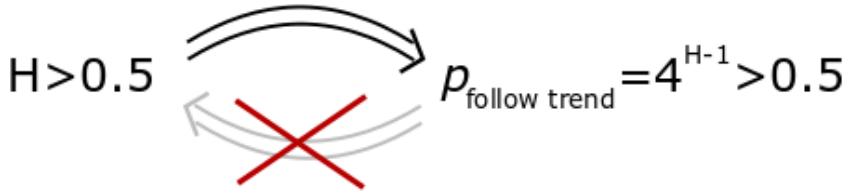


Figure 7.6: **Following the trend.** When $H > 0.5$ then there is more than a 50% chance that the next move is in same direction as the previous move. However, the reverse is not true! A tendency to follow the trend typically implies a random walk with a longer “persistence length,” i.e. a longer time before the walk changes direction. On this longer timescale the walk will still be a random walk. Thus follow on one timescale does not mean the walk is trend follower on all larger timescales.

The numerator in eq. 7.7 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. 7.7 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:

$$\text{Probability (follow the trend)} \sim 4^{H-1}, \quad (7.8)$$

a statement that qualitatively should be true for all time intervals where the walk can be characterized by H (see also Fig. 7.6). Importantly, the equation **cannot be reversed to calculate H** given the probability to follow trend!. This is because the “follow the trend” probability have to be true on widely different scales, say from days to years. 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 [135]. Thus for $H > 1/2$ one should:

$$\text{Buy at } t \text{ if } s(t-T) < s(t) \quad (7.9)$$

$$\text{Sell at } t \text{ if } s(t-T) > s(t) \quad (7.10)$$

whereas this strategy should be reversed in a $H < 0.5$ market, see Fig. 7.5. Noticeably, electricity markets have $H = 0.40$ [137, 138]. Again we emphasize that this buy-sell strategy would work on with trading intervals anywhere inside the time-scale where the walk is characterized by the Hurst exponent H .

Finally, as a small notice, then an walk with Hurst 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)

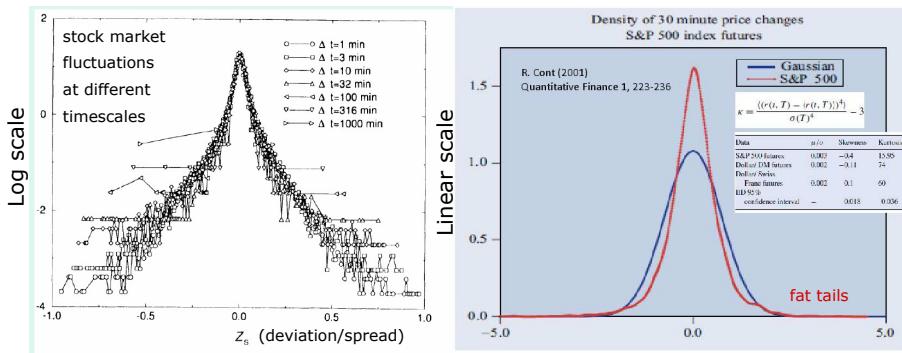


Figure 7.7: **Fat tails:** Distribution of short time-scale fluctuations exhibit fat tails. The left panel show short timescale fluctuations of an index, re-scaled with the timescale over which one examine the fluctuations. In right panel, the red and blue curve have same variance, but different Kurtosis. 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 $p(\text{tail}) \propto 1/\Delta s^\tau$, with $\tau \leq 5$.

- **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:

7.1) 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).

Qlesson: Any finite $p > 1/2$ still leads to a random walk, just with a correlation time that increases with decreasing $1 - p$.

7.2) Simulate a random walk of uncorrelated up and down movements of s , where step length 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. If one instead considered a narrower step size distribution $1/\delta^{3.5}$ then the variance would be well defined and $H=1/2$ would be simply obtained.

7.3) Plot eq. 7.7 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 average profit per investment step for an $H = 0.4$ market.

Qlesson: Act as if tomorrow would be opposite to today.

7.4) 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).



Figure 7.8: **Fear and Panic in 1929.** The crash in 1929 ultimately resulted in a factor 10 reduction of average stock values.

7.2 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 [139]. 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.

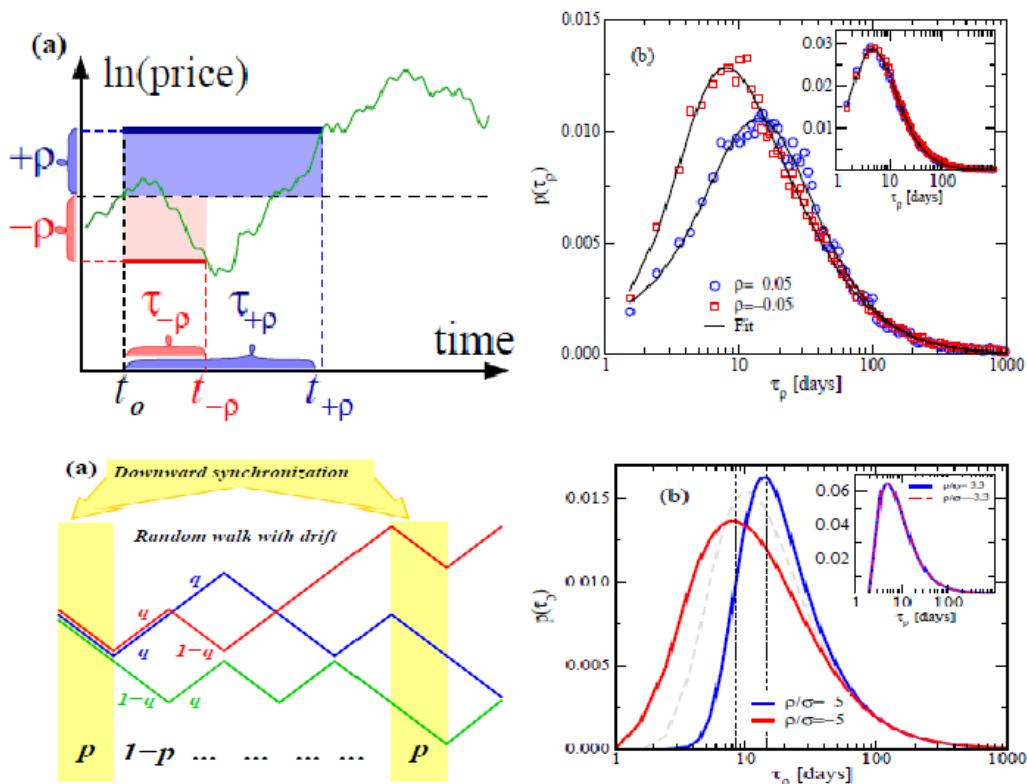


Figure 7.9: Inverse statistics and Fear-factor model. The upper two panels show the definition of "strike price", and the distribution as measured from the de-trended Dow-Jones index. The blue curves show the number of days when the price first exceeds the current price by 5%, the red when it first time become 5% below its current price (The inset illustrate 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.

The data set used is the daily close of the DJIA covering its entire history from 1896 until today. Fig. 7.9 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 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 [140].

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. 7.8. In essence, these collective drops are the cause of the asymmetry in the index.

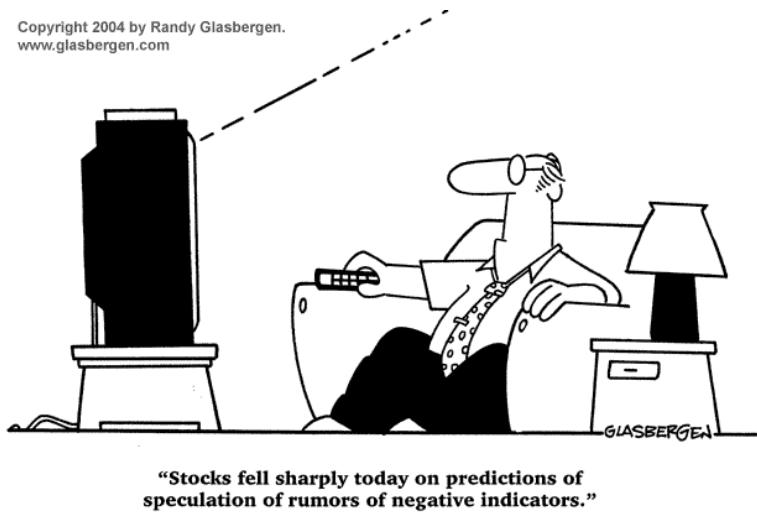


Figure 7.10: Fear is a strong driver at stock markets.

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

$$s_i(t+1) = s_i(t) + \epsilon_i(t)\delta, \quad i = 1, \dots, N, \quad (7.11)$$

where $\delta > 0$ denotes the common fixed log-price increment (by assumption), and $\epsilon_i(t) = \pm 1$ 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:

$$(1 - p) \cdot q = p + (1 - p) \cdot (1 - q), \quad (7.12)$$

fixing q in terms of the probability for overall fear:

$$q = \frac{1}{2 \cdot (1 - p)}. \quad (7.13)$$

A value of $q > 1/2$ for $p > 0$ represent a “compensating” upwards drift that governs the quiet periods between collapses¹. From the price realizations of the N single stocks, one may construct the corresponding price-weighted index, like in the DJIA, according to

$$I(t) = \frac{1}{N} \sum_{i=1}^N v_i(t) \quad (7.14)$$

and investigate inverse statistics for this (Fig. 7.9). 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 $\rho = 5 \cdot \sigma$, 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. 7.10).

- **Summary:** Crowd behavior in short periods of panic has an impact on first passage time statistics.

Questions:

7.5) 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.

Qlesson: The ups and downs of the average is asymmetric, but the average still follows a walk characterized by a with Hurst exponent of 1/2.

¹Note 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

7.3 Modeling Economic Bubbles (Not Pensum)



Figure 7.11: **Herding.** A market driven by herding can give huge volatility (Kalton, in the Economist). Volatility is defined as standard deviation of $\log(s(t)/s(t-1))$ where time unit 1 represent the period we measure volatility over.

At one hand a stock market resembling a random walk seems quite natural as a fixed point, since correlations in the walk give fairly obvious opportunities for profit. Thus people would trade on these opportunities, and presumably reduce them. However, imagine a Hurst exponent larger than 0.5, where an increase would lead to more buying, thus increasing the trend. Thus, the pure random walk seems like an unstable fixed point. It is in this spirit one may view the following model for Economic bubbles.

There have been multiple examples of economic bubbles [141, 142], including the Dutch Tulips (1637), the South sea company (1711-1720), and the Japanese stocks (in the 1980's) and perhaps the contemporary phenomenon of bitcoins [142, 143]. While there exist some considerations on positive feedback [144, 145], bubbles [146] and depressions [147], the economic literature is not settled on the non-equilibrium nature of value [148, 149, 150, 151]. Following Cecilie Toftdahl Olesen & Kim Sneppen (2019) we here present a discrete 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 [152].

Fig. 7.12 illustrates our model in terms of the amount of memory M that people allocate to enthusiasm, while S is associated to the memory associated to common sense. 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 negative feedback.

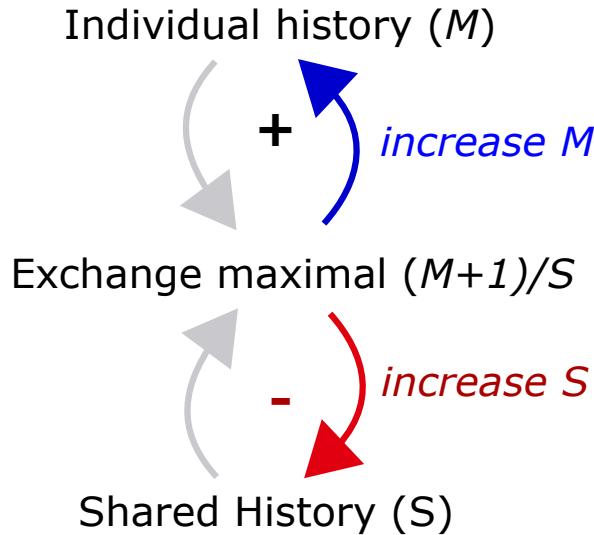


Figure 7.12: **Bitcoin model.** Model with positive (blue) and negative feedbacks (red) on the trading activity of products. The positive feedback amounts to peer-to-peer communication, whereas negative feedbacks may come from a common history of all past transactions that is associated to some sort of fundamental “common sense”.

Consider D different goods. The goods do not need to exist physically, and are only recorded in terms of two types of memory:

A memory that consists of μ slots which each can be assigned a number corresponding to one of the D product types. If several memory slots M is assigned to the same good, it is more likely to be talked about next.

Second, there is a memory of how much each product was traded during the last τ timesteps.

The memories are implemented in lists:

- Perception: The “slots” m_j , $j = 1, \dots, \mu$ that each can be assigned one of the D product types.
- Reality: The slots s_j associated to $j = 1, 2, \dots, \tau$ earlier trading events. Each of these global trading positions is also assigned one of the D product types.

The model is executed in steps. At each step, one select good k with probability:

$$p_k \propto \left(\frac{1 + \sum_{l=1}^{\mu} \delta(m_l - k)}{\sum_{l=1}^{\tau} \delta(s_l - k)} \right)^{\gamma} = \left(\frac{1 + M(k)}{S(k)} \right)^{\gamma}, \quad (7.15)$$

and replicate this memory in both M and S . 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 $M(k)$ that the product k occurs in the memory. The added number 1 in the numerator avoids absorbing states where

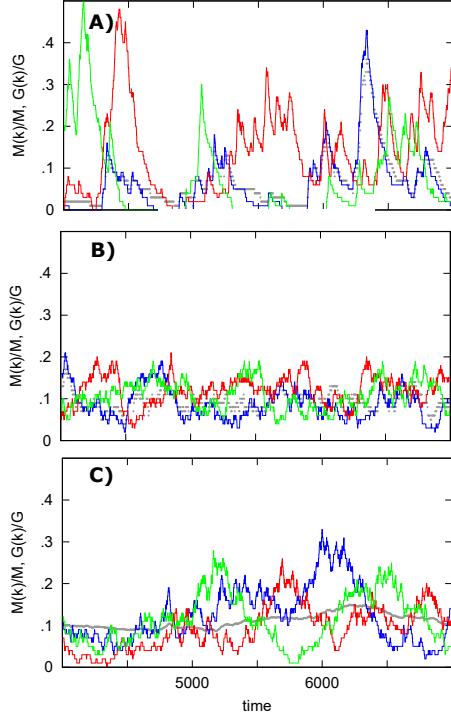


Figure 7.13: **Model simulation:** All panels show 3 products out of $D = 10$. Grey dots show $S(\text{blue})/S$. A) Model with $\mu = \tau = 100$ and infinite γ (select maximum). B) Linear model where copied memory is selected proportional to $(1 + M(k))/S(k)$, i.e. $\gamma = 2$. Again $\mu = \tau = 100$. C) Linear model as in B) but with $\mu = 100$ and $\tau = 1000$.

a product is absent in M . When several products has same maximal p_k 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 memory M by inserting the chosen product k in one randomly chose place x , $m_x = k$.
- Second, one adjusts the shared memory S by inserting the chosen product k in one randomly chosen place y , $s_y = k$.

When the new memory is inserted, an old memory “bit” is discarded. Thereby, τ defines the characteristic time for adjustment of common sense, whereas μ is the lifetime of the individual memory.

The global trade activity of a product k , $S(k) = \sum_{u=1}^{\tau} \delta(S_u - k)$, reflects the common sense memory. 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. 7.13A) show that even similar length of the two memories, $\tau = \mu$, may lead 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. 7.13B,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?.

In our model the amplitude of the fashions depend on a relative slow negative feedback (and on the degree of cooperativity, i.e. whether we select proportionally to p_k or the max p_k). Thus the time-series 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 [153]. In this light “Brand names” allow tightly controlled production, and thereby a slower negative feedback.

The here presented model suggest 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.

Mini tutorial: How could the presented model be implemented as a model of an excitable medium (in a 2 dimensional space)?.

Questions:

7.6) Simulate 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

7.7) 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:

$$\begin{aligned} \frac{dm_1}{dt} &= \left(\frac{m_1 + \epsilon}{s_1}\right)^\gamma \cdot (1 - m_1) - \left(\frac{m_2 + \epsilon}{s_2}\right)^\gamma \cdot (1 - m_2) \\ \theta \cdot \frac{ds_1}{dt} &= \left(\frac{m_1 + \epsilon}{s_1}\right)^\gamma \cdot (1 - s_1) - \left(\frac{m_2 + \epsilon}{s_2}\right)^\gamma \cdot (1 - s_2) \end{aligned} \quad (7.16)$$

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 $m_1 \rightarrow m_1 + \epsilon$ with probability $1 - m_1$ and a change $s_1 \rightarrow s_1 + \epsilon/\theta$ 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 $\epsilon = 0.04$.

7.4 Bet hedging

7.4.1 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, i.e. noise that are larger if the value v larger):

$$dv = \mu \cdot v \cdot dt + \sigma \cdot v \cdot dw \quad (7.17)$$

Here v is the value of a particular stock, and μ the average interest rate. σ sets the size of the noise term, while dw represents the normalized noise and is calculated as the change that happens in the future:

$$dw = w(t + dt) - w(t) \quad (7.18)$$

where w is a random walk process that is normalized such that

$$\langle(w(t) - w(t'))^2\rangle = |t - t'| \quad (7.19)$$

This would be equivalent with $dw/dt = \eta(t)$, where $\langle\eta(t)\eta(t')\rangle = \delta(t - t')$ and $\langle\eta\rangle = 0$ because such a w would follow a random walk moving a step up of dw at each time instant. Thus the typical total increment squared over a 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

$$\begin{aligned} ds &= \log v(t + dt) - \log v(t) \\ &= \log(v + dv) - \log(v) \\ &= \frac{1}{v} \cdot dv - \frac{1}{2v^2}(dv)^2 \\ &= \frac{1}{v}(\mu \cdot v \cdot dt + \sigma v \cdot dw) - \frac{1}{2v^2}(\sigma \cdot v \cdot dw)^2 \end{aligned} \quad (7.20)$$

where we expanded to first order in dt (equal second order in dw). This then give us:

$$\begin{aligned} ds &= \left(\frac{1}{v} \cdot \mu \cdot v - \frac{1}{2v^2} \cdot \sigma^2 \cdot v^2\right) \cdot dt + \frac{1}{v} \cdot \sigma \cdot v \cdot dw \\ &= \left(\mu - \frac{\sigma^2}{2}\right) \cdot dt + \sigma \cdot dw \end{aligned} \quad (7.21)$$

Notice that this last equation only have additive noise, and not multiplicative noise! By integration this equation one obtain

$$s(t) = s(0) + (\mu - \sigma^2/2) \cdot t + \sigma \cdot (w(t) - w(0)) \quad (7.22)$$

where w was a simple random walk and s was the log of the value of your assets.

Notice that the log of your average capital (in principle averaged over many parallel time courses, or as before by starting on many different initial points):

$$\langle s(t) - s(0) \rangle / t = \frac{\langle \log(v(t)/v(0)) \rangle}{t} = \mu - \frac{\sigma^2}{2} \quad (7.23)$$

easily may decrease in value, even if the average return looked positive ($\mu > 0$). [Or in other terms $\langle \exp(\sigma w t) \rangle = \exp(\sigma^2 t / t)$.] Thus even if the average looks good ($\mu > 0$, but the actual long term performance will be bad. Risk (σ) cost on your average performance, because a 50% down movement, followed by a 50% up movement does not cancel out ($0.5 \times 1.5 < 1$). 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. 7.20). The above equation are one cornerstone in evaluating the value of risk, used for setting price on future expectations, see Fig. 7.14.

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

$$\text{growth rate} = \mu \cdot (1 - x) - \frac{\sigma^2}{2} \cdot (1 - x)^2 \quad (7.24)$$

which for small $1 - x$ will be positive if $\mu > 0$. Optimal investment fraction can then be found by differentiation.

$$1 - x = \frac{\mu}{\sigma^2} \quad (7.25)$$

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 if market goes down, then you need to take from your safe money to keep the fraction constant. Reversely, if market values increase, 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.

Noticeably, the above formalism is sensitive to the assumption about Gaussian uncorrelated random walks, where steps is always small, i.e the assumption that

$$\langle (w(t) - w(t'))^2 \rangle = |t - t'| \quad (7.26)$$

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).

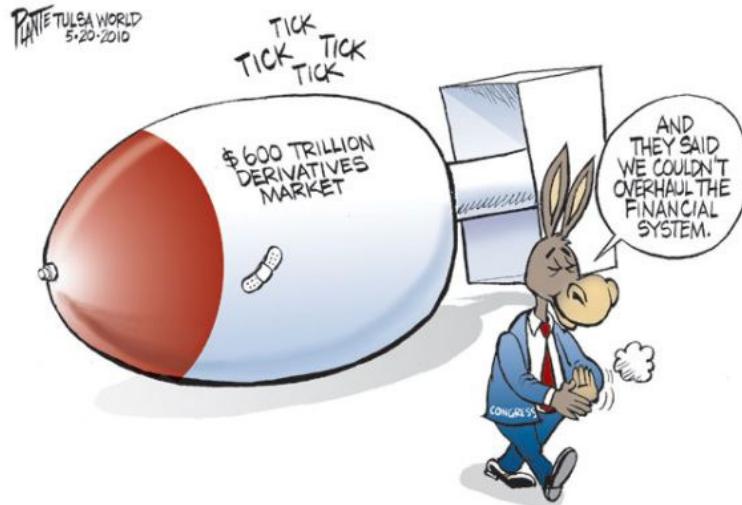


Figure 7.14: **Bet hedging.** The equation that relates risk (volatility, $= \sigma^2$) and average rent μ is a cornerstone in valuating the price of stocks in future, where one for example can buy the right to sell a given share at a given price half a year from now (without buying the share). This has become a huge market, with potential instabilities.

7.4.2 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 reflects optimization of long term growth rate in much the same way as gamblers on horse races may optimize their portfolio [154], and have lately been applied to biology by for example Bergstrøm and Lachman [155] and Kussel and Leibler [156, 157]. Here we use the simplified formulation of [158], directly applicable to simple win-lose games.

Consider a game with two outcomes, one good event where everything invested gets 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 $1 - p$. The game is illustrated in Fig. 7.15. 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

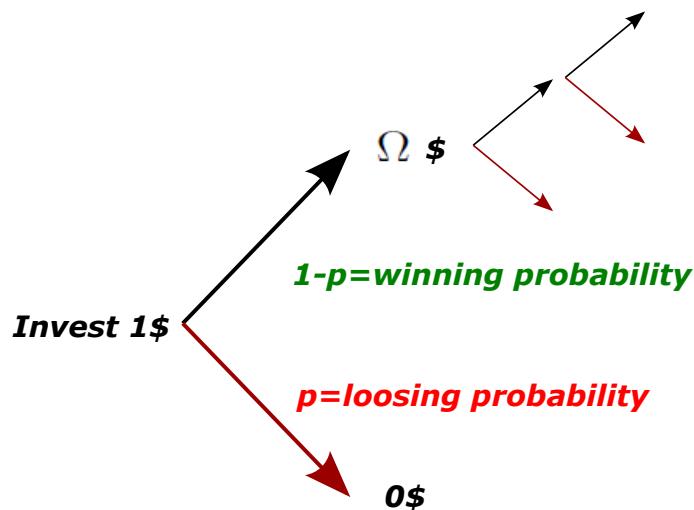


Figure 7.15: **Quit versus gain by factor Ω** An extreme version of quit-double game where p for loose may differ from 50 and gain Ω may differ from a factor 2. We will assume that the game is played many times.

of unit of money is

$$(1 - p) \cdot \Omega + p \cdot 0 = (1 - p) \cdot \Omega \quad (7.27)$$

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

$$\text{propability}(\text{solvent}|t) = (1 - p)^t \rightarrow 0 \quad (7.28)$$

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. This is illustrated in Fig. 7.16. 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 loses the other half. Therefore the typical gain after two games will be the product of returns for a win game and a loose game:

$$\text{Capital} \propto \text{Win} \cdot \text{Loose} \quad (7.29)$$

and if $\text{Loose} = 0$ you will typically have zero money after an equal number of wins and losses.

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 fraction x is a constant throughout all repeated rounds, and thus specifies a strategy.

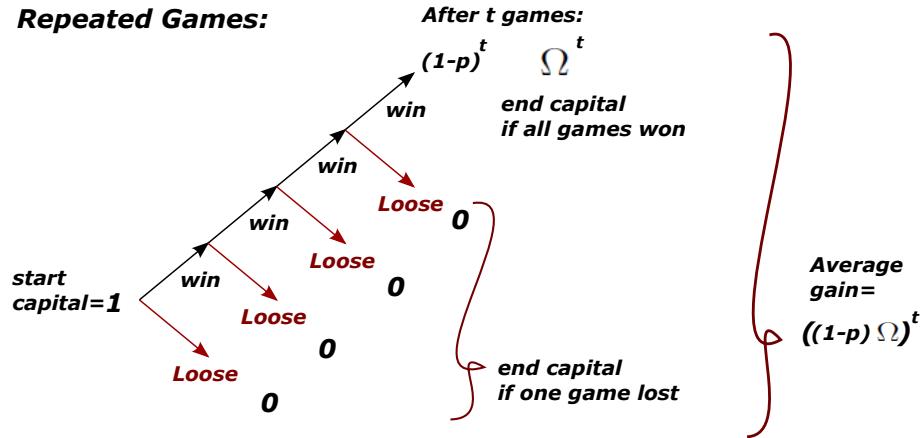


Figure 7.16: **To win repeated games you need extreme luck:** Figure illustrate that the average gain comes about from a very unequal distribution between 1 very lucky player and $(1/p)^t - 1$ normal players.

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

$$\text{Capital} \propto (\Omega(1 - x) + x) \cdot x \quad (7.30)$$

representing the product of a winning situation and a loosing situation. This “Capital” function 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 the Capital:

$$Cap(t) = Cap(0) \cdot \sum_{b=0}^t \left(\frac{t!}{(t-b)!b!} \right) \cdot p^b \cdot (1-p)^{t-b} ((1-x)\Omega + x)^{t-b} x^b$$

where $t! / ((t-b)!b!)$ 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 maximal of the binomial part of the above equation predict an expected number of bad events

$$b = p \cdot t \quad (7.31)$$

Thus we simplify the sum to only contain the term with this expected average. I.e. we focus on the typical behavior. This lead to optimization of

$$\begin{aligned} Cap(t) &\propto ((1-x)\Omega + x)^{t-p} x^p \\ &= (win^{1-p} \cdot loose^p)^t \\ &= (((1-x)\Omega + x)^{1-p} x^p)^t = e^{t\Lambda(x)} \end{aligned} \quad (7.32)$$

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 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 [154].

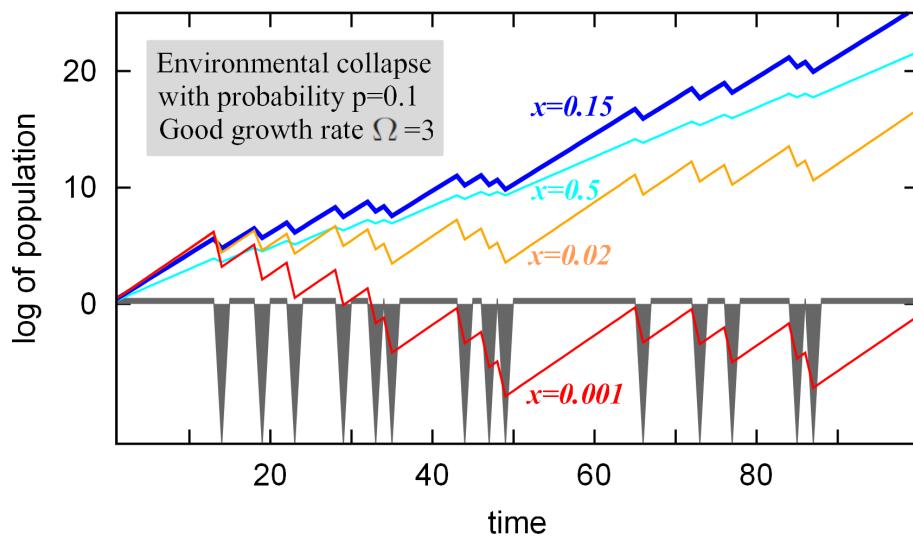


Figure 7.17: **Dynamics of capital.** One can play a game where one wins by factor $\Omega = 3$ by probability $1 - p = 0.9$, and loses all investment with probability $p = 0.1$. The blue curve is the growth of the Kelly-optimal strategy with a fraction $x^* = p \cdot \Omega / (\Omega - 1) = 0.15$ in 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 $x = 0.5$.

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 [154]. 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 [159]. 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 ($\Omega \gg 1$), the optimal strategy is to maintain a safe fraction which is equal to the probability that you loose.

Put in practical use: Imagine that an investment agent suggest you a 12% per year 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 substantial fraction of your capital in your bank account reflecting the probability that the investment agent/company is a crook. If this probability is 50%, keep at least 50% of your money in a safe deposit.

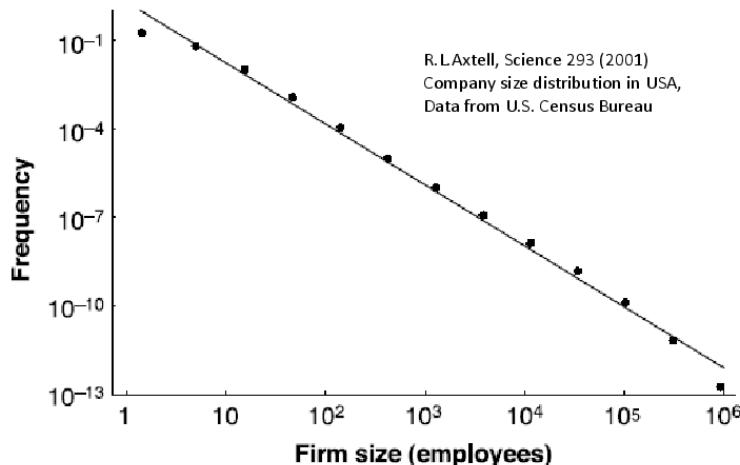


Figure 7.18: **Company size distribution in the USA.** This distribution exhibits a scale free behavior with exponent -2 (data from R.L. Axtel, 2001). A similar exponent may be found by extrapolating between the 7 richest persons in the world owning the same property as the poorest 3.5 billion on the planet.

- **Summary:** Hedge your bets

Questions

7.8 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.

7.9 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 $\omega = 0.5$. 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.

7.10) 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, ($C_i \rightarrow 0$)

That is consider N=100 companies with sizes C_1, C_2, \dots, 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 ($p(C \rightarrow \gamma) \propto C^{-0.2}$)

Qlesson: catastrophes on single company scale transcends into system wide collapses/revolutions.

7.11) 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.

7.12) 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 $s \rightarrow 2 \times s$, respectively $s \rightarrow 1$ with equal probability. Calculate analytically this distribution and compare with Fig. 7.18. 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.

7.13) 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 will be identical to the 1000 agents, because they are anyway non-interacting.

Qlesson: 2^{-t} . Very short lifetime.

Lessons:

- Persistent walks corresponds to Hurst exponent $H > 1/2$, quantified by a correlation coefficient

$$\begin{aligned} C &= \frac{\langle -\Delta s(-T) \cdot \Delta s(T) \rangle_t}{\langle (\Delta s(T))^2 \rangle_t}, \text{ where} \\ \langle (\Delta s(T))^2 \rangle &= \langle (s(t+T) - s(t))^2 \rangle_t \propto T^{2H} \end{aligned}$$

- Stocks are more correlated when the market falls, than otherwise. This is captured by the Fear-Factor model that assigns a probability p to a collective fall and a compensating drift term for the random walk when it is not behaving collectively.
- Bet hedging is a way to deal with the fact that the average return is larger than the typical return in a fluctuating market (and that investments typically(!) follow the typical return). For equal chance the win or loose, the optimal bet hedging fraction x can be calculated from:

$$Capital \propto Win \times Loose = (\Omega(1-x) + x) \cdot x$$

where Ω is the gain factor when winning.

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.

Chapter 8

Dynamic Fronts

“Big whirls have little whirls, That feed on their velocity; And little whirls have lesser whirls, And so on to viscosity.”

— Lewis Fry Richardson

8.1 Burgers Equation

This is a chapter about non-linearities in large spatial systems, and how these give rise to non-equilibrium phenomena, such as shocks and growth. We first describe a motivation to these studies in terms of the classical Burgers equation.

The classical large-scale chaotic system is three-dimensional turbulence. The main complication is here that the large-scale chaotic systems is generated by the Navies-Stokes equations containing the non-linear advection term, $u \cdot \partial u / \partial x$ where u is the velocity field.

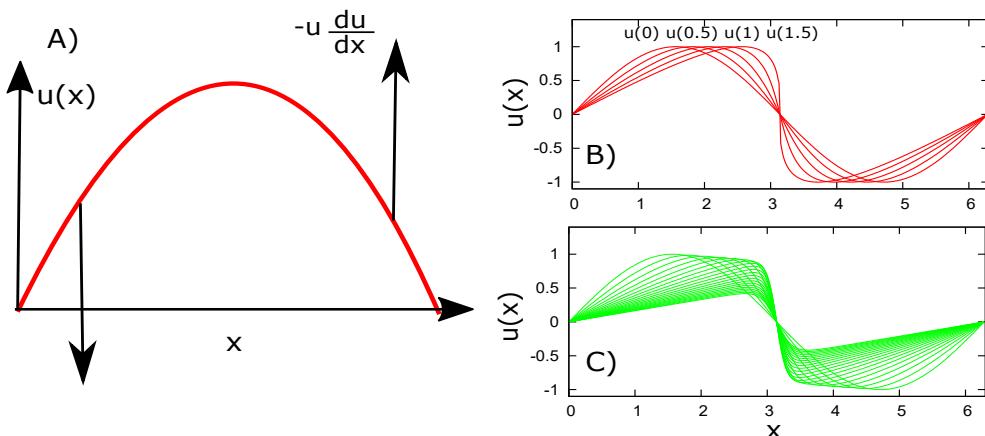


Figure 8.1: A) The term $-u \cdot (\partial u / \partial x)$ makes an initial sin-wave change as indicated with the respective arrows. B) u field for 4 subsequent times ($t = 0, 0.5, 1, 1.5$) for equation $\partial u / \partial t = -u \partial u / \partial x$. C) u field for 20 subsequent times ($t = 0, 0.25, 0.5, \dots, 4.75$) for equation $\partial u / \partial t = -u \cdot \partial u / \partial x + 5(\partial^2 u / \partial x^2)$.

For a velocity field $u(x, t)$

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} \quad (8.1)$$

represent the non-linearity in the Navier-Stokes equation. This term in itself makes initial waves develop into shocks as described in Fig. 8.1. In discrete terms, the above equation is iterated by:

$$\begin{aligned} \frac{\partial u}{\partial t} &= u(x) \cdot \frac{u(x+1) - u(x-1)}{2 \cdot \Delta x} \\ &= -\frac{(u(x+1))^2 - (u(x-1))^2}{4 \cdot \Delta x}, \end{aligned} \quad (8.2)$$

where Δx is the lattice spacing and the last rewrite emphasizes the development of u in terms of gradients in kinetic energy.

The shocks in the above equation are softened when one includes a dissipation term. Thereby we obtain the Burgers equation (see J. M. Burgers, Proc. Acad. Sci. Amsterdam **43** (1940) 2):

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} + \nu \frac{\partial^2 u}{\partial x^2}. \quad (8.3)$$

The dynamics of Eq. 8.3 is illustrated in Fig. 8.1C). One can see that the advection term first lets the shock with sharp gradient develop (until a time $t \sim 1.5$). The development of the shock is associated to transfer of energy from the relatively long initial wavelength of the full system to shorter wavelength associated to the shock. When strong gradients have developed, there is also a cusp with large $\partial^2 u / \partial x^2$ which subsequently dissipates the energy associated to the front. In wave-number space this corresponds to diffusion that quickly dissipates short wavelengths.

The Burgers equation contains both an advection term and a dissipative term with ν as a viscosity coefficient. Compared to the standard diffusion equation, which is linear and thus solvable, the Burgers equation is, as mentioned, non-linear due to the advective term. In the so-called inviscid limit, $\nu \rightarrow 0$, then Burgers equation conserves all moments $u(x, t)^m$ in the velocity field. This is easily seen by integration over an interval. The existence of these conserved moments stands in contrast to the Navier-Stokes equation, which describes a real fluid where only the moments $m = 1$ (momentum) and $m = 2$ (kinetic energy) are conserved. Thus, in that sense, the Burgers equation is rather un-physical.

One can find an exact static solution to (8.3) of the form

$$u(x) = -U \cdot \tanh\left(\frac{Ux}{2\nu}\right) = U \cdot \frac{e^{x \cdot U/2\nu} - e^{-x \cdot U/2\nu}}{e^{x \cdot U/2\nu} + e^{-x \cdot U/2\nu}}, \quad (8.4)$$

where U can take any value (arbitrary velocity scale). To prove this, set

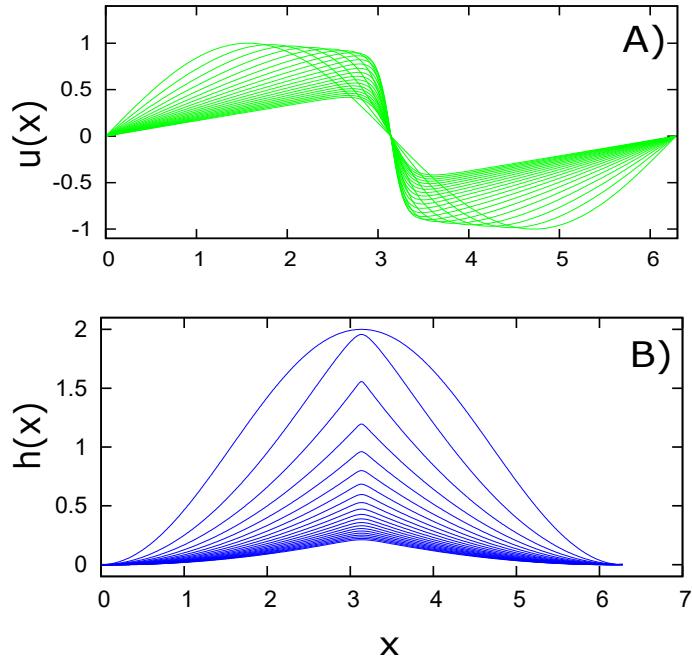


Figure 8.2: **Dynamics in the Burgers and KPZ equations.** A) Burgers equation. u field for 20 subsequent times ($t = 0, 0.25, 0.5, \dots, 4.75$) for equation $\partial u / \partial t = -u \partial u / \partial x + 5(\partial^2 u / \partial x^2)$. B) KPZ equation. $h = \int u dx$ for 20 subsequent times ($t = 0, 0.25, 0.5, \dots, 4.75$) for equation $\partial u / \partial t = -u \partial u / \partial x + 5(\partial^2 u / \partial x^2)$. Thus $\partial h / \partial t = -(\partial h / \partial x)^2 + (5/2)(\partial^2 h / \partial x^2)$.

$y = x \cdot U/2\nu$ with $dy/dx = U/2\nu$ and

$$\begin{aligned} \frac{du}{dx} &= U \cdot \frac{U}{2\nu} \frac{(e^y + e^{-y})^2 - (e^y - e^{-y})^2}{(e^y + e^{-y})^2} = \frac{U^2}{2\nu} \frac{4}{(e^y + e^{-y})^2} \\ \frac{d^2u}{dx^2} &= -\frac{2U^3}{\nu^2} \frac{e^y - e^{-y}}{(e^y + e^{-y})^3} \\ u \frac{du}{dx} &= \frac{U^3}{\nu} \frac{2(e^y - e^{-y})}{(e^y + e^{-y})^3}, \end{aligned}$$

from which one sees that the latter two terms cancel each other (when one multiplies the dissipation term with ν), thus proving that the proposed u from eq. 8.4 is a stationary solution to the Burgers equation. This proposed u changes from $-U$ to $+U$ when x changes from $-\infty$ to $+\infty$, and it changes most strongly within the length scale ν/U . This length scale approaches zero when viscosity approaches zero.

It is convenient to convert the Burgers equation (8.3) into dimensionless form by introducing dimensionless variables $\hat{u} = \frac{u}{U}$, $\hat{x} = \frac{x}{L}$, $\hat{t} = \frac{t}{T}$, thus defining typical velocity, length and time scales U, L, T with $U = \frac{L}{T}$. Inserting this into (8.3) we find

$$\frac{\partial \hat{u}}{\partial \hat{t}} + \hat{u} \frac{\partial \hat{u}}{\partial \hat{x}} = \frac{1}{Re} \frac{\partial^2 \hat{u}}{\partial \hat{x}^2}, \quad (8.5)$$

where we have obtained a dimensionless quantity, the Reynolds number $Re = \frac{UL}{\nu}$, which is a measure of the strength of the motion (and subsequently the “turbulence”).

Notice that the Reynolds number equals the system size divided by the scale at which substantial changes takes place in the u -field.

In 1950, Hopf and Cole found a beautiful exact and general solution to the Burgers equation defining one of the few exactly solvable non-linear field theories (others being exact soliton solutions to non-linear Sine-Gordon and Korteweg-de Vries (KdV) equations). If $\phi(x, t)$ is a solution to the linear diffusion equation

$$\frac{\partial \phi}{\partial t} = \frac{1}{Re} \cdot \frac{\partial^2 \phi}{\partial x^2} \quad (8.6)$$

then an exact solution to the non-linear Burgers equation (with ‘hats’ deleted) is obtained by a logarithmic derivative (Hopf-Cole transformation)

$$u(x, t) = -\frac{2}{Re} \cdot \frac{\partial \ln \phi}{\partial x} = -\frac{2}{Re} \cdot \frac{1}{\phi} \frac{\partial \phi}{\partial x} \quad (8.7)$$

To prove this, we need to insert this u into the Burgers equation. Here the right hand side equals:

$$\frac{\partial}{\partial t} u = \frac{2}{Re} \cdot \frac{1}{\phi^2} \frac{\partial \phi}{\partial t} \frac{\partial \phi}{\partial x} - \frac{2}{Re} \cdot \frac{1}{\phi} \frac{\partial}{\partial x} \frac{\partial \phi}{\partial t} = \frac{2}{Re} \cdot \frac{1}{\phi^2} \frac{\partial \phi}{\partial t} \frac{\partial \phi}{\partial x} - \frac{2}{Re^2} \cdot \frac{1}{\phi} \frac{\partial^3 \phi}{\partial x^3}$$

and the non-linear term equals

$$u \frac{\partial u}{\partial x} = -\frac{2}{Re} \cdot \frac{1}{\phi} \frac{\partial \phi}{\partial x} \cdot \left(\frac{2}{Re} \cdot \frac{1}{\phi^2} \left(\frac{\partial \phi}{\partial x} \right)^2 - \frac{2}{Re} \cdot \frac{1}{\phi} \frac{\partial^2 \phi}{\partial x^2} \right),$$

whereas the diffusion term equals

$$\begin{aligned} \frac{\partial^2 u}{\partial x^2} &= \frac{\partial}{\partial x} \left(\frac{2}{Re} \cdot \frac{1}{\phi^2} \left(\frac{\partial \phi}{\partial x} \right)^2 - \frac{2}{Re} \cdot \frac{1}{\phi} \frac{\partial^2 \phi}{\partial x^2} \right) \\ &= -\frac{4}{Re} \cdot \frac{1}{\phi^3} \left(\frac{\partial \phi}{\partial x} \right)^3 + \frac{4}{Re} \frac{1}{\phi^2} \left(\frac{\partial \phi}{\partial x} \right) \frac{\partial^2 \phi}{\partial x^2} + \frac{2}{Re} \cdot \frac{1}{\phi^2} \left(\frac{\partial \phi}{\partial x} \right) \frac{\partial^2 \phi}{\partial x^2} - \frac{2}{Re} \cdot \frac{1}{\phi} \frac{\partial^3 \phi}{\partial x^3}. \end{aligned}$$

Putting these terms together one can prove that the Hopf-Cole transformation indeed works (prove this yourself). The diffusion equation for ϕ gives the decay of an initial state $\phi = \sum_{n=1}^{\infty} a_n \cdot \cos(n \cdot x)$ as

$$\phi = \sum_{n=1}^{\infty} a_n \cdot e^{-n^2 \cdot t / Re} \cdot \cos(n \cdot x), \quad (8.8)$$

where a_n are coefficients associated to the Fourier decomposition of the initial state ($t = 0$). Prove this by inserting into eq. 8.6. The corresponding u then is

$$u(x, t) = -\frac{2}{Re} \cdot \frac{1}{\phi} \frac{\partial \phi}{\partial x} = \frac{2}{Re} \cdot \frac{\sum_{n=1}^{\infty} n a_n \sin(nx) e^{-n^2 \cdot t / Re}}{\sum_{n=1}^{\infty} a_n \cos(nx) e^{-n^2 \cdot t / Re}}. \quad (8.9)$$

The modes in this sum are damped exponentially by the factor $\exp(-n^2 t/Re)$. This dampening originates from the second spatial derivative in the diffusion equation. In practice one would only need to include a few of the lowest modes, that is, $n = 1, 2, \dots$.

Now we turn to the connection to the Kardar–Parisi–Zhang (KPZ) equation. The KPZ equation with $\lambda = -1$ is in fact an integrated version of the noisy Burgers equation for the variable $u = dh/dx$:

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial x^2} - u \frac{\partial u}{\partial x} + \frac{\partial}{\partial x} \eta \quad (8.10)$$

where η is white (uncorrelated) Gaussian noise. The differentiation of this noise term η leads to a noise which is conserved noise, that is, one adds and subtracts changes close to each other in space. The above identification is proved by inserting $u = \partial h/\partial x$:

$$\frac{\partial}{\partial x} \frac{\partial h}{\partial t} = \nu \frac{\partial}{\partial x} \frac{\partial^2 h}{\partial x^2} - \frac{\partial h}{\partial x} \frac{\partial}{\partial x} \frac{dh}{dx} + \frac{\partial}{\partial x} \eta. \quad (8.11)$$

Using that $(\partial h/\partial x) \cdot \partial^2 h/\partial x^2 = (1/2) \cdot \partial/\partial x \cdot (\partial h/\partial x)^2$, we can identify

$$\frac{\partial h}{\partial t} = \nu \frac{\partial^2 h}{\partial x^2} - \frac{1}{2} \left(\frac{\partial h}{\partial x} \right)^2 + \eta + \text{constant}. \quad (8.12)$$

Thereby, the shocks of the Burgers equation translate into discontinuous derivatives of h . This latter equation is called the KPZ equation. In the following we will develop the formalism to deal with this equation.

Questions:

7.1) Consider a random walker in velocity space. What would be the scaling of its displacement in normal space with time. (use that $x(t) = \int dt' (t - t') \eta(t')$ with $\langle \eta(t) \eta(t') \rangle = u_0 \delta(t - t')$ in the calculation of the ensemble averaged $\langle x(t) x(t') \rangle$).

Qlesson: The displacement is faster than ballistic! It recapitulates the famous Richardson scaling from turbulence. In fact also Kolmogorov scaling is consistent with parts that just do uncorrelated random walks in velocity space. **7.20)** Show that $d/dt \int u^m dx = 0$ if u evolves as described by the $\nu = 0$ limit of the Burgers equation, when fixing the boundary $u(0) = 0$ and $u(1) = 0$.

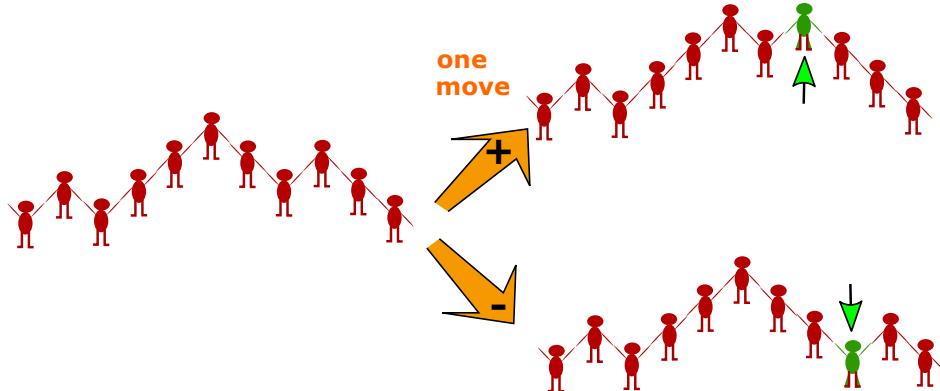
Qlesson: As a corollary, momentum and energy are conserved when there is no dissipation.

7.21) Show that the Burgers equation for the field $u(x, t)$ is invariant under the transformation $x - ct \rightarrow x$, $t \rightarrow t$ and $u - c \rightarrow u$ where c is a fixed velocity.

Qlesson: When one moves along the interface with velocity c the equations remain unchanged. This is called Galilean invariance.

7.22) If U is a typical velocity and L a typical length scale, then show that the Burgers equation can be transformed into

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \frac{1}{Re} \cdot \frac{\partial^2 u}{\partial x^2},$$

String of people holding hands, moving randomly:

Edwards-Wilkinson: Movements in both directions (detailed balance, equilibrium)

KPZ: Only Forwards movement (non equilibrium)

Figure 8.3: **Illustration of the Edwards-Wilkinson and KPZ equations.** The points of the interface are people that hold hand with their neighbors. Each person may move, but only under the constraint of holding hands with their neighbors. When people can move both forward and backward equally, the string of people can be described by the Edwards-Wilkinson equation. When people only are allowed to move forward, then the universality class changes. This change is reflected in addition of the non-linear term that is characteristic of the KPZ equation.

where the variables u, x, t are now dimensionless and $Re = U \cdot L/\nu$ is the Reynolds number.

Qlesson: *Re is the number that matters, as it is the ratio of largest scale to dissipative scale.*

7.23) Show that $\varphi(x, t)$ is a solution to the diffusion equation:

$$\frac{\partial \varphi}{\partial t} = \frac{1}{Re} \cdot \frac{\partial^2 \varphi}{\partial x^2} .$$

Then show that the full solution to the dimensionless Burgers equation is

$$u(x, t) = -\frac{2}{Re} \cdot \frac{\partial \ln(\varphi)}{\partial x} .$$

This is called the Hopf-Cole transformation.

8.2 Noise-driven Interface Dynamics

8.2.1 Linear Interface Dynamics

In this chapter we will study the system that can be illustrated as in Fig. 8.3. That is, we are going from just considering one random walker to considering a string of walkers that update randomly under certain rules.

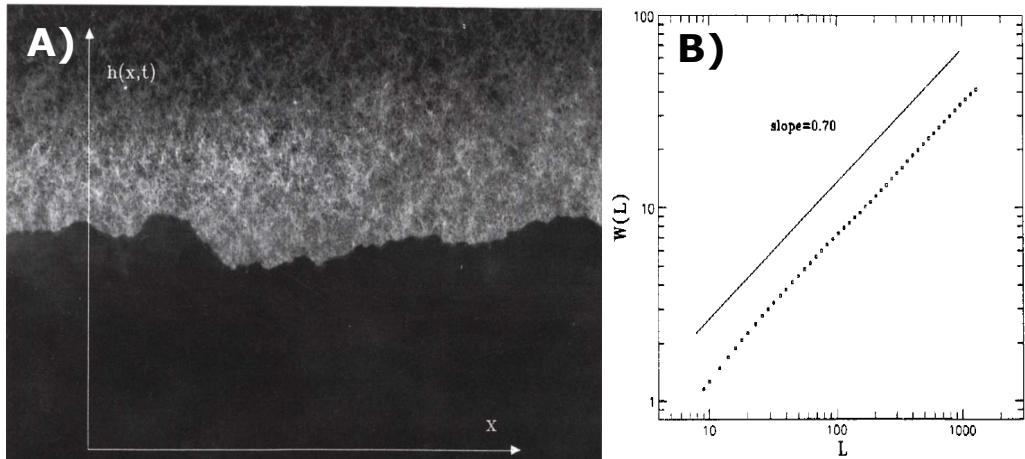


Figure 8.4: **Fire propagation.** A) 8.5 cm segment of burning paper (total width 46cm). Fire propagates upwards. B) Variation of the width of the front height (the standard deviation of $h(x, t)$) with the length of the segment that one measure this standard deviation over (from Jun Zhang et al., Physica A 189, 383 (1992)).

Edwards-Wilkinson equation (Random strings): We consider a one-dimensional interface that evolves due to stochastic noise, *see* Fig. 8.4, where a piece of paper is burning from below, and a interface is formed at the edge of the fire.

In general such an interface represents the separation between two widely different media. A system could be the front formed by random deposition and or evaporation of grains to an aggregate. The interface then represents the border of the aggregate. The advance of the interface represent the addition of grains. The random addition of particles can occur very slowly (diffusion limited) or very fast (reaction limited). The case of random fast addition is also called *Eden growth*.

Throughout these notes, we restrain ourselves to interfaces that have no overhangs, and thus can be described as a single valued function $h(x, t)$ of position x and time t . This means that there will be two independent scaling exponents: for h as function of x and for h as function of t .

The analysis of these types of interface dynamics is done in terms of its width w and how this scales with extension L and time t at which we allow it to evolve, *see* Fig. 8.5. In Fig. 8.4B) one sees that a burning paper develop a front with a width that scale as $W \propto L^{0.7}$.

The general types of interfaces we will consider can be described as "roughening of strings" where dh/dt does not depend on h itself, and where the growth is symmetric under left-right symmetry $x \rightarrow -x$. This means we consider equations of the form:

$$\frac{\partial h}{\partial t} = f\left(\frac{\partial h}{\partial x}\right) + \eta \quad (8.13)$$

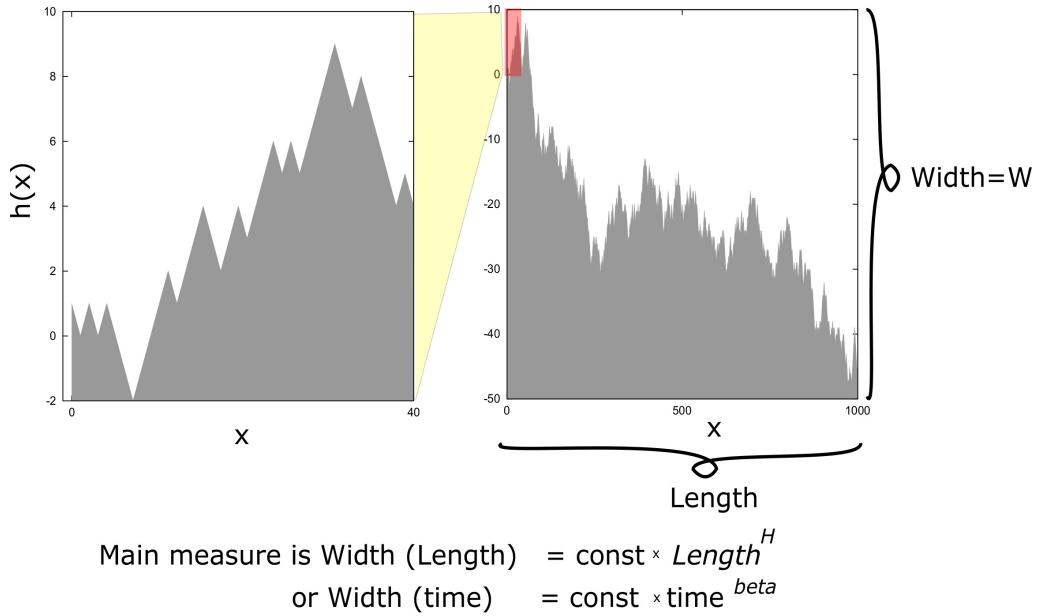


Figure 8.5: **Exemplifying the Hurst exponent.** Interface at two different length scales, with the horizontal axis scaled by a factor 25 and the vertical (height) axis scaled by a factor $\sqrt{25} = 5$. The scaling of width W with length (after a long time) is measured using the Hurst exponent $H = \chi$ and the scaling of how the width W initially increases with time is quantified through β .

where f contains only terms with an even number of differentiations with respect to x , and where η is some random noise (to be specified further below).

An example of a stochastic equation is the Edwards-Wilkinson equation (EW):

$$\frac{\partial h}{\partial t} = \nu \frac{\partial^2 h}{\partial x^2} + v + \eta(x, t), \quad (8.14)$$

which was suggested to model how the height h of e. g. a pile of sand varies, when one distributes sand randomly with rate $v + \eta(x, t)$ over the system (this is called "Eden Growth"). The constant velocity term v can be eliminated by using the transformation $h \rightarrow h - vt$, and thus we will always work with $\langle \eta \rangle = 0$. If further $\eta = 0$ for all x and t , then the diffusion term $\nu \partial^2 h / \partial x^2$ will eliminate height fluctuations. This is also seen from writing the linear diffusion equation (eq. 3) for an arbitrary wavenumber k : $h = h_k(t) \exp(ikx)$: Fig. 8.6 shows the effect of the main terms of the interface equation, as well as the term we will add in the subsequent KPZ equation.

$$\frac{dh_k}{dt} = -\nu k^2 h_k \quad (8.15)$$

giving that waves with wavelength $\lambda = 2\pi/k$ decay exponentially: $h_k(t) = \exp(-\nu k^2 t)$. Without noise the systems dissipates, fastest for small length scales, and generally reaches a steady flat state without height fluctuations.

When we include the noise term η , there will be established a balance between the dissipation (due to d^2/dx^2) and the excitation due to η . This balance obviously depends on the value of η . In simple modeling one may assume that η is random Gaussian with $\langle \eta \rangle = 0$ and with finite correlation length (σ_x) and finite correlation time (σ_t). That the noise is Gaussian means that it has a characteristic size, which we set to a fixed value Γ . This noise can be written formally as a covariance of η :

$$\begin{aligned} \langle \eta(x, t)\eta(x', t') \rangle_{ensemble} &= \frac{\Gamma}{4\sigma_x\sigma_t} e^{-|x-x'|/\sigma_x} e^{-|t-t'|/\sigma_t} \\ &\approx \Gamma \delta(x - x')\delta(t - t') \end{aligned} \quad (8.16)$$

where $\langle \cdot \rangle_{ensemble}$ represents an ensemble averaging (average over many systems)¹. The last step represents the shape of the noise when its correlation length is much smaller than the scale under which we study the fluctuations of our system. On that large scale one may say that the noise is random Gaussian and uncorrelated in space and time.

Technical solution of linear interface model: Consider the evolving interface within an interval $[0, L]$, and assume periodic boundary conditions. The Fourier modes of h are defined as

$$h_k = \frac{1}{\sqrt{L}} \sum_{x=1}^L e^{-ikx} h(x) \quad (8.17)$$

$$h(x) = \frac{1}{\sqrt{L}} \sum_{q=-\pi}^{\pi} e^{iqx} h_q = \frac{\sqrt{L}}{2\pi} \int_{-\pi}^{\pi} dq e^{iqx} h_q \quad (8.18)$$

where $k = -\pi, \dots, \pi$ in increments of $2\pi/L$, i. e. we have the following relation between summation and integration

$$\frac{2\pi}{L} \sum_k = \int dk$$

The EW for the Fourier modes reads:

$$\frac{dh_k}{dt} = -\nu k^2 h_k + \eta_k(t). \quad (8.19)$$

Using

$$\sum_{x=1}^L e^{i(k+q)x} = L\delta(k + q)$$

¹Notice that the finite range expression of the noise is non analytic in 0. In reality the behaviour at 0 should be analytic. The above expression has the right long range form for a noise with a finite correlation length σ_x . Another type of noise could be the Gaussian shaped

($= \Gamma \frac{\pi}{2\sigma_x\sigma_t} \exp(-(x - x')^2/2\sigma_x^2) \exp(-(t - t')^2/\sigma_t^2)$) but this would not have the exponential tail of noise with the standard definition of correlation length.

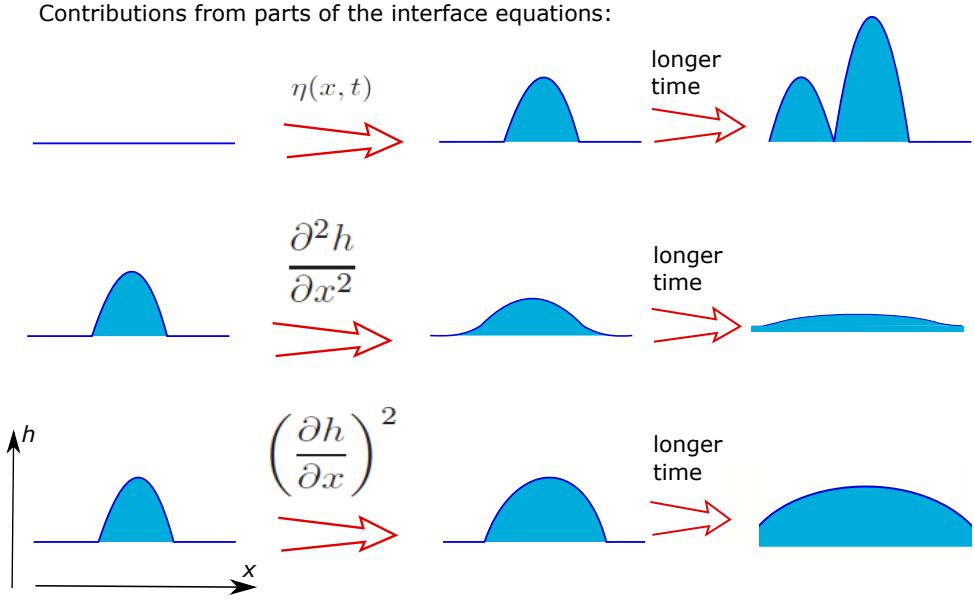


Figure 8.6: **Illustrating the different terms of the interface equations.** Contribution to $h(x, t)$, where the top panel illustrates the noise, always added at random points, the middle panel illustrates the diffusion terms, that tend to flatten the interface (under conservation of the average h), and the lowest panel, which illustrates the effect of the non-linear terms, that nearly always add, but also tend to make interface flatter.

and using eq. 8.17 for the noise we get the correlation of the Fourier components of the noise:

$$\begin{aligned}
 \langle \eta_k(t) \eta_q(t') \rangle_{ensemble} &= \left\langle \frac{1}{L} \sum_{x', x''} e^{-i(kx' + qx'')} \eta(x', t) \eta(x'', t') \right\rangle_{ensemble} \\
 &= \frac{1}{L} \sum_{x', x''} e^{-i(kx' + qx'')} \langle \eta(x', t) \eta(x'', t') \rangle_{ensemble} \\
 &\stackrel{(8.16)}{=} \frac{1}{L} \sum_{x', x''} e^{-i(kx' + qx'')} \Gamma \delta(x - x') \delta(t - t') \\
 &= \frac{1}{L} \Gamma \delta(t - t') \sum_x e^{-i(k+q)x} \\
 &\stackrel{(8.20)}{=} \Gamma \delta(k + q) \delta(t - t'). \tag{8.20}
 \end{aligned}$$

I.e. Fourier modes at k and $-k$ provide "kicks" simultaneously, and are uncorrelated from previous kicks.

The solution to the first order linear equation (8.19) is given as the sum of exponentially decaying contributions from each kick from the noise:

$$h_k(t) = \int_0^t dt' \eta(k, t') \exp(-\nu k^2(t - t')) .$$

The mean height, $\langle h \rangle_x$ is determined by h_0 :

$$\langle h \rangle_x = \left\langle \frac{1}{\sqrt{L}} \sum_{q=-\pi}^{\pi} e^{iqx} h_q \right\rangle_x = \frac{1}{\sqrt{L}} \sum_{q=-\pi}^{\pi} \langle e^{iqx} \rangle_x h_q = \frac{1}{\sqrt{L}} h_0 . \quad (8.21)$$

The corresponding width expressed in sum of only positive wavenumbers k is (using Equations (8.19), (8.20) for the second equality):

$$\begin{aligned} w^2(L, t) &\stackrel{\text{def}}{=} \frac{1}{L} \left\langle \int_0^L (h(x) - \langle h \rangle_x)^2 dx \right\rangle_{\text{ensemble}} \\ &= \frac{1}{L} \int_0^L (\langle h(x)^2 \rangle_{\text{ensemble}} - \langle \langle h \rangle_x^2 \rangle_{\text{ensemble}}) dx \\ &= \frac{2}{L} \sum_{k=2\pi/L}^{\pi} \langle h_k h_{-k} \rangle_{\text{ensemble}} = \frac{1}{\pi} \int_{2\pi/L} S(k) dk , \end{aligned} \quad (8.22)$$

where the last sum excludes the $k = 0$ mode because it is canceled by the average term squared. Here the structure function, $S(k)$, is given by ensemble average of the time dependent h_k multiplied with h_{-k} (use delta function of noise correlation with time Eq. (8.20)):

$$\begin{aligned} S(k) &= \langle h_k h_{-k} \rangle_{\text{ensemble}} \\ &= \left\langle \int_0^t dt' \eta(k, t') e^{-\nu k^2(t-t')} \int_0^t dt'' \eta(k, t'') e^{-\nu k^2(t-t'')} \right\rangle_{\text{ensemble}} \\ &= \int_0^t dt' \int_0^t dt'' \langle \eta(k, t') \eta(k, t'') \rangle e^{-\nu k^2(t-t')} e^{-\nu k^2(t-t'')} \\ &= \Gamma \int_0^t dt' e^{-2\nu k^2(t-t')} \\ &= \frac{\Gamma}{2\nu} k^{-2} (1 - e^{-2\nu k^2 t}) , \end{aligned} \quad (8.23)$$

where the last integration can be done by substituting $u = 2\nu k^2(t' - t)$ with lower boundary $u = -2\nu k^2 t$ and upper boundary $u = 0$. This in turn gives that the width in the average stationary state for large times t , i.e. where $\sqrt{t\nu} \gg L$:

$$w^2(L, t) = \frac{\Gamma}{2\pi\nu} \int_{2\pi/L}^{\infty} k^{-2} (1 - e^{-2\nu k^2 t}) dk \approx \frac{\Gamma}{2\pi\nu} [-k^{-1}]_{2\pi/L}^{\infty} \propto L \quad \text{for } t \text{ large} \quad (8.24)$$

or $w(L, t = \infty) \propto L^\chi$ with $\chi = 1/2$. This means that a static snapshot of the interface looks like a random walker along the x axis.

For times t where $\sqrt{t\nu} \ll L$ the width changes in time (substitute $y = k\sqrt{t\nu}$):

$$w^2(L, t) = \frac{\Gamma}{2\pi\nu} (2\nu t)^{1/2} \int_{2\pi\sqrt{t\nu}/L}^{\infty} y^{-2} (1 - e^{-2y^2}) dy \quad (8.25)$$

$$\begin{aligned} &\propto \sqrt{t} \int_0^{\infty} dy (1 - e^{-2y^2}) / y^2 \\ &\propto t^{1/2} = t^{2\beta} \text{ for small } t, \end{aligned} \quad (8.26)$$

where the first approximation is valid when $y \ll kL$ and the second expresses the whole width in terms of a function g that is constant for small arguments (and decease for large arguments). In this equation $\beta = 1/4$ and $z = 2$ comes from the lower end of the integral when $\sqrt{t\nu}/L$ becomes size ably. Thus the interface roughens with time as $w \propto t^{1/4}$. The cutoff exponent z set by $\sqrt{t\nu}/L \sim 1$ or $t \propto L^2 = L^z$ describes how long time it takes before saturation is reached.

Lesson: Here we have a simple remarkable result. When looking at a point of the interface (or the relative movement of two well separated points), then the movement is random, but not in the sense of having the Hurst exponent of a random walk, but rather with

$$H = \beta = 1/4 \quad (8.27)$$

Thus a point on the interface performs an **anti-persistent** random walk, it will repeatedly revert previous deviations, and will do so on all scales. This anti-persistence comes about because of the infinite memory in the surrounding interface.

The time roughing exponent $\beta = 1/4$ of the interface may be understood as a "random walker on a random walker". I. e. defining nodes of the interfaces where $h - \langle h \rangle$ changes sign, then these nodes perform a random walk. On top of this slow walk, the interface rapidly roughens to a width that for each node configuration is given approximately by the square root of the typical distance between the nodes.

The scaling in eq. 8.25 is occasionally reformulated as:

$$w(L, t) = L^x f(tL^{-z}) \quad (8.28)$$

where $f(x) = \text{const} > 0$ for $x \rightarrow \infty$ and $f \approx x^\beta$ for $x \rightarrow 0$. Here $z = 2$ is the *dynamic exponent*, meaning the exponent that relates space scaling to time scaling.

The lesson of the dynamics of the EW equation is that there are two regimes, a **transient** regime where the width increases $W \propto t^{1/4}$. For a finite system of length L this increase cannot continue, but saturates when one reaches a width $W \propto L^{1/2}$. Obviously it takes a time $t \propto L^z$ with $z = 2$ to

reach this saturated steady state regime. In this **saturated regime** the interface fluctuates around a statistically stationary state where height in positions that are distances x apart typically differ by $x^{1/2}$. Again, as in the transient case, the time for a complete reorganization of the interface scale as L^2 .

Physical solution of linear interface model: The scaling of the saturated height is simply the scaling of profiles of a random string with hamiltonian:

$$H = \frac{\nu}{2} \sum \left(\frac{\partial h}{\partial x} \right)^2 \quad (8.29)$$

where now ν plays the role of a string tension prohibiting large stretching of the string.

To get the equation of motion for a system with hamiltonian $H(h, x)$ one may use to equation (analog to $dx/dt \propto -dV/dx$ for strongly damped motion in potential $V = V(x)$):

$$\frac{dh}{dt} = -\frac{\delta H}{\delta h} \quad (8.30)$$

where the functional derivative of H w. r. t. h at a given point i is calculated from

$$\frac{\delta H}{\delta h_i} = \frac{\delta}{\delta h_i} \sum_j (h_j - h_{j-1})^2 \quad (8.31)$$

$$= 2(h_i - h_{i-1}) - 2(h_{i+1} - h_i) = 2(h_{i+1} - 2h_i + h_{i-1}) \quad (8.32)$$

where one only obtains contributions from the two terms where h_i enters). The configuration space of a string with hamiltonian that secures finite uncorrelated steps is the space of all random directed strings, i. e. the configuration space of 1-d random walks (Boltzmann weighting and replacing \sum with \int in (8.30)):

$$P_{EW}(\{h\}) = \exp \left(-\frac{\nu}{\Gamma} \int dx \left(\frac{\partial h}{\partial x} \right)^2 \right), \quad (8.33)$$

where ν now plays the role of string tension and Γ is an effective temperature, constraining the typical string slopes to be of size $dh/dx \approx \sqrt{\Gamma/\nu}$. Random strings without finite bounded slopes meander as a random walk, thus we see also directly that $\langle (h - \langle h \rangle)^2 \rangle^{1/2} \propto L^{1/2}$.

The mapping of saturated states of EW to the equilibrium states of a random string, can also be seen from simulating its dynamics on a lattice:

Define h_i as integer values for $i = 1, \dots, L$. Set initially $h_i = 0$ for all i . At each time-step select a random position i . Then select with probability 1/2 whether the move at i should be $h \rightarrow h' = h + 1$ or $h \rightarrow h' = h - 1$. If the new h' has a slope to one of the neighboring sites which exceeds 1 then reject the move and select new position i . Otherwise accept the move and select a new position i for the next move.

In the stationary state this algorithm does fulfill detailed balance (see chapter 1): the ensemble averaged probability to move from local configuration a to configuration b equals the probability to move in the opposite direction (prove it). When the algorithm is simulated on a computer, an initially flat configuration indeed roughens to a self-affine structure with scaling exponents $\chi = 1/2$.

8.2.2 Non-Linear Interface Dynamics

Karder-Parisi-Zhang equation: The lattice deposition model introduced in the previous chapter does not address the situation where grains are only added, never removed. Obviously this would be a more realistic description of a growth process. If we only add grains then detailed balance will be violated: Consider the three neighboring heights $(h, h, h + 1)$ that can transition to $(h, h + 1, h + 1)$, whereas the opposite is impossible, although both states are allowed by the dynamics. Therefore lateral growth breaks detailed balance and our interface is not described by equilibrium dynamics.

The simplest non-linear addition to the EW equation leads us to the KPZ equation (after Kardar-Parisi-Zhang [160]):

$$\frac{\partial h}{\partial t} = \nu \frac{\partial^2 h}{\partial x^2} + \frac{\lambda}{2} \left(\frac{\partial h}{\partial x} \right)^2 + \eta(x, t), \quad (8.34)$$

where the velocity term is again removed by appropriate shifting of the coordinate system. The overall direction of the motion imposed by the non-linear term opens for non-equilibrium sampling of state space.

The scaling behavior of KPZ is somewhat different from that of the EW. In particular, assume the scaling relations:

$$x \rightarrow b x, \quad h \rightarrow b^\chi h, \quad t \rightarrow b^z t \quad (8.35)$$

and consider equation on rescaled length $bL \gg L$:

$$b^{\chi-z}(dh/dt) \approx \nu b^{\chi-2} \frac{\partial^2 h}{\partial x^2} + \frac{\lambda}{2} b^{2\chi-2} (\partial h / \partial x)^2 + b^{-1/2-z/2} \eta \quad (8.36)$$

From this re-scaled equation we immediately read off several lessons:

First, in the absence of the non-linearity all remaining terms are equally important, if $z = 2$. Therefore, we can in fact determine all scaling exponents of the Edwards-Wilkinson equation by noting that it samples random strings, with roughness exponent as that of the random walk: $\chi = 1/2$. With the dynamic exponent $z = 2$ this gives a temporal scaling $W \propto t^{\chi/z} = t^{1/4}$.

Second, in the KPZ equation with a non-linear term, on large scales this non-linearity will dominate the linear term ($b^{2\chi-2} \gg b^{\chi-2}$). The temporal growth

therefore scales as the non linear growth, implying the “Galilean invariance” scaling relation ²:

$$\chi - z = 2\chi - 2 \Rightarrow z + \chi = 2 \quad (8.37)$$

which in fact is true in all dimensions of x (notably also in two spatial dimensions).

Third, if we include higher order linear or non-linear terms, then these will not contribute (on large scales). This is because each x -derivative gives an additional scaling factor $b^{\chi-1}$, and the contributions would thus be insignificant on large enough scales ($b \gg 1$) when $\chi < 1$!! Thus higher order terms might alter the small scale shape of the propagating front, but will not alter its scaling.

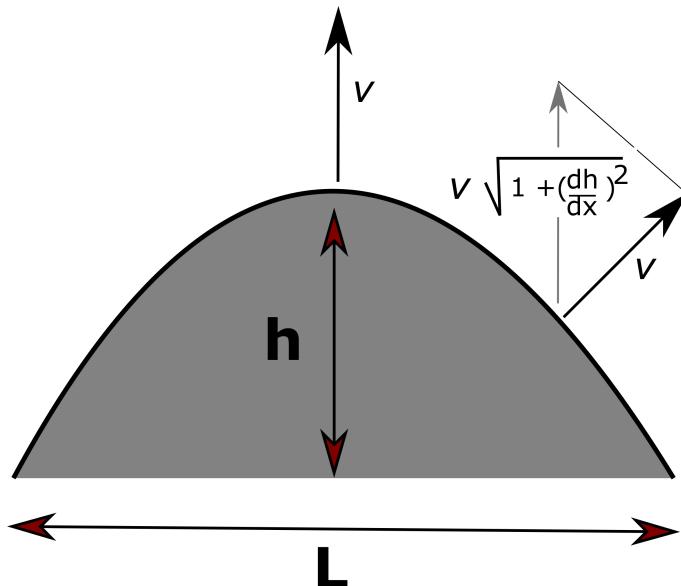


Figure 8.7: **Bulge dynamics.** Growth perpendicular to the local interface results in growth along pre-defined vertical axis $dh/dt = v \cdot (1 + dh/dx)^{1/2}$ where v is the typical velocity of the front.

The scaling of the noise term

$$\langle \eta(x, t)\eta(x', t') \rangle = \frac{\Gamma}{4\sigma_x\sigma_t} e^{-|x-x'|/\sigma_x} e^{-|t-t'|/\sigma_t} \quad (8.38)$$

²Galilean invariance refer to invariance of description under tilting the interface an angle ϵ and simultaneously describing its evolution in a coordinate system that moves with constant velocity along the interface:

$h' = h + \epsilon x$, $x' = x - \lambda_{net}$ and $t' = t$ then for the tilted interface:

$dh'/dt' = d^2h'/dx'^2 + \lambda/2(dh/dx)^2 + \eta'(x' + \lambda\epsilon t', t')$, just as for the original interface. See Medina et al. Phys. Rev. **A39** 3053

with $x \rightarrow bx$ consist of a scaling $\sigma_x \rightarrow b\sigma_x$ and $\sigma_t \rightarrow b^z\sigma_t$ implying that the total noise on scale bL scales as $b^{-(z+1)/2}\eta$ (i.e. $\langle \eta\eta \rangle \propto b^{-1-z}$). For EW the noise remains on the same scale as the linear term. For the KPZ equation the noise diminishes on large scales (but is still important for providing the seeds from which the lateral growth can roughen the interface).

We need to find one other exponent relation (beyond $\chi + z = 2$). We will prove that $\chi = 1/2$ is the same as for the EW equation. This is done by showing that the equilibrium distribution $P(\{h\})$, see Equation (8.33), given by the EW equation is stationary under the non-linear term:

$$\begin{aligned} \frac{d}{dt} P_{EW}(\{h\}) &= \frac{d}{dt} \exp\left(-\frac{\nu}{\Gamma} \int dx (\frac{\partial h}{\partial x})^2\right) \\ &= -\frac{\nu}{\Gamma} \left(\int dx \left(\frac{d(\frac{\partial h}{\partial x})^2}{dt} \right)_{NL} \right) \cdot \exp\left(-\frac{\nu}{\Gamma} \int dx (\frac{\partial h}{\partial x})^2\right) \\ &\propto \int dx \frac{\partial h}{\partial x} \frac{d}{dx} \left(\frac{\partial h}{\partial t} \right)_{NL} P_{EW}(\{h\}), \end{aligned} \quad (8.39)$$

where the h change due to the non-linear term $(\partial h/\partial t)_{NL} = \lambda/2(\partial h/\partial x)^2$ means that we should evaluate:

$$\int dx \frac{\partial h}{\partial x} \frac{\partial}{\partial x} \left(\frac{\partial h}{\partial t} \right)_{NL} = \int dx \frac{\partial h}{\partial x} \frac{\partial}{\partial x} \left(\frac{\partial h}{\partial x} \right)^2. \quad (8.40)$$

Partial integration transforms the above into

$$= - \int dx \frac{\partial^2 h}{\partial x^2} \left(\frac{\partial h}{\partial x} \right)^2 = -\frac{1}{3} \int dx \frac{\partial}{\partial x} \left(\frac{\partial h}{\partial x} \right)^3 = 0, \quad (8.41)$$

where we in the final step used the periodic boundary conditions (string of length L , then $h(L+1) = h(1)$, $h(L+2) = h(2)$). Thus, the static picture of a saturated interface of the KPZ equation equals that of the EW equation.

Notice that one can repeat the above argument by using the Laplacian on the distribution of strings. Then one will see a strictly negative term, i.e. that interfaces tend to become flatter with time. It is the combination of the Laplacian and noise that maintain the random strings in absence of the non linear term. The above calculation demonstrate that the non-linear term subsequently does not change this distribution.

In summa for KPZ we have the scalings:

$$W(L, t) \propto L^\chi \quad \text{for } t \gg L^z \quad (8.42)$$

with roughness exponent $\chi = 1/2$ and

$$W(L, t) \propto t^\beta \quad \text{for } t \ll L^z \quad (8.43)$$

with temporal exponent $\beta = 1/3$ and dynamic exponent $z = \chi/\beta = 3/2$.

Lesson: Here we have another remarkable result. When looking at a point of the interface (or relative movement if two well separated points), then the movement is random, but not as a random walk, but rather with a Hurst exponent

$$H = \beta = 1/3 \quad (8.44)$$

This behavior is **anti-persistent** as it was in the case of the linear interface, but *less* anti-persistent. A deviation will revert back, but the surrounding interface is dragged more with the movement of a single point (because of the non-linearity) and anti-persistence is therefore weakened.

The KPZ equation can be simulated by a variety of deposition rules. Thus the details of the growth rule do not matter, as long as the up down symmetry is broken, as expected from **universality**. In Fig. 8.8 we use:

Define a discrete interface $h(x)$ on a chain $x = 1, 2, 3, \dots, L$. The chain is updated by selecting a random site. At this site one unit is added to h . Then neighbouring sites are sequentially ($y = x+1, x+2, \dots$ and $x = x-1, x-2, \dots$) adjusted upwards ($h \rightarrow h + 1$) precisely until all slopes $|h(y) - h(y-1)| \leq 1$.

Questions:

7.3) Simulate discrete interface model:

Define a discrete interface $h(x)$ on a chain $x = 1, 2, 3, \dots, L$. The chain is updated by selecting a random site. On this site one unit is attempted to be added to h . If, after the addition, the neighbor sites fulfill $|h(x) - h(x+1)| \leq 1$ and $|h(x) - h(x-1)| \leq 1$ then the attempted addition is accepted. Else the addition is rejected (interface is not advanced) and one proceeds to a new attempt at adding at another position.

Simulate the dynamics for a system of size $L = 100$ with periodic boundary conditions. Calculate the width (W) of the interface as a function of time since it was flat. Calculate the width when the interface has reached saturation. Repeat the simulation for the system size $L = 10000$.

Qlesson: It takes time to reach steady state: The needed time is proportional to $L^{3/2}$ updates per lattice site ($L^{5/2}$ updates). Notice that for a 100 times larger system the width should be increased by a factor 10 ($\chi = 1/2$).

7.4) Repeat the above simulation, but now either randomly adding a unit to h , or randomly subtracting a unit from h . Examine again the width as function of time.
Qlesson: Notice that the symmetric model roughens more slowly (with exponent 1/4) than the model where we only add to h . Notice that the same behavior will be obtained by e.g. the model in end of chapter 7.2.1.

7.5) What happens if instead of selecting a move up or down with equal probability, one selects always both an up and a down move, separated by the finite lattice spacing 3.

Qlesson: This corresponds to noise that is exactly conserving. The interface rough-

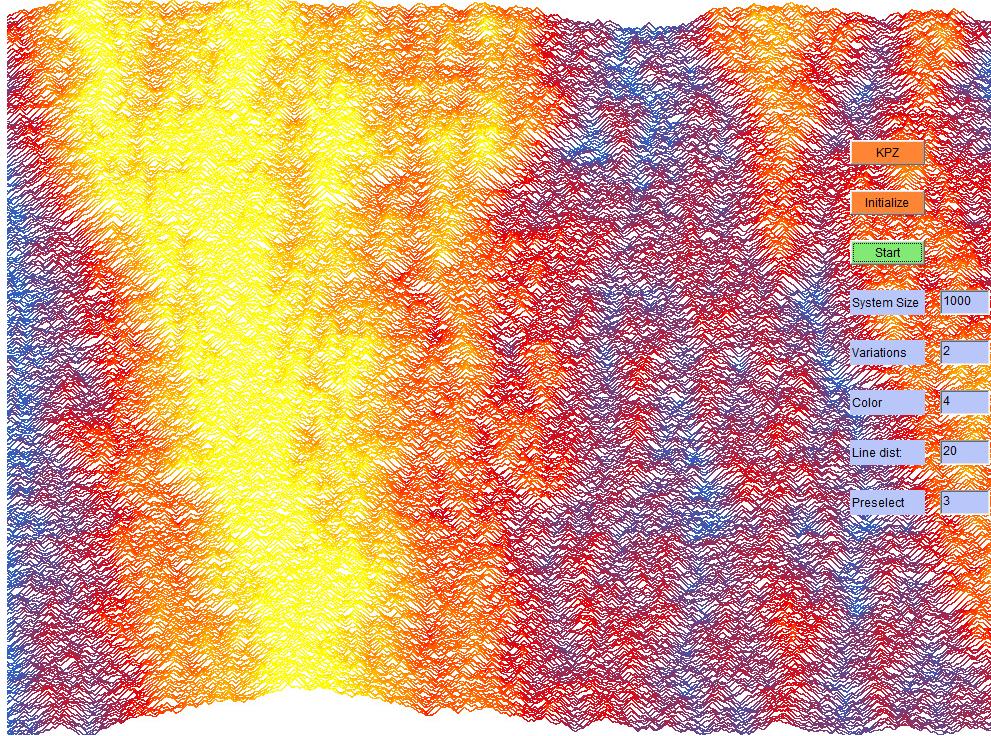


Figure 8.8: Artistically colored propagation of a stochastic cellular automata version of the KPZ equation, with yellow marking relative large h values, and blue the parts of front that is behind. The lateral growth due to the $(dh/dx)^2$ term is seen as a tendency that the yellow area is spreading.

ens much more slowly.

7.6) What happens if $\partial^2 h / \partial x^2$ is replaced by $-\partial^4 h / \partial x^4$ in EW (conserved noise).

Qlesson: The Fourier method works, the interface roughens much more slowly.

7.7) How will the EW equation roughen in 2+1 dimensions? (random surfaces without overhangs).

Qlesson: Much weaker (logarithmically), use Fourier decomposition.

7.8) How does the cross over time between linear ($w \propto t^{1/4}$) and non linear growth ($w \propto t^{1/3}$) scale with the size of diffusion ν and the non-linear term λ ? How large a system is needed to see the effect of a non-linearity?

Qlesson: Use ν in front of the Laplacian, and compare scaling $\nu l^{\chi-2}$ with $\lambda l^{2\chi-2}$, i.e $\nu \sim \lambda l^\chi$ giving $l \sim (\nu/\lambda)^{1/\chi} = (\nu/\lambda)^2$. Thus, for doubled diffusion coefficient the scale for the cross over is four times longer..

7.9) What happens if one adds $(dh/dx)^4$ to the right hand side of the KPZ equation? What happens if instead one replaces $\lambda/2(dh/dx)^2$ with $\lambda/2(dh/dx)^4$ in the KPZ equation?

Qlesson: Basically nothing..., convince yourself

7.10) The KPZ equation should supposedly describe universality for a quite large class of growth phenomena. However experiments consistently give a roughness exponent $W \propto L^\chi$ with $\chi = 0.6 \rightarrow 0.8$ [?], whereas $\chi_{KPZ} = 1/2$. What could be reasons to the apparent failure of KPZ?

Qlesson: Don't believe in all you are told.

8.3 Deterministic Interface Dynamics with Chaos

Kuramoto-Sivashinsky Equation: Can the noisy KPZ behavior be obtained from a deterministic equation? In that case the noise should be caused by an intrinsic chaotic evolution of the front. We therefore now turn our attention to an interface equation with an intrinsic instability. Such instability might occur if the part of the interface that is ahead tends to propagate faster than the parts of the interface that are behind. An example could be the fire front of a burning paper, where there will be most material to burn for the parts of the interface that are ahead. To lowest order this is a linear instability of the form:

$$\frac{\partial h}{\partial t} = -\frac{\partial^2 h}{\partial x^2} + \dots \quad (8.45)$$

This lowest order description captures the dynamics of an evolving interface that is diverging at all wavelength and unlimited fast for very short wavelength. Adding a short wavelength cut off (e. g. due to a minimal ignition heat for the firefront) and the simplest possible non-linearity one obtains the deterministic Kuramoto-Sivashinsky (KS) equation:

$$\frac{\partial h}{\partial t} = -\frac{\partial^2 h}{\partial x^2} - \frac{\partial^4 h}{\partial x^4} + \left(\frac{\partial h}{\partial x} \right)^2, \quad (8.46)$$

where the constants in front of the 3 terms are rescaled away by appropriate scaling of space x , time t and height h . In Fourier space the KS equation reads:

$$\frac{dh_k}{dt} = k^2(1 - k^2)h_k - \sum_q q(k - q)h_q h_{k-q}. \quad (8.47)$$

The difference $k - q$ in final term arises from:

$$\begin{aligned} \left(\frac{\partial h}{\partial x} \right)^2 &= \sum_{q'} iq' e^{iq'x} h_{q'} \sum_{q''} iq'' e^{iq''x} h''_q \\ &\stackrel{q'+q''=k}{=} - \sum_k e^{ikx} \sum_{q'} q'(k - q') h_{q'} h_{k-q'}, \end{aligned}$$

when identifying terms with the same k -number. One notices that all k vectors contribute to the evolution of a given wavevector, e.g., the large wave vector modes will get many random kicks from interfering short modes around the most unstable wave vector. The understanding of this non-linear term poses one of the main problems in non-linear physics today; in a slightly disguised form it also appears in the Navier Stokes equation.

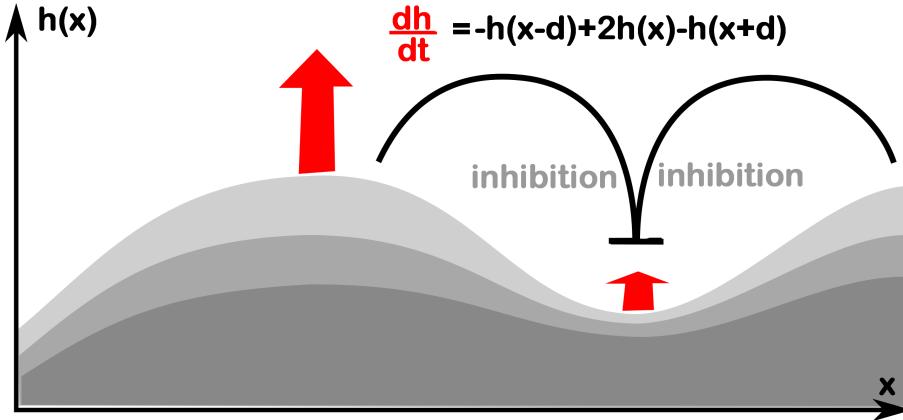


Figure 8.9: **Growth with an instability.** The instability is caused by an ‘unfair’ advantage to the part of the growing colony that is ahead of rest. The front at some horizontal coordinate x will grow at some speed (here denoted v) plus a correction associated to whether the neighborhood is ahead or behind $h(x)$.

From Eq. (8.47) we see that the fastest linearly growing mode is with $k^2 = 1/2$ or wavelength $\lambda = 2\pi\sqrt{2}$. This mode will dominate the early development. However it will not explode completely, because the non-linear term takes over when slopes become sufficiently large. This is seen from the two-mode solution

$$h = a \cos(kx) - b \cos(2kx) \quad (8.48)$$

for $k < 1$ and $2k > 1$, where we allow ourselves to ignore higher modes $3k, 4k\dots$ because of their huge dissipation (by the d^4/dx^4 term). The equations for the amplitudes a and b respectively are:

$$\frac{da}{dt} = k^2(1 - k^2)a - 2k^2 a b \quad \text{and} \quad \frac{db}{dt} = 4k^2(1 - 4k^2)b + \frac{1}{2}k^2 a^2, \quad (8.49)$$

which have the stationary ($da/dt = 0$ and $db/dt = 0$) solutions:

$$(a, b) = (2\sqrt{(4k^2 - 1)(1 - k^2)}, (1 - k^2)/2). \quad (8.50)$$

For k between $1/2$ and 1 , Eq. 8.50 is a stable fixed point (prove this by investigating solutions $a + \delta a, b + \delta b$ to first order in $\delta a(t)$ and $\delta b(t)$). The average growth velocity of this simple two wave solution is

$$\frac{\partial h}{\partial t} = \left\langle \left(\frac{\partial h}{\partial x} \right)^2 \right\rangle_{ensemble} = \frac{1}{2}a^2 k^2 + 2b^2 k^2 = \frac{1}{2}k^2(1 - k^2)(7k^2 - 3), \quad (8.51)$$

which is maximal for $k = 0.8751$. That is, for this k value the stationary solution propagates fastest, and regions with this k vector may get ahead and spread (very loose argument).

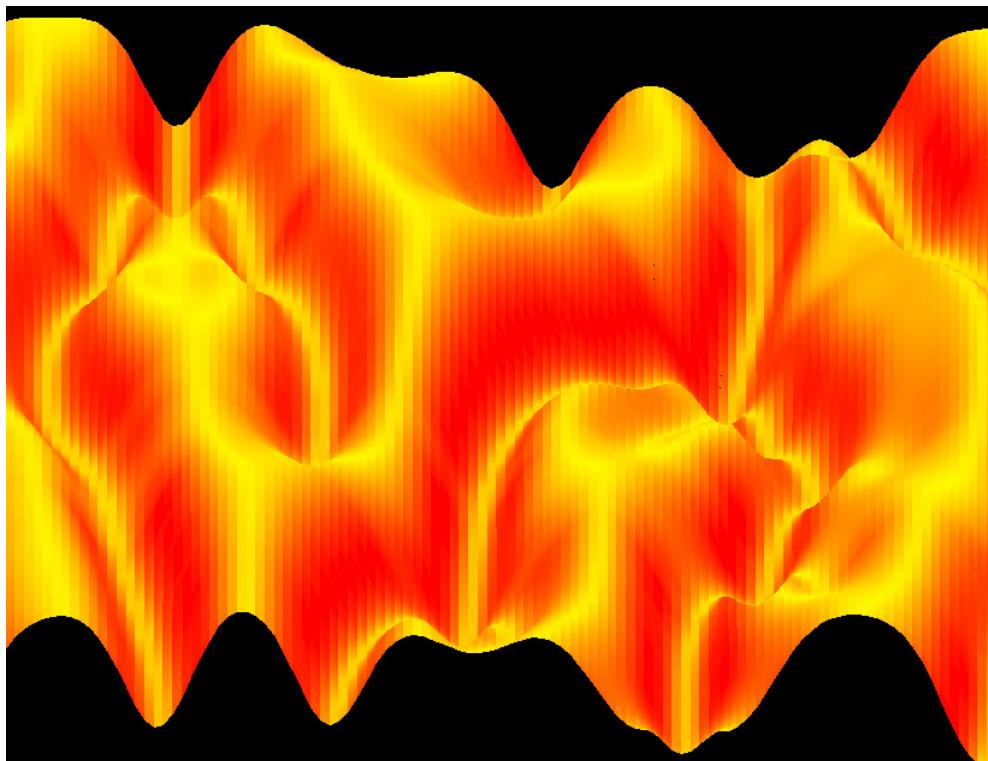


Figure 8.10: *Artistically colored propagation of dynamics of the Kuramoto-Sivashinsky equation. Notice the tip splitting dynamics, and the valley merging.*

In fact the two mode solution is a metastable solution to the KS equation for $k \in [0.845, 0.912]$. However it constitutes a smaller fraction of phase space as we consider larger evolving systems. The largest part of solutions to KS are known to be chaotic. Positive Lyapunov exponents would lead to noisy behavior, and de-correlate the dynamics between regions that are separated by distances much larger than a typical wave velocity divided by the Lyapunov exponent. From this one can argue that one may represent their relative movement by a Gaussian uncorrelated noise.

Using the basic symmetry of the KS equation with the expected Gaussian noise behavior, we expect that the large scale behavior of the Kuramoto-Sivashinsky equation is in the same [universality class](#) as the KPZ equation (This behavior was first suggested by V. Yakhot using some renormalization arguments). In 1-d this conjecture indeed seems to hold true. However as the characteristic cell size has wavelength $\ell \approx 2\pi/k_{max}$ and as the characteristic time to develop such a cell is about $\delta t = 1/(k_{max}^2(1-k_{max}^2))$ with $k_{max} \approx \sqrt{1/2}$, then the effective diffusion constant $\nu = \ell^2/\delta t \approx 15$.

Because ν is very large, the “viscous” term will dominate the non-linear term up to quite large length scales. One has to consider quite large systems before the non linearity can influence the scaling.

In order to visualize the KS equation and demonstrate how one numerically may extract effective parameters governing large scale behaviors, we quantify

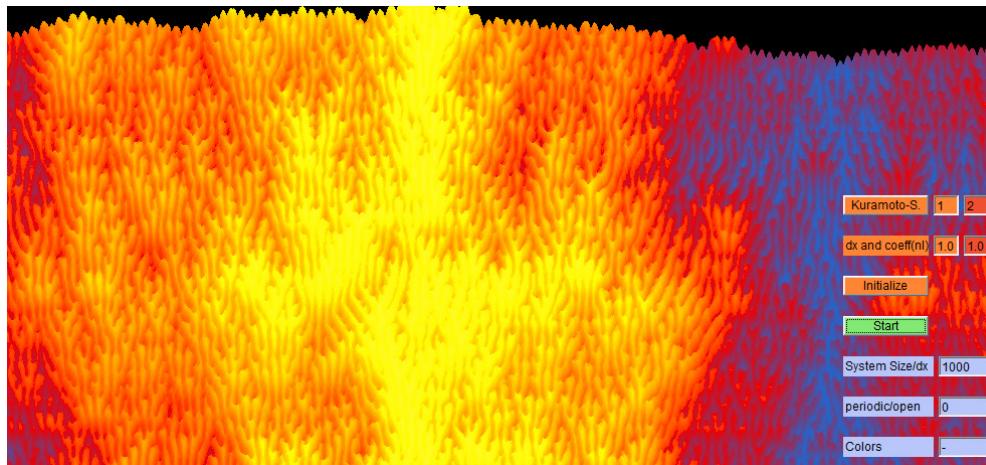


Figure 8.11: Artistically colored propagation of dynamics of the Kuramoto-Shivashinsky equation. Notice that on scales larger than the $L = 1000$ system shown here it starts behaving as the KPZ equation.

the relation between KS and KPZ by a numerical analysis, where the KS is simulated as described in Question 7.9 using a spatial discretization with unit $dx = 1$ (and $dt = 0.1$).

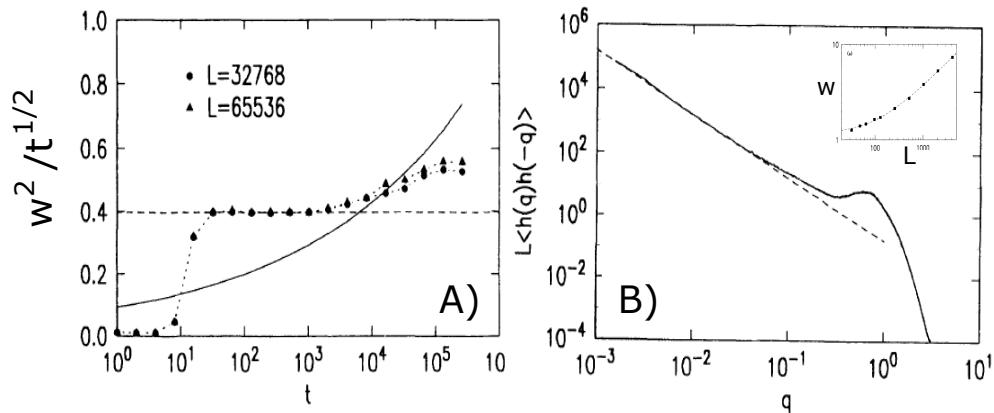


Figure 8.12: Dynamics of the Kuramoto-Sivashinsky (KS) Equation. A) Time development of variance (width squared) of an interface driven by the KS equation. The width is normalized by the expectation of the Edwards-Wilkins equation, and one notices that the non-linearity first sets in after 1,000 updates. B) Fourier spectra of the KS equation obtained for an $L = 4,096$ system in steady state. The dashed line indicates the scaling $1/q^2$ expected for an interface that looks like a random walk when going along it (both KPZ and Edwards Wilkenson "look" like that). The peak at $q = 1/\sqrt{2}$ corresponds to the maximal growth of the linear term in the KS equation ($q^2(1 - q^2)$). The inset shows how the width actually changes when one considers different systems sizes.

Assume now that the KS equation on some large scale behaves like

$$\frac{\partial h}{\partial t} = \nu \frac{\partial^2 h}{\partial x^2} + \lambda \left(\frac{\partial h}{\partial x} \right)^2 + \eta. \quad (8.52)$$

Then the coefficients can be determined numerically by measuring three quantities: the overall width of the saturated interface ($W(L) = (L\Gamma/2\pi^2\nu)^{1/2}$), the increase in width as a function of time for small times ($W(t) = \Gamma^{1/2}(t/2\pi\nu)^{1/4}$) and finally the velocity dependence $\langle dh/dt \rangle$ as a function of an overall tilt of the interface.

The result of these numerical measurement was (Sneppen et al. PRE 1992) that $\nu = 10.5$, $\Gamma = 3.2$ and $\lambda = 2.3$. The large value of ν can be understood by noting that it has the dimensions of a length squared divided by a time scale (see earlier). The estimated $\lambda = 2.3$ is a result of a faulty discretization. In reality, with fine enough discretization $\Delta x < 0.6$ instead of $\Delta x = 1$, one obtains $\lambda = 1$ as expected, because the overall advance of the interface in any case should be given only by the non-linear term.

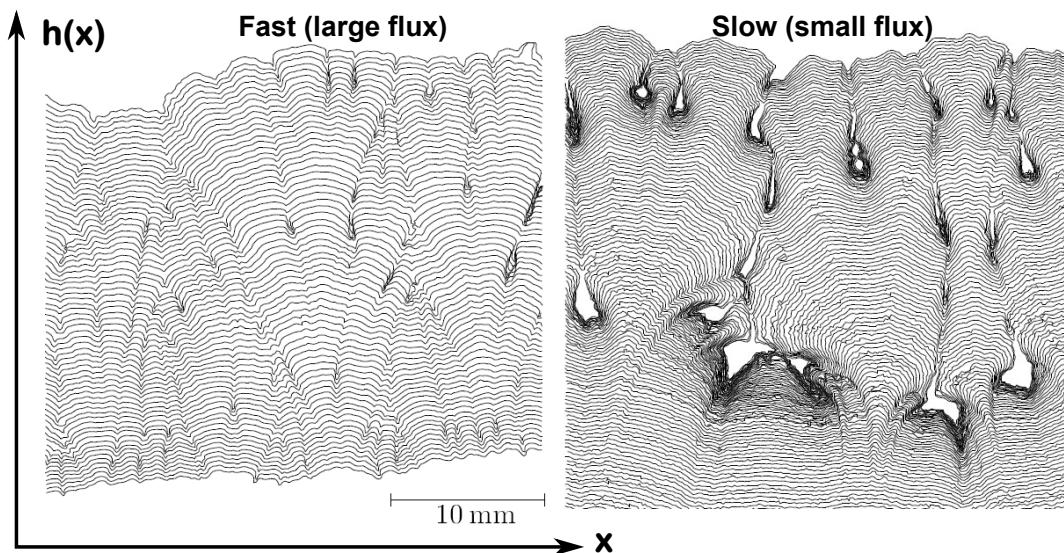


Figure 8.13: Kuramoto-Shivashinsky-like behaviour in an actual experiment. The dynamics was modified to allow the front to develop overhangs. The pictures show experimentally measured growth of a yeast colony, where each interface corresponds to the edge of colony at a particular time. Subsequent lines are separated by one hour. For the fast growth case, the colony is growing on a 1 mm thick agar layer that separates it from a replenishing liquid. For slow growth the growing yeast is separated from by 2.6 mm from this continuous food supply and waste product removal (T. Sams et al, PRL 1997).

In conclusion, the temporal roughening of the Kuramoto-Sivashinsky equation exhibits four different regimes: Initially an exponential growth to an intrinsic width dominated by the two-wave expansion. In second regime the

roughening develops according to the linear EW equation, $w \propto t^{1/4}$. Then at a (much) later time and for sufficiently large systems, one observes a cross over to a faster roughening where $w \propto t^{1/3}$. Finally, for large systems, at very late times, a saturated stage is reached, where the width fluctuates around a statistically stationary value.

Questions:

7.11) Solve the KS by the simplest possible discretization [?]:

$$dh = (\mathcal{L} + \mathcal{N})dt, \quad (8.53)$$

where

$$\mathcal{L} = -[h(x-2) - 3h(x-1) + 4h(x) - 3h(x+1) + h(x+2)] \quad (8.54)$$

and

$$\mathcal{N} = [(h(x+1) - h(x-1))/2]^2 \quad (8.55)$$

with $h(x) = h(x, t)$ and $h(x, t + dt) = h(x, t) + dh$. For time step $dt = 0.1$ the above equation can be iterated ad infinity. Use a lattice of $L = 100$ with periodic boundary conditions and initial condition $h(x) = \sin(2\pi \cdot x/100)$.

Qlesson: A simple integration scheme is giving basic phenomenology. Its not truly exact, in particular the choice of lattice spacing 1 tends to give dynamics that resembles a more true scheme with a factor 2.3 large non-linear term

7.12) Simulate variations of the main idea in KS using the simple integration scheme from above:

$$\begin{aligned} \frac{\partial h}{\partial t} &= -\frac{\partial^2 h}{\partial x^2} - \frac{\partial^4 h}{\partial x^4} + \left(\frac{\partial h}{\partial x}\right)^2 \\ \frac{\partial h}{\partial t} &= -\frac{\partial^2 h}{\partial x^2} - \frac{\partial^4 h}{\partial x^4} + \left(\frac{\partial h}{\partial x}\right)^4 \\ \frac{\partial h}{\partial t} &= \frac{\partial^4 h}{\partial x^4} + \frac{\partial^6 h}{\partial x^6} + \left(\frac{\partial h}{\partial x}\right)^2 \end{aligned}$$

starting with $h(x, t = 0) = \sin(x\pi/50)$ in a $L = 300$ system with periodic boundary conditions.

Qlesson: Some of the above behave similar to the KS equation, with the exception of the case with $(dh/dx)^4$ which tends to freeze the interface around the minima. Thus, higher order terms in the KS equation can change the large scale dynamics, in contrast to what we would have guessed from normal scaling arguments.

7.13) Redo the linear stability analysis for the KS equation. Describe how the non-linear term stabilizes the equation.

Qlesson:... connect unstable and stable modes,... secure transfer to cusps where the high-order term can dissipate gradients and curvatures.

7.14) If h fulfills the Kuramoto-Sivashinsky equation, what is the equation for its derivative: $\theta = \partial h / \partial x$. Which symmetries differ between the original equation and the equation for its derivative? What is the roughness exponent χ for $\theta(x)$.

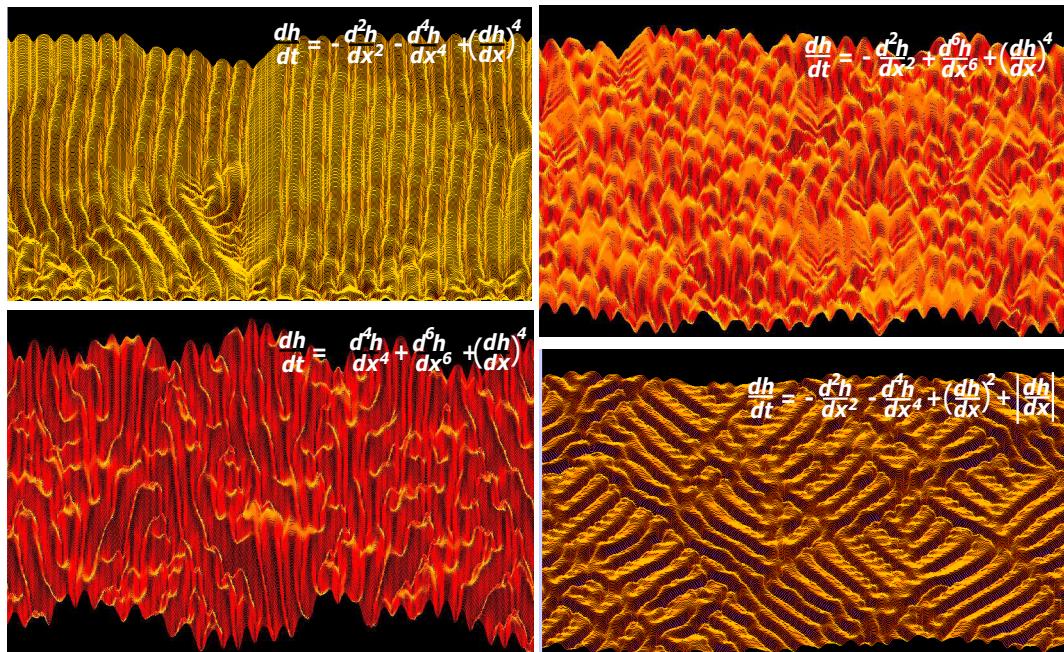


Figure 8.14: **Playing with models inspired by the KS-equation.** All simulations are performed by first order iteration, using a grid $dx = 1$ and a sufficiently small time step to avoid instabilities.

Qlesson:... KS is left-right symmetric, the equation for the derivative is ...

7.15) Calculate the cross over time from EW growth to KPZ growth for the KS equation (eventually see [?]).

8.4 Self-Organized Critical Interfaces

Experimentally determined χ for 1-dimensional interfaces ranges from $\chi = 0.55 \pm 0.06$ in electrochemical deposition (Galathara et. al. *PRL* **68** 3741) $\chi = 0.78 \pm 0.07$ in growth of bacterial colonies (Vicsek et al. *Physica A* **167** 315) to a χ of 0.73 ± 0.03 (Rubio et al. *PRL* **63** 1685) and ≈ 0.81 (Horváth et al. *PRL* **65** 1388) measured for fluid invasion of porous media. In all experiments the measured χ are above the one predicted by the KPZ universality class.

It is known that noise plays a big role in the growth of an interface. Parisi suggested to study interfaces with quenched noise η

$$\frac{\partial h}{\partial t} = \nu \Delta h + \frac{\lambda}{2} \left(\frac{\partial h}{\partial x} \right)^2 + \eta(x, h) + F_c, \quad (8.56)$$

where $\langle \eta(x, h) \eta(x', h') \rangle_{ensemble} = \Gamma \delta(x - x') \delta(h - h')$. This form of the noise is physically appealing because, e. g., a liquid penetrating a porous medium is influenced by a local resistance which does not change explicitly with time, but only with location (x, h) . This quenched noise has the property that if

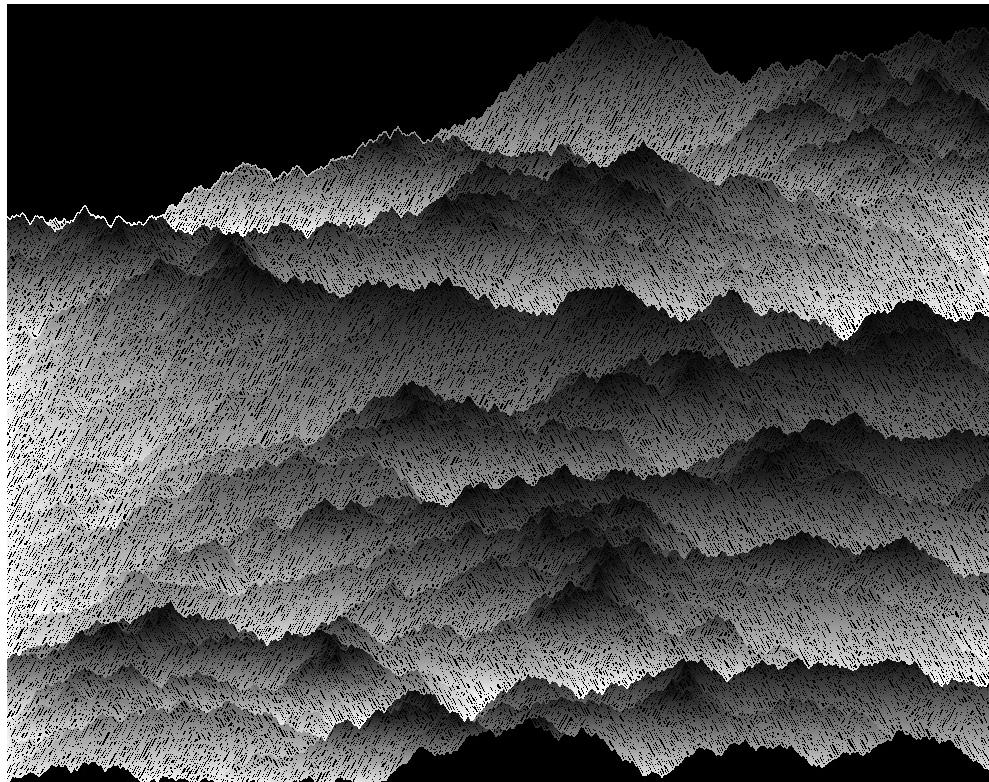


Figure 8.15: Interface driven by extremal dynamics to the critical point of directed percolation.

the interface has zero velocity, it typically also has zero velocity next time (except when neighbors have moved). Thus it is easily pinned. For a suitably fine-tuned pinning force $F = F_c$, simulations of the above equation give more dramatic roughening, reflecting that part of the interface gets stuck in regions with large resistance.

There is however another approach, in which the pinning force F does not need to be fine-tuned, but where the growth at a given time occurs at the site where the pinning force is the global minimum. Thus we effectively want to simulate the quenched noise growth equation of Parisi by subsequently updating only one site at each step: the site where the right hand side of the equation is minimal. A one-dimensional lattice model for such dynamics is the “Sneppen model” (Sneppen, *PRL* **69**, 3539):

Define a discrete interface $h(x)$ on a chain $x = 1, 2, 3, \dots, L$, and a string of Gaussian-distributed random uncorrelated local pinning forces $\eta(x, h)$. The chain is updated by finding the site with the smallest pinning force $\eta(x, h)$ among all sites. On this site one unit is added to h . Then neighbouring sites are sequentially ($y = x + 1, x + 2, \dots$ and $y = x - 1, x - 2, \dots$) adjusted upwards ($h \rightarrow h + 1$) precisely until all slopes $|h(y) - h(y-1)| \leq 1$. New random noise $\eta \in [0, 1]$ is assigned to all adjusted sites.

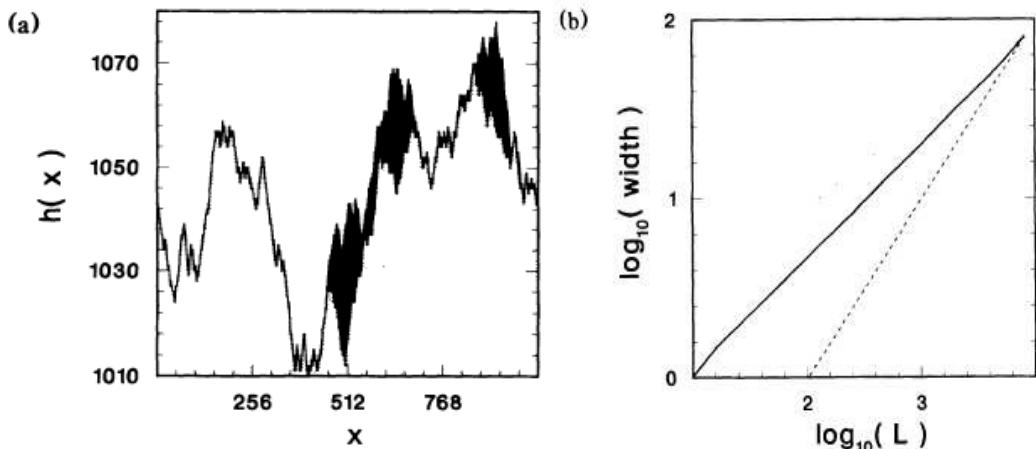


Figure 8.16: Simulated interface dynamics. a) Snapshot of two configurations of the interface generated by model. The black area indicates the advance of the interface. Notice that activity appears highly irregular, in "bursts". Please notice that the Y-axis has a different scale than the X-axis. b) Scaling of width of interface as function of length that we investigate. The dashed line has slope 1, whereas the actual interface width scales as length to exponent 0.63.

Numerical simulations at the statistically stationary state give a roughness exponent $\chi = 0.63 \pm 0.02$, an exponent that is in fact given by the ratio of two exponents in directed percolation ($\chi = \nu_{\perp}/\nu_{\parallel}$). The connection to directed percolation comes from the model's tendency to persistently remove small η , and only leave stretches of large η that percolate along the interface. When there is a connected path of such $\eta > 1 - p_c$ for directed percolation, this path can only be broken by selecting an η that is just at p_c .

It is stressed that there are limitations to the applicability of the above model:

First of all it simplifies the global equilibrium by assuming that the chance to advance only depends on pinning forces, but is completely independent of the overall shape of the interface. In this way the model will not apply to situations where long range instabilities caused by screening can occur (e.g the Mullins Serkerka instability [*J. Appl. Phys.* **34** 323 (1963); **35** 444]). Rather, the model may be applicable in situations where we push a high viscous liquid into a liquid of low viscosity.

Second, as the model assume that pinning forces are compared globally before movement takes place, the model is valid only for experimental situations, where the time for deciding which site to advance next is much smaller than the time of the actual movement. If, reversely, one can only compare over a small restricted region of neighbors, then on large scales the behavior will be governed by local rules.

Third, as is the case with the KPZ dynamics, the model assumes an explicit breaking of rotational symmetries in order to obtain anything self affine.

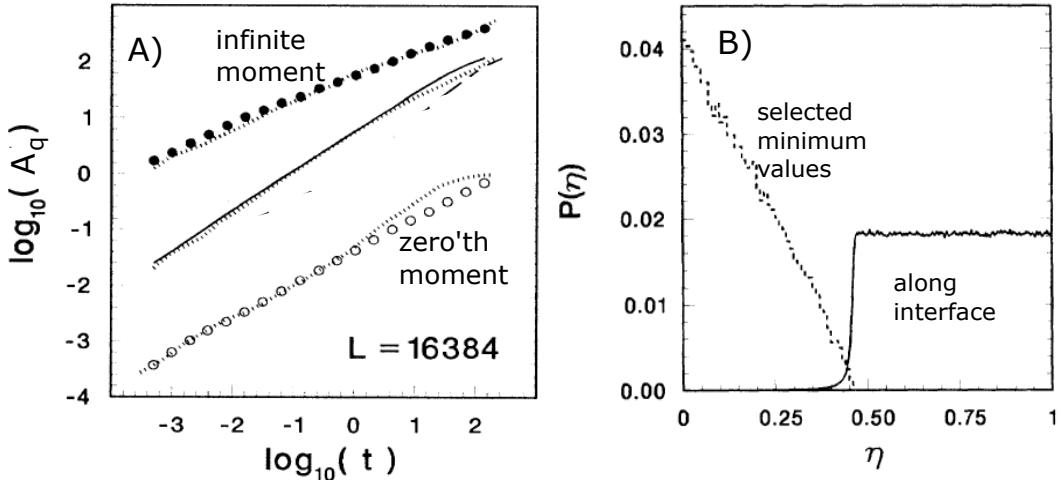


Figure 8.17: **Exemplifying interface dynamics.** A) Dynamics of interface analyzed in steady state, quantified in terms of A_q with infinite moment measuring the maximal advance at any point as function of some elapsed time at steady state, whereas zeroth moment measures the number of sites that have changed value. The solid line in the middle measures the second moment, corresponding to the spread of the advance along the interface. B) Values of noise terms along interface, and of selected active sites, sampled over long time.

From fig. 8.16 one sees that activity is localized, or what we will call burst-like. To quantify the bursts, consider the scaling of the accumulated activity (ignoring the insignificant average displacement):

$$A_q(t_0, t) = \langle (h(x, t_0 + t) - h(x, t_0))^q \rangle_{ensemble,x}^{1/q} \quad (8.57)$$

which at saturation may be ensemble averaged by averaging over different initial times t_0 . In the above equation A_0 defines the spatial spreading of activity with time and A_∞ measures the recurrence of activity at a given time. The scaling of A_q is shown in Fig. 8.17.

The interface advances from one interface to the other, which both are characterized by the same roughness exponent χ leads to a simple expression for the scaling of A_q :

$$A_q(t) = \langle (h(x, t + t') - h(x, t'))^q \rangle^{1/q} = (r_\perp^q \cdot r_\parallel + 0 \cdot (L - r_\parallel))^{1/q} \quad (8.58)$$

where $r_\parallel(t)$ denotes the number of sites that have been active until time t and r_\perp measures the height of the ‘‘bump’’. $L - r_\parallel$ is the number of sites which have not changed (alas $r_\perp = 0$). As $r_\perp \propto r_\parallel^\chi$ and as $r_\parallel r_\perp \propto t$, where time t counts number of updates, then:

$$A_q(t) \propto t^{\beta_q}, \quad \text{where } \beta_q = \frac{q\chi + 1}{q\chi + q}, \quad \text{and } \beta_0 = \frac{1}{1 + \chi}, \quad (8.59)$$

derived by Olami et. al. in PRE **49** (1994) 1232. The equation for β_0 is derived separately by noting that the zeroth moment simply counts the number of

activated sites. Notice that $\beta_0 + \beta_\infty = 1$ and that $\beta_\infty/\beta_0 = \chi$ whereas the exponent $1/\beta_0 = z$ plays the role of the dynamic exponent z . The term $1 + \chi$ counts the mass of the burst as measured along the system(interface), and is the dimension D of the burst.

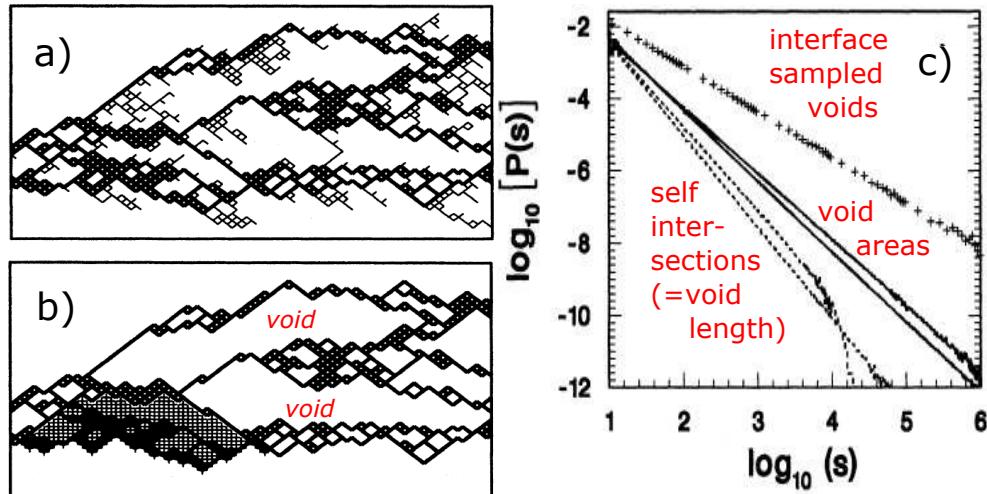


Figure 8.18: **Exemplifying interface dynamics.** a,b) Snapshot of moving interface, while highlighting backbone of lattice points with resistance η larger than the critical threshold number. c) distribution of voids in backbone of DP-avalanches in model.

Although we have obtained the exponents χ and β , as we did for the KPZ equation, the model is much richer than that. One signal of this is seen in figure 8.17B), where we sample the sizes of η values along the interface. One sees that these η 's are nearly always above some critical threshold value η_c . $1 - \eta_c$ in fact turn out to be the critical percolation threshold for blocking the propagation of the one-dimensional interface. This is in close analogy to the evolution model from chapter 2. The model is yet another example of self-organized criticality.

Questions:

- 7.16)** Write a computer program to simulate the Self-organized critical interface model. Estimate the roughness exponent of the interface in steady state.

Qlesson: The interface is rougher than a random walk.

- 7.17)** Formulate a variant of the above SOC model where, instead of moving neighbor sites such that all gradients are less than or equal to unity, one adjusts neighboring sites such that all curvatures are less than or equal to unity. Simulate this model. **Qlesson: The interface is very very rough, and in fact the total width of the interface grows more than linearly with system size (see also Fig 7.1).**

- 7.18)** If the signal $q(t)$ exhibits fractional Brownian motion with exponent H , show that the power spectrum $S(f) \propto f^{-2H-1}$. Noisy signals with $H = 1/2$ are called Brown noise.

Qlesson: $H = 1/2$ gives $S(f) = 1/f^2$ whereas $H = 0$ gives $S(f) = 1/f$.

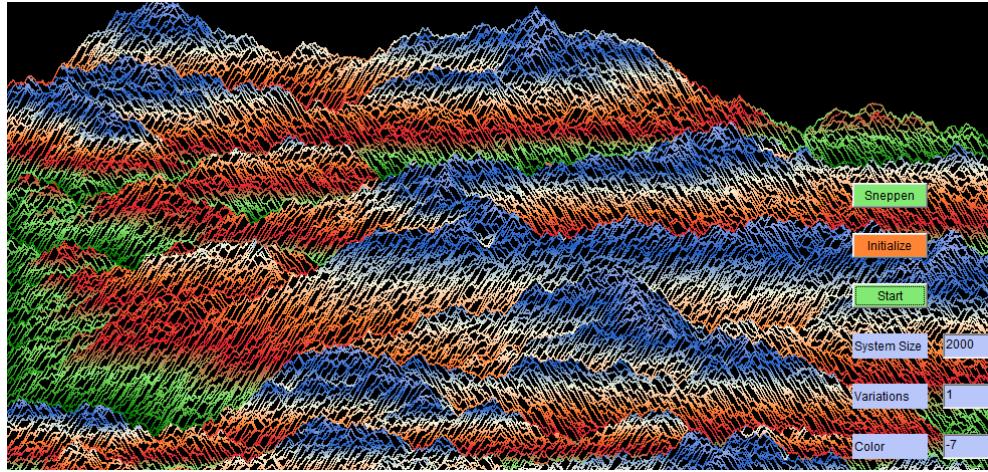


Figure 8.19: Artistically colored propagation of a stochastic version of the “Sneppen model”.

7.19) Consider the above SOC interface model. Monitor the value of the selected minimal $\eta(x)$ at each time, and observe that these minimal values never becomes larger than a critical value η_c , see Fig. 8.17. Avalanches can then be defined a number of updates between times where selected η is above a threshold $\eta_t \sim \eta_c$. Measure this avalanche size distribution numerically.

Argue that:

$$\begin{aligned} P(s) &\propto \frac{1}{s^\tau} \\ \tau &= 1 + \frac{1}{1+\chi} \left(1 - \frac{1}{\nu_\parallel}\right) = 1 + \frac{\nu_\parallel - 1}{\nu_\perp - 1} = 1.259 \end{aligned} \quad (8.60)$$

in terms of exponents from directed percolation ($\chi = \nu_\perp/\nu_\parallel$). (from Maslov & Paczuski, PRE 50, R643). This equation is associated to the correlation length $\ell \propto \epsilon^{-\nu_\parallel}$ that defines the length between points that are within ϵ from the critical $p = p_c$ for directed percolation. These points are punctuated by the interface dynamics when all η 's along the interface are above $p_c - \epsilon$. The relation between these avalanches and the properties are visualized in Fig. 8.18. [Qlesson: the critical thresholds \$\eta_c\$, above which \$\eta\$ value does not matter...](#)

Lessons:

- Interfaces easily look like random walks, but the individual parts make anti-persistent walks around their mean drift.
- In theory, the KPZ equation should be universally correct for a large class of interface like systems.
- The non-linear term in the KPZ equation transfer noise to lateral growth and dominate the dynamics of the interface.

Supplementary reading:

Halpin-Healy, Timothy, and Yi-Cheng Zhang. "Kinetic roughening phenomena, stochastic growth, directed polymers and all that. Aspects of multidisciplinary statistical mechanics." Physics reports 254.4-6 (1995): 215-414.

Bohr, Tomas, et al. Dynamical systems approach to turbulence. Cambridge University Press, 2005.

Chapter 9

Diversity

“survival of the fittest”

Herbert Spencer (summarizing C. Darwin’s work)

9.1 The living world is rich in diversity

Life presents an astounding diversity of discrete states or species that coexist with each other over long time intervals. At the sub-cellular scale, molecular competition and positive feedback maintain cells in specialized epigenetic states, allowing for embryonic development of complex multicellular organisms. On larger length scales, stable yet dynamic ecosystems emerge from competition between different species. Ecosystems with many species, in spite of the idea of survival of the fittest, that rather should favour a single unique super species as a global winner.

Diversity calls for an understanding of discrete states, including how they sustain themselves and how they interact with each other. Statistical mechanics of complex systems provides a framework for studying universality of collective and cooperative phenomena, usually through repeated action of identical agents. This methodology can be extended by allowing these agents to take diverse types and analyze biological phenomena in the perspective of effective rules of interaction.

9.2 Balance versus Dominance

Consider a system of D active species that compete across a common intermediate species (IS) [161]. The model with $D = 3$ is shown in Fig. 9.1, and is defined by many sequential updates between N agents that each can be in one of $D+1$ states. At each time step the system is updated by a recruitment step (r) and a noise step n :

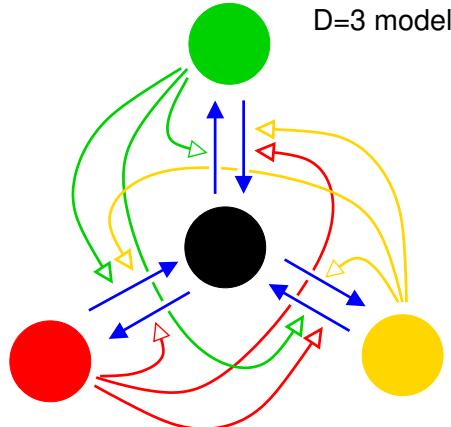


Figure 9.1: Model with $D = 3$ active states/species that attack each other across an intermediate inactive state (IS).

- r) At each time step a random site $j \in \{1, 2 \dots N\}$ is selected. If it is in one of the active species it attempt a conversion of another randomly selected site k : If k is one of j 's antagonistic species, it is reset to the IS. If k is IS, then it is converted to the same state as j . If state of k is equal the state of j , then no change is made.
- n) Select a random site l with probability β and change its state: If l is active it becomes IS, and if l is IS it becomes one of the randomly chosen active species.

The role of the parameter β is to parametrize all the behaviour that is not associated to the directed interactive step (r). It may for example represent noise events due to an outside of the considered system. Although the system we simulate is not in equilibrium and does not have an entropy, then β is somewhat similar to a temperature, with high β implying that the effective interactions between the agents plays a smaller role.

Multistability and the Mixed metastable state: Figure 9.2 shows trajectories for the $D = 3$ model. Notice a transition from the mixed state (MS) at $\beta > 0.31$ to a tri-stable case with alternating dominance state (DS) of one of the 3 active species at $\beta < 0.28$. Remarkably, however, for $\beta \sim 0.295$ the system can be trapped in the meta-stable mixed state (MMS). This is a peculiar behaviour, which is also seen in the equilibrium Potts model, that share the multi-state options with the current model.

The deterministic version of the $D + 1$ model is given by changes in respec-

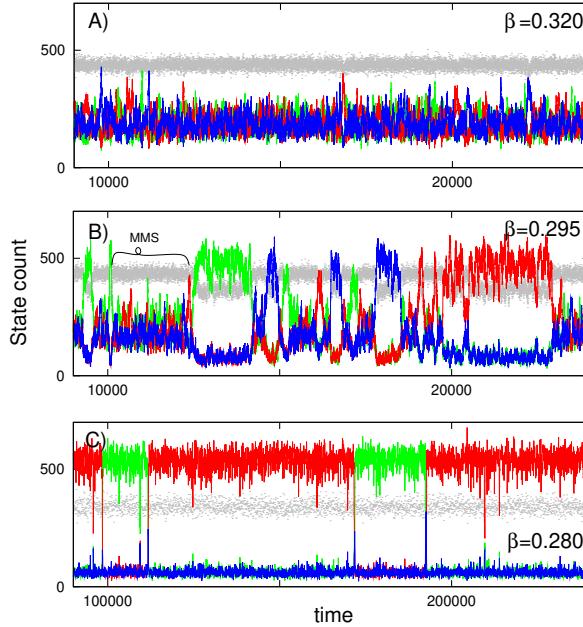


Figure 9.2: *Dynamical simulation of the $D = 3$ model with $N = 1000$. The grey dots referring to the IS. Time is counted in number of attempted pair interactions per site. Panel B) show occasional appearance of the MMS. Panel C) examines a 10 times longer times than A) and B), showing that stability depend on β .*

tive the active species n_j and the intermediate state (n_0):

$$\begin{aligned} \frac{dn_j}{dt} &= n_j n_0 - n_j \cdot \sum_{k>0, k \neq j} n_k - \beta n_j + \beta \frac{n_0}{D} \\ &= n_j(2n_0 - 1) + n_j^2 - \beta n_j + \beta \frac{n_0}{D} \end{aligned} \quad (9.1)$$

$$\frac{dn_0}{dt} = (1 - n_0 + \beta)(1 - 2n_0) - \sum_{j>0} n_j^2 \quad (9.2)$$

with $\sum_{j>0} n_j = 1 - n_0$, where n_j ($j = 1, 2, \dots, D$) denotes the fractional occupation of the active species. The coupling between the active species occur through depletion of n_0 . Notice that eq.(9.2) is redundant if we express n_0 in terms of n_j with $j > 0$ using $\sum_{j=0}^D n_j = 1$.

Figures 9.3 A-B) show trajectories and fixed points for the $D = 10$ model. We see that the model opens for a DS state at low β , whereas it gives monostability of of a mixed state (MS) at high β .

Fixed point analysis: For $D \geq 3$, the metastable states are always either a MS or a state where one species dominates (DS). In such a case, eqs. (9.1)-(9.2) can be solved analytically for the fixed points as well as their linear stability. As this is complicated we refer to question section, and here proceed to a simple and solvable variant model.

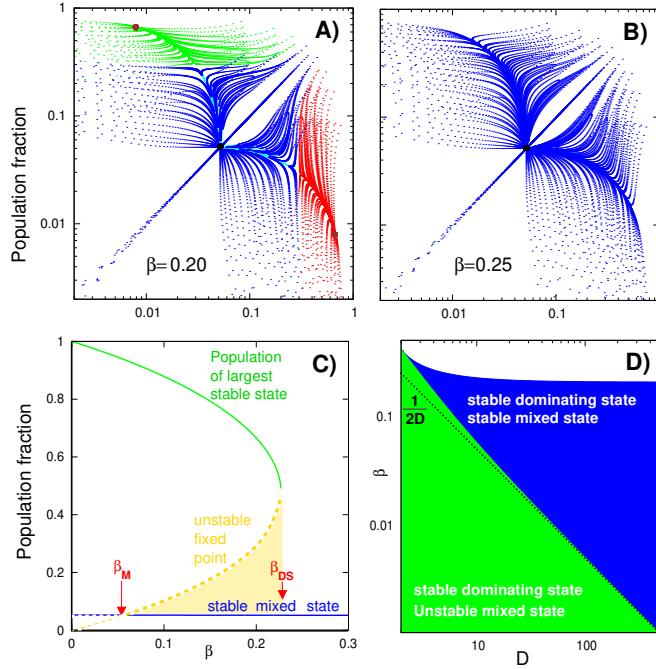


Figure 9.3: Deterministic dynamics for $D = 10$ model. A,B) Trajectories starting from uniformly distributed n_0, n_1, n_2 with all other n_i equal, shown for $\beta = 0.20$ in A) and $\beta = 0.25$ in B). Trajectories are colored after their final fixed point. Fixed points are shown with solid red and black dots. Initial conditions fix $n_i, i \geq 3$ to be equal and thus exclude trajectories ending in DS with $i \geq 3$. C) Population size of two of the stable states as function of β for $D = 10$. The dashed “gold” colored curve is the unstable fixed point that separate the MMS and the DS. D) The parameter region with MSS (blue region) and DS (green+blue region).

Variant (and easier solvable) model: Qualitatively, all of the above behavior can also be obtained in a variant of the model where there is no intermediate state (no IS), but where the active modifications instead consist of selecting two sites i and j , and only if they are the same species $S_i = S_j = S$ one recruit a random site k and set $S_k = S$ [109, 162]. The noise term is here most easily formulated by taking a random agent, and set it to a random state (which can be its original state with probability $1/D$). The deterministic counterpart of this model is given by the equation:

$$\begin{aligned} \frac{dn_i}{dt} &= n_i^2(1 - n_i) - n_i \sum_{j \neq i} n_j^2 + \beta \left(\frac{1}{D} - n_i \right) \\ &= \sum_{j=1}^D \left(n_i n_j - \frac{\beta}{D} \right) (n_i - n_j) = F_i(\vec{n}) \end{aligned} \quad (9.3)$$

The second step in this equation uses that $\sum_j n_i^2 n_j = \sum_i n_i^2$ and that $\sum_j n_i = n_i \cdot D$. Please convince yourself that the above equations indeed represent the proposed model. As we will now show, then this model supports a DS for $\beta < \frac{D}{4(D-1)}$ and a MMS in for $\beta \in [\frac{1}{D}, \frac{D}{4(D-1)}]$.

First of all, remember the overall conservation

$$\sum_{j=1}^D n_j = 1. \quad (9.4)$$

which was in fact used to obtain the above equation.

The fixed points in the form

$$n_1 = a, \quad n_2 = \dots = n_D = b \quad (9.5)$$

should satisfy

$$a + (D - 1) \cdot b = 1, \quad ab = \beta/D. \quad (9.6)$$

where the first equation comes from the overall conservation, an the second equation from setting $da/dt = 0$ and investigating solutions where $a \neq b$) Inserting $b = \beta/(D \cdot a)$ into the first equation we get:

$$a = \frac{1 \pm \sqrt{1 - 4\beta(1 - 1/D)}}{2}, \quad b = \frac{\beta}{a \cdot D}. \quad (9.7)$$

we obtain real positive solutions for a provided that

$$\beta \leq \frac{1}{4 \cdot (1 - 1/D)}, \quad (9.8)$$

Thus for these small β values one obtain solutions with one dominant solution. For larger β there simply does not exist a solution with one dominant species. The noise will drive all opinions to become equally present.

Note that one trivially, always have the uniform solution:

$$n_1 = \dots = n_D = \frac{1}{D}. \quad (9.9)$$

where all populations are equal.

In Fig. 9.4 we plot $a(D)$ from eq. 9.7 and compare with the uniform solution (grey in the figure). We see that the minus branch of the solutions in eq. 9.7 is unstable and will collide with the uniform solution at

$$\beta = \frac{1}{D}, \quad (9.10)$$

below which the uniform solution becomes unstable. For $1/D \leq \beta \leq \frac{1}{4 \cdot (1 - 1/D)}$ there is a meta-stable solution where all species are equally populated. Intuitively the existence of the metastable uniform solution (MMS) relies on the gap between an opinion having $1/D$ followers, and the need to obtain a majority, i.e. more than half the possible followers.

Real world interpretations: For $\beta = 0$, $D = 2$ and reduced cooperativity ($n^2 \rightarrow n^{1.3}$) eq.(9.3) simplifies to the language competition model of [163]

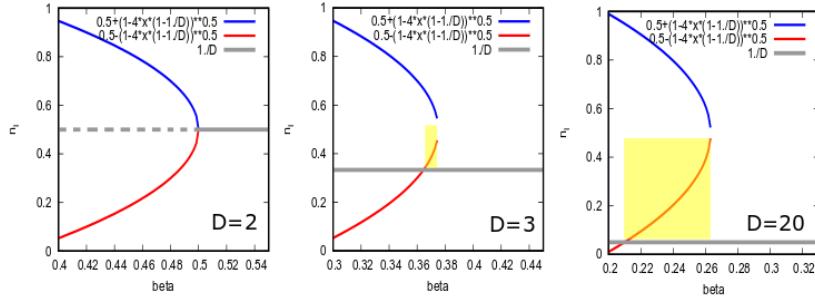


Figure 9.4: Fixed points for multi-state model as function of "noise" parameter β for 3 different diversities. For $D = 2$ the uniform solution (grey) becomes unstable at $\beta < 0.5$, and the system chooses one of two metastable states (both with one dominating species). For $D > 2$ the red branch becomes unstable for some intermediate β values and the system then have to chose between the high diversity state (grey), or a state dominated by one state (blue). For lower β values, it is again the extreme blue branch that becomes stable (and the red then correspond to the density of the $D - 1$ minority species). As D is increased, there is a larger range of β where the high diversity state is metastable (yellow region in the above figures).

that was used to describe global decay of language diversity during the last century. Our analysis shows that appropriate noise or spontaneous fluctuation also allows eq.(9.3) to describe metastable coexistence of multiple languages.

The model could also be considered in the context of a society with antagonistic political fractions. In this interpretation, the MMS could correspond to a representative democracy with many balanced interest groups, whereas the extreme states correspond to one-party systems.

Noticeably, a model of dynamically evolving networks proposed by ref. [164] shows similar multistability as found here. In that model, agents in the form of nodes on a network can each have one of D opinions. These opinions are updated by voter dynamics on the network (repeated dynamics where one random agent convince a neighbor agent about his opinion). In addition the networks is sometime rewired by a dynamics where agents accept connections to other agents provided that they have equal opinions. When links were removed randomly and agents with no connections randomly was assigned new opinions, the model obtained a multi-stable system where the MMS sometimes persists for some periods and may be associated to a network of disconnected agents and small clusters of agents. In this interpretation, the MMS correspond to a dissolved society where nobody communicates and everybody takes random opinions.

Questions:

- 8.1)** Simulate the $D=10$ model for a $N=100$ system and find a β value where the mixed meta-stable state can exist for some time. Hint: assign each of the 100 sites on a 1-d line one of the D values. At each time take two sites, and if they are equal take a third site and set its number (identity) equal to that on the two first sites. Subsequently take a random site and sets its identity to a random value among the D possible ones.

8.2) Fill out the intermediate steps in the Fixed point analysis in the text (alternatively make the fixed point analysis of the model in eq. 9.3).

8.3) Simulate a spatial variant of the above D=3 model model where conversion from IS is local and the general “kill” global. Keep noise n unchanged, and let the recruitment step r be: r) At each time step a random site j is selected and if it is in one of the active species then one of two moves is attempted: i) With probability $\frac{1}{2}$ a neighbor k to j is selected, and if k is in IS it is converted to species of j . ii) Else a random site k is selected, and if k is one of j ’s antagonistic species, it is changed to the IS. If selected j is not active or k is not fulfilling the corresponding requirements, then no change is made. Show dynamics for $D = 3$, $\beta = 0.098$ and $N = 100$.

8.4) Stability of uniform solution of the directly cooperative model. It is easy to calculate the stability of the uniform solution (9.9). Show that:

$$\frac{\partial F_i}{\partial n_i} = - \sum_{j \neq i}^{D-1} n_j^2 - \beta - \left(\sum_{j \neq i}^{D-1} n_j \right)^2 - 2(2n_i - 1) \sum_{j \neq i}^{D-1} n_j - 6n_i^2 + 6n_i - 1, \quad (9.11)$$

For $k \neq i$,

$$\frac{\partial F_i}{\partial n_k} = -n_i \cdot n_k - 2n_i^2 + 2n_i - 2n_i \sum_{j \neq i}^{D-1} n_j.$$

By substituting the uniform solution (9.9), one should get the Jacobian around this solution (all $n_i = 1/D$):

$$\begin{aligned} \frac{\partial F_i}{\partial n_i} &= D \left(\frac{1}{D^2} - \beta/D \right), \\ \frac{\partial F_i}{\partial n_k} &= 0. \end{aligned}$$

This matrix have only diagonal terms, and its stability is therefore determined by the sign of the (the diagonal terms are automatically the eigenvalues, and a stable solution needs (the real part of) all eigenvalues to be negative). Thus the uniform solution is stable for $\beta > 1/D$.

9.3 Emergence and Decline of Wrong Paradigms

Human history contains a number of epochs, 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.**

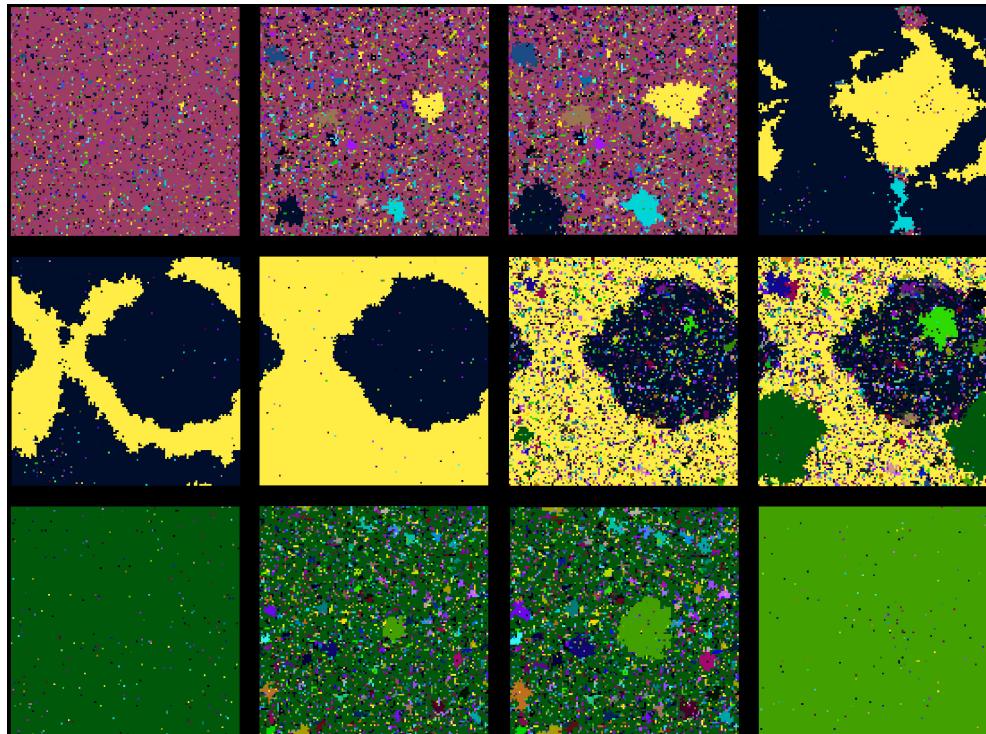


Figure 9.5: Model for idea spreading. 12 consecutive snapshots of a $N = 128 \times 128$ system with $\alpha = 25 \times 10^{-6}$. The time intervals between the pictures are not equidistant, as can be seen from the times given for each panel (times correspond to Fig. 2b). Time is measured in units of 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?**

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

$$\frac{dn}{dt} = \left(\frac{n}{N}\right)^2 \propto n^2 \quad (9.12)$$

$$n(t) = \frac{1}{t_c - t} \quad (9.13)$$

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. 9.5 shows 12 subsequent states of a system driven by the model. The snapshots reflect states of the simulation shown in Fig. 9.6b, 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

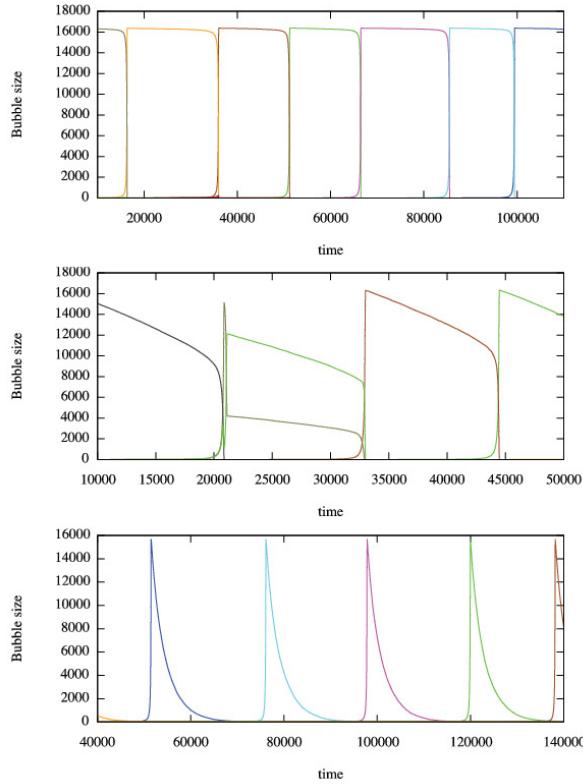


Figure 9.6: **Three time series of the sizes of the dominant states of the system.** At $\alpha = 0.4 \times 10^{-6}$, $\alpha = 25 \times 10^{-6}$, and $\alpha = 400 \times 10^{-6}$, respectively. Time is measured in units of sweeps (updates per agent). Notice that the length of each period does not change substantially with α , and in fact becomes more regular with larger α .

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 9.6 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. 9.6 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 to 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.

gent 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 9.7: Space influence biodiversity, as it prevent some predators from fully exploiting their prey. Here a picture from a bird colony in India. The birds build nests on islands, while their eggs are protected from land predators by alligators in the water.

9.4 Diversity from Cyclic Competition

Organisms cooperate or compete with each other and form complicated ecological systems that often sustain themselves over long time intervals. This stability is not easily understandable because exponential growth with competition tend to destabilize large coupled dynamical systems [165]. Competition may however be limited by spatial segregation [166, 167, 168, 169, 170, ?]. In particular, formation of spatially segregated regions may be facilitated by life itself. An example of is given in Fig. 9.7, where a population of one species

(aligators) form an environment for birds that is protected from other species (monkeys).

As a scenario for self-organized niche formation we now consider populations on a 2-dimensional surface. One may for example think about lichen growing on a surface of rock. The important assumptions are that the species are non-motile, that interactions are random, and that there is no particular hierarchy among the assigned strength of invasions between the different species. We will review this model by following the analysis in ref. [?, ?].

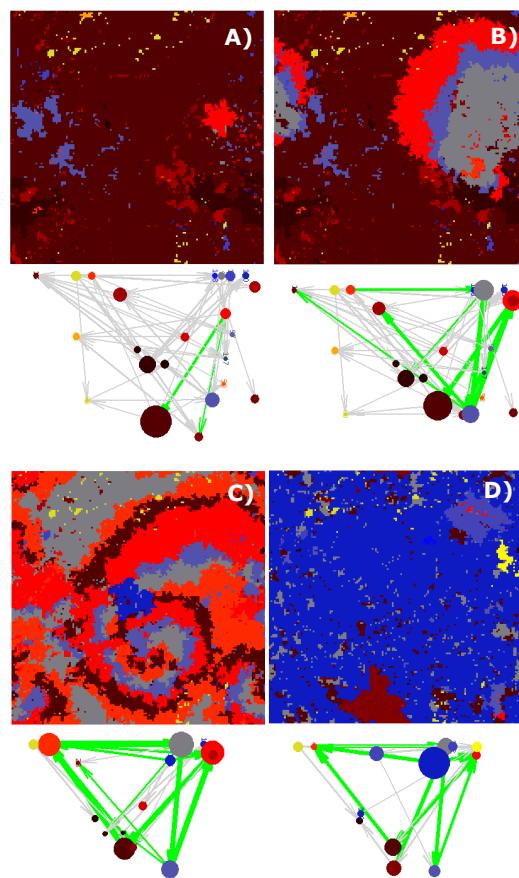


Figure 9.8: Simulation of model from ref. [170] with parameters $\gamma = 0.1$ and $\alpha = 0.1$ for an $N = 200 \times 200$ system. The small networks show contemporary interactions with the thickness of green arrows as indicators of frequency of active invasions. The grey arrows mark potential interactions that are inactive because the species are not in contact.

In the model each site i can at most be occupied by one individual. The model further allow new species to arrive or evolve with a rate α . The limit $\alpha \rightarrow 0$ is interesting, as it represents a clear time-scale separation between population dynamics and the evolutionary dynamics. In real ecosystems this evolutionary timescale could also represent the arrival frequency of new seeds from outside the system.

The one remaining model parameter is the probability γ that one species is able to invade another specific species. If $\gamma = 0.10$ a species i is assigned the property that it can invade species j with probability $\gamma = 0.10$. However the invasion can only take place when these two species are neighbors in the physical 2-dimensional space.

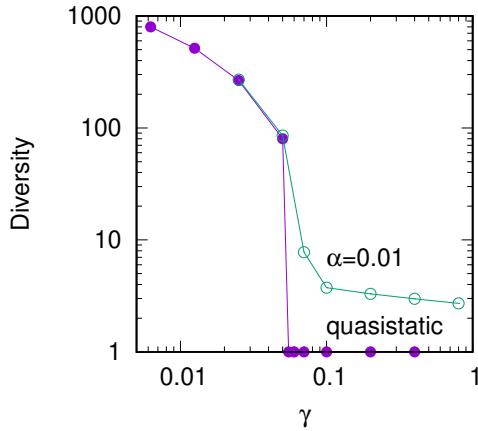


Figure 9.9: Diversity D as function of γ , for different values of α . Thick black curve reflect behavior for $\alpha \rightarrow 0$, a limit that can be obtained in a quasistatic simulation where species are only added when all dynamics are frozen. Figure from [170].

The model is implemented on a 2-dimensional square lattice with 4 neighbors per lattice location. It is defined in terms of discrete updates of single sites. At each update:

- Select a random site i and one of its nearest neighbors j . If the species $s(i)$ at site i can invade the species $s(j)$ on site j , i.e. $\Gamma(s(i), s(j)) = 1$, then the occupation on site j is updated: $s(j) = s(i)$.
- With probability α a new random species s is introduced at a random point j and assigned random interactions $\Gamma(s, u)$ and $\Gamma(u, s)$ with all existing species u in the system. Each of these interactions are assigned the value 1 with probability γ , or otherwise set = 0. If the introduced species s cannot invade the previous species at the site j , $s(j)$, $(\Gamma(s, s(j)) = 0)$ the insertion is aborted.

Notice that the interaction between any pair of species is fixed in their entire existence time. New interactions only appear with new species.

Fig. 9.8 illustrates how species emerge, spread and collapse (follow for example the fate of the red species). The different colors represent different species. Because most of these cannot invade each other, then parts of the lattice is frozen. In A) a new red species is invading the dominating brown species, which in turn lead to a boundary with the grey-blue species. This

grey-blue species then invade the brown in B). Sometimes cycles appear, for example the cycle of 5 subsequently invading species in C): *lightred* → *grey* → *greyblue* → *red* → *brown* → *lightred*. Notice in the small blue species in C) spreads unhindered, leaving many separated patches of the grey species. In general, when a cycle collapse, some of the surviving species will be fragmented into disconnected patches.

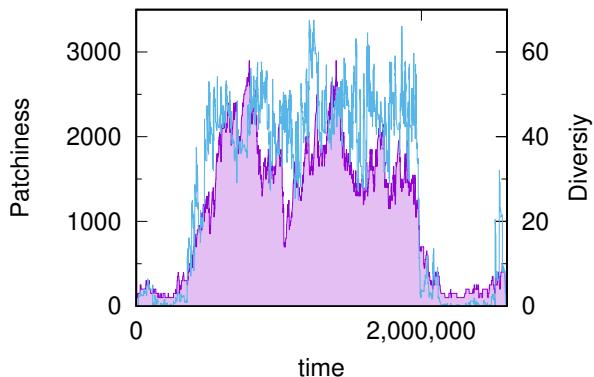


Figure 9.10: Patchiness→ Diversity in the model [?]. The time evolution of the patchiness (blue line) and diversity (pink shaded area) in a simulation with model parameter $\gamma = 0.07$ and system size $N = 200 \times 200$.

When the interaction density γ is decreased below a critical $\gamma = \gamma_c \approx 0.06$ the system is found to exhibit a sharp transition, see Fig. 9.9. That is undergo a first order like transition from a state that is dominated by one-species ($D \sim 1$) to a state with high diversity D ($D > 20$).

The model could easily be compared to its non-spatial counterpart where an individual on any site can attempt invasion of any other site. This “infinite-dimensional” model always collapse to very low diversity. In fact, independent of the value of γ this random neighbor variant always predict collapse to diversity $D = 1$ with $\alpha \rightarrow 0$.

In the 2-d version, the species diversity can only increase if an invading species cannot access the full population of its prey. This will occur when the prey is subdivided into spatially separated regions. Accordingly, habitat fragmentation into isolated patches is a prerequisite for speciation. Fig. 9.10 examines typical switching between low and high diversity states. The figure show an increase in the number of patches before the diversity increase. In contrast, when the diversity collapse, it does so together with the ”patchiness”.

Creation of patches require the spatial reshuffling events that occur when cycles are formed and subsequently break down. In fact cycles are essential for obtaining high diversity. This is demonstrated in Fig. 9.11 and ref. [171] by simulations where one artificially prohibit addition of a species when this would participate in any cycle that has a length which is shorter than seven.

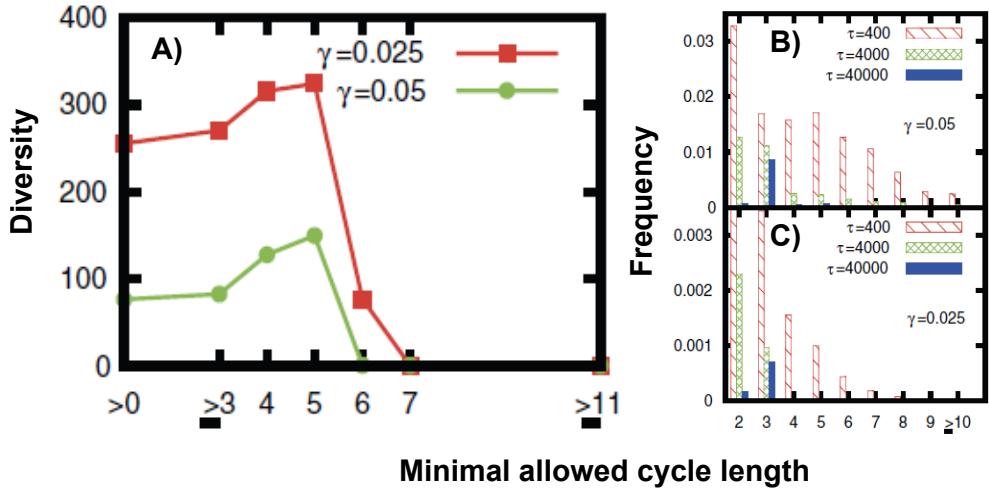


Figure 9.11: **Cycles → Patchiness**. A) Quasistatic simulation ($\alpha \rightarrow 0$) of $L=200$ system where species are only introduced when they do not form cycles shorter than specified on the x-axis. B,C) Show that frequency of eventual cycles is low, and that especially the long cycles only exist shortly. However, diversity can be obtained from cycles of length 5 in spite of their fast decay. Figure from [171].

In that case, one never obtain the high diversity state, irrespective of the value of γ .

Overall, the model provides co-existence of many species through the positive feedback in Fig. 9.12. This drive towards mere species is constantly exposed to an opposing force where patches merge when one expanding species encounter a new species that it can invade, or which can invade it. With such mergers a larger patch is created with more possibilities for subsequent mergers. This can potentially initiate an overcritical cascade of mergers forming ultimately one giant coherent occupied area. One observe that collapse of the high diversity state is rapid and characterized by passing a barrier in patch extension of about 5000 to 10000 lattice sites.

The co-existence of multiple species depends on both the limitations imposed by the interaction network (γ), and on spatial separation. If either are absent, the diversity of species collapses. “Fitness” in this model is truly context dependent, and it may collapse just as a species expand and thereby encounter a new predator.

Cycles, or non-hierarchical relationships, are observed in several ecosystems, from seaweeds and coral reefs to lichen on alpine rocks [174, 175, ?, 176]. In particular [175] argued that cyclic dependencies among coral reef species allowed for longer coexistence than hierarchical relationships. The model above takes this proposal further, emphasizing that even transient cycles can contributing hugely to sustainable long term diversity. The model is an idealization in the sense that it assume that non-interacting species form fixed boundaries, and in assuming that there are close to zero long range invasions. In regards to the motivating picture from Fig. 9.7, then diversity may further

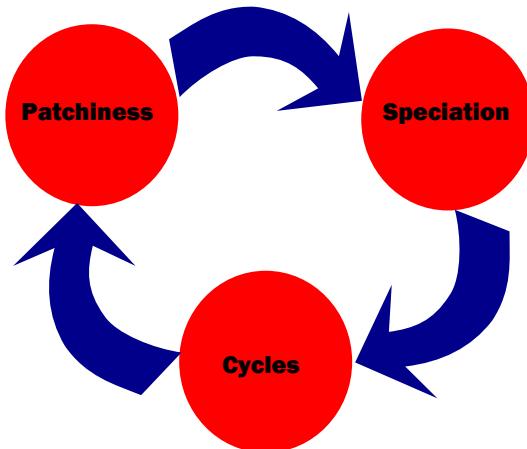


Figure 9.12: *Positive feedback for creation of diversity:* Patches of species create the heterogeneous environment that is needed for new speciation. Given separated patches, then both a new “mutant” and the original species can remain separated. Thereby diversity can increase in overall system. If many different species are present then some of them may form a cyclic relationship. When these cycles collapse the surviving species generate a number of spatially separated patches. This then close the overall cycle of events that support high diversity. Figure reprinted from ref. [172, 173].

be supported through a positive feedback that exploits the opportunities of a heterogeneous physical environment.

9.5 Collapse & Exponential growth

Unlimited exponential growth is not possible. Following [177] we here explore the interplay between

- Exponential growth
- Finite world
- Bad things happens

which are interconnected in the sense that exponential growth is only possible when there is available resources. These in turn could come from “bad things” that happened to “somebody else”.

Let $s(t)$ be the size of a single growing population in a growth-limiting environment. The time development is traditionally captured by Verhulst’s [178] logistic equation

$$ds/dt = s \cdot (1 - s) \quad (9.14)$$

with a carrying capacity of the environment that here is set to 1. This formalism can be extended to N different populations that are constrained by the

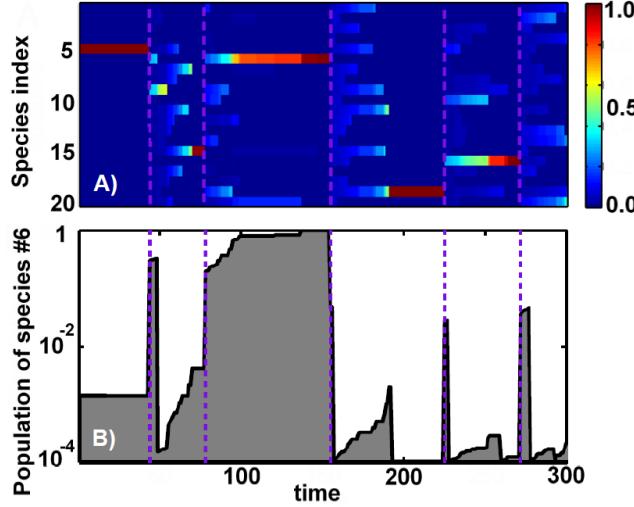


Figure 9.13: Dynamics of $N = 20$ populations and $\gamma = 10^{-4}$, reprinted with permission from [177]. (A) Time-courses of population sizes, with the color scale shown on the right. Note five diversity waves ending at purple dashed lines. (B) The time-course of population # 6 shown with logarithmic y-axis. Between time steps 100 and 150 this population nearly dominate. Its local extinction at the time step 154 ended the third diversity wave and started the fourth one. Note the erratic yet distinctly exponential growth of this population. The timescale of growth is set by the frequency of collapses of other populations ($= 1/N$).

same limited environment, and further are exposed to episodic independent collapse events.

For simplicity we will assume that all growth rates of individual populations are identical. And we will assume that they each are exposed to collapse with equal rate. A time-step in this collapse driven model for growth consists of:

- 1) Select a random population and replace it with a population of a size $\gamma \ll 1/N$.
- 2) Immediately after this collapse the freed-up resources leads to the transient exponential growth of all remaining s_i , until carrying capacity is reached. This is done by re-scaling all populations $s_i \rightarrow s_i(\text{new}) = s_i / \sum_j s_j$.

Figure 9.13A shows time-courses for the different populations in a system with $N = 20$ and $\gamma = 10^{-4}$. We in the following analyze the time development of this “collapse model” in some detail.

At certain periods then the entire system is nearly fully occupied by just one dominant population. A population that then have size $s_{max} \simeq 1$ as seen from Figure 9.13A. When such a dominant population collapses, then large resources becomes available and all smaller populations are increase by the large factor $1/(1 - s_{max})$.

The collapse of the last "big" population marks the end of one "wave", and the start of another. The start is characterized by many relatively equally sized populations and a high value of the size weighted diversity $D(t) = 1/(\sum_{i=1}^N s_i(t)^2) \sim N$. In the course of the wave then these populations are eliminated one-by-one by random collapses, causing D to decline (not shown here). Within each of these diversity waves, the ongoing collapses makes D declines as $\exp(-t/N)$. The typical duration, t_{wave} of such a wave is given by the number of events t_{wave} that is required for all populations to go extinct one-by-one: $N \cdot \exp(-t_{\text{wave}}/N) \sim 1$ or $t_{\text{wave}} \sim N \cdot \log_e N$. At around this time the last big population collapses again.

Figure 9.14 shows the time-average distribution of populations. This distribution exhibits two pronounced peaks. The upper peak consists of the populations that have not yet collapsed. The lower peak correspond the already collapsed ones. The log binned tail of the distribution of the upper tail of the distribution scale as $\propto 1/s^{0.7}$. This correspond to observing a given populations with size between $[s, s + ds]$ with probability $\propto 1/s^{0.7+1}$.

At different times within each wave, the population distribution shifts. To better understand this dynamics Fig. 9.14 also show the distribution of population sizes at the very end of a diversity wave (green line) and at the beginning of the next wave (red line).

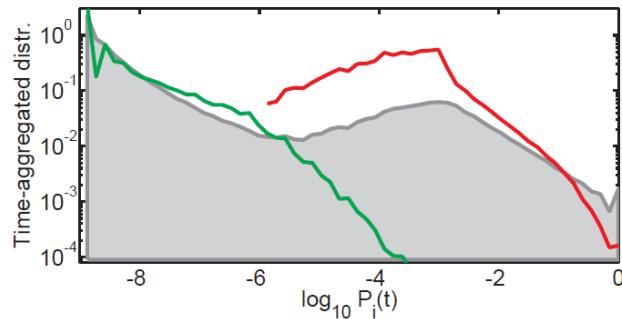


Figure 9.14: The grey shaded area shows the time averaged distribution of population sizes, $\pi(s) = d\text{Prob}(s_i(t) > s)/d(\log_{10} s)$ in the collapse model with $\gamma = 10^{-9}$ and $N = 1000$. The green line show the size distributions collected at the very end of each wave. The red line show the size distributions collected at the very beginning of the next wave. Figure reprinted with permission from [177].

To gain insight into the origins of the scale-free tail in Fig. 9.14 consider now the simplified model where all s_i are set to be equal at the start of each wave. As time t progresses the number of surviving populations decreases as $N_{\text{surv}}(t) = N \exp(-t/N)$. These then all have the same size $s = 1/N_{\text{surv}}(t)$ jointly filling up the carrying capacity of the market (we assume that $\gamma \ll 1/N$). The time-averaged probability to observe a size $s_i > s = 1/N_{\text{surv}}(t)$ is

therefore given by $N_{surv}(t)/N \propto 1/s$:

$$\begin{aligned} \text{Prob}(s_i > s) &\propto \frac{1}{s} \Rightarrow \\ \text{Prob}(s_i = s) &= \frac{d\text{Prob}(s_i > s)}{ds} \propto \frac{1}{s^2} \end{aligned}$$

The exponent $\tau = 2$ deviates substantially from the exponent $\tau \simeq 1.7$ from Fig. 9.14. This deviation is the result of unequal sizes at the beginning of a new wave. These inequality in turn is because they started to grow at different times during the previous diversity wave.

The assumptions and behavior of the collapse driven model is in contrast to the steady state perspective obtained from a long simulation of extended Lotka-Volterra equations. By its near mythical perspective (Trimurti, from the Indian world of Goods) on the cycle of introduction of new, growth+maintenance and collapse the model emphasize dynamics on a timescale where severe reordering takes place. And a reordering that also causes larger scale (diversity) cycles. In microbial ecology the collapse-creation analogy may correspond to an interpretation of the collapse in terms of extinction and subsequent introduction of a new species with a tiny population.

The collapse model is clearly simplified, for example in assuming that there is only one limiting resource to be redistributed, that all species have same growth rates, and in assuming that collapse frequency is independent of population size. All of these assumptions can be relaxed, while still maintaining the main prediction: That on going external driven extinctions facilitates a very wide distribution in population sizes, with an on going shift in who is dominating. The last feature is observed in some microbial ecological data, for example in industrial wastewater [179] and algea bloom [180, 181]. Perhaps more important, the collapse scenario put the rare but dramatic events in focus, a type of event that is ignored if one solely focus on the typically observed behavior.

Lessons:

- “Survival of the fittest” is context dependent, and context may be under selective evolution as well.

Supplementary Literature:

Sneppen, Kim, and Stefan Bornholdt. ”Globalization in a nutshell.” Physical Review E 98.4 (2018): 042314.

Sneppen, Kim, and Namiko Mitarai. ”Multistability with a metastable mixed state.” Physical review letters 109.10 (2012): 100602.

Bornholdt, Stefan, Mogens Høgh Jensen, and Kim Sneppen. ”Emergence and

decline of scientific paradigms.” Physical review letters 106.5 (2011): 058701. Haerter, Jan O., Namiko Mitarai, and Kim Sneppen. ”Theory of invasion extinction dynamics in minimal food webs.” Physical Review E 97.2 (2018): 022404.

Mitarai, Namiko, Joachim Mathiesen, and Kim Sneppen. ”Emergence of diversity in a model ecosystem.” Physical Review E 86.1 (2012): 011929.

Maslov, Sergei, and Kim Sneppen. ”Diversity waves in collapse-driven population dynamics.” PLoS computational biology 11.9 (2015): e1004440.

Bibliography

- [1] Philip W Anderson. More is different. *Science*, 177(4047):393–396, 1972.
- [2] Wilhelm Lenz. Beiträge zum verst̄ndnis der magnetischen eigenschaften in festen körnern. *Physikalische Z*, 21(613-615):1, 1920.
- [3] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.
- [4] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.
- [5] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.
- [6] Gyan Bhanot. The metropolis algorithm. *Reports on Progress in Physics*, 51(3):429, 1988.
- [7] Persi Diaconis and Laurent Saloff-Coste. What do we know about the metropolis algorithm? *Journal of Computer and System Sciences*, 57(1):20–36, 1998.
- [8] Ulli Wolff. Collective monte carlo updating for spin systems. *Physical Review Letters*, 62(4):361, 1989.
- [9] Helmut G Katzgraber. Introduction to monte carlo methods. *arXiv preprint arXiv:0905.1629*, 2009.
- [10] Stephen G Brush. History of the lenz-ising model. *Reviews of modern physics*, 39(4):883, 1967.
- [11] WP Wolf. The ising model and real magnetic materials. *Brazilian Journal of Physics*, 30(4):794–810, 2000.
- [12] PH Lundow and K Markstrom. The critical behaviour of the ising model on the 4-dimensional lattice preprint. *arXiv preprint arXiv:1202.3031*, 2004.
- [13] E Ao Guggenheim. The principle of corresponding states. *The Journal of Chemical Physics*, 13(7):253–261, 1945.
- [14] RHM Morais, JP Santos, S Oliveira, RGB Mendes, DS Rosa, RM Francisco, and FC Sá Barreto. Magnetocaloric effect in the potts model based on the effective-field theory. *Physics Letters A*, 424:127844, 2022.

- [15] FY Wang and ZG Dai. Self-organized criticality in x-ray flares of gamma-ray-burst afterglows. *Nature Physics*, 9(8):465, 2013.
- [16] V. Pareto. La legge della domanda. *Giornale degli Economisti*, 10(59-68):691–700, 1895.
- [17] 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.
- [18] H. A. Simon. On a Class of Skew Distribution Functions. *Biometrika*, 42:425–440, 1955.
- [19] G. K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, Massachusetts, 1949.
- [20] Simon R Broadbent and John M Hammersley. Percolation processes: I. crystals and mazes. In *Mathematical proceedings of the Cambridge philosophical society*, volume 53, pages 629–641. Cambridge University Press, 1957.
- [21] Muhammad Sahimi. *Applications of percolation theory*. CRC Press, 1994.
- [22] MEJ Newman and Robert M Ziff. Efficient monte carlo algorithm and high-precision results for percolation. *Physical Review Letters*, 85(19):4104, 2000.
- [23] Takashi Hara and Gordon Slade. Mean-field critical behaviour for percolation in high dimensions. *Communications in Mathematical Physics*, 128(2):333–391, 1990.
- [24] 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.
- [25] Lene Oddershede, Peter Dimon, and Jakob Bohr. Self-organized criticality in fragmenting. *Physical review letters*, 71(19):3107, 1993.
- [26] Stephen Wolfram. Cellular automata as models of complexity. *Nature*, 311(5985):419–424, 1984.
- [27] John Von Neumann, Arthur W Burks, et al. Theory of self-reproducing automata. *IEEE Transactions on Neural Networks*, 5(1):3–14, 1966.
- [28] Martin Gardner. The fantastic combinations of jhon conway’s new solitaire game’life. *Sc. Am.*, 223:20–123, 1970.
- [29] Stephen Wolfram. Statistical mechanics of cellular automata. *Reviews of modern physics*, 55(3):601, 1983.
- [30] K. Jacobs. *Stochastic Processes fro Physicists*. Cambridge University Press, 2010.
- [31] NS Bingham, S Rooke, J Park, A Simon, W Zhu, X Zhang, J Batley, JD Watts, C Leighton, KA Dahmen, et al. Experimental realization of the 1d random field ising model. *Physical Review Letters*, 127(20):207203, 2021.

- [32] Vidar Frette, Kim Christensen, Anders Malthe-Sørenssen, Jens Feder, Torstein Jøssang, and Paul Meakin. Avalanche dynamics in a pile of rice. *Nature*, 379(6560):49–52, 1996.
- [33] N. Eldredge. *Life Pulse, Episodes from the story of the fossil record*. Facts on File Publications (New York), New York, 1987.
- [34] J. J. Sepkoski. Ten years in the library: new data confirm paleontological patterns. *Paleobiology*, 19:43–51, 1993.
- [35] 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.
- [36] L. W. Alvarez. Mass extinctions caused by large solid impacts. *Physics Today*, pages 24–33, 1987.
- [37] K. J. Kauffman, P. Prakash, and J. S. Edwards. Advances in flux balance analysis. *Current Opinion in Biotechnology*, 14:491–496, 2003.
- [38] P. Bak and K. Sneppen. Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.*, 71:4083, 1993.
- [39] 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.
- [40] N. Eldredge and S. J. Gould. Punctuated equilibria: An alternative to phyletic gradualism. In T. J. M Schopf, J. M. Thomas, and S. Francisco, editors, *Models in Paleobiology*. Freeman and Cooper, 1972.
- [41] S. J. Gould and N. Eldredge. Punctuated equilibrium comes of age. *Nature*, 366:223–227, 1993.
- [42] G. G. Simpson. *Tempo and Mode in Evolution*. Columbia Univ. Press, New York, 1944.
- [43] G. G. Simpson. *The Major Features of Evolution*. Columbia Univ. Press, New York, 1953.
- [44] K. Sneppen. Extremal dynamics and punctuated co-evolution. *Physica A*, 221:168, 1995.
- [45] M. Paczuski, S. Maslov, and P. Bak. Avalanche dynamics in evolution, growth, and depinning models. *Phys. Rev. E*, 53:414–443, 1996.
- [46] H. Flyvbjerg, K. Sneppen, and P. Bak. Mean field model for a simple model of evolution. *Phys. Rev. Lett.*, 71:4087, 1993.
- [47] 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.
- [48] Kim Sneppen. Self-organized pinning and interface growth in a random medium. *Physical Review Letters*, 69(24):3539, 1992.

- [49] Y. G. Jin et al. Pattern of marine mass extinction near the permian-triassic boundary in south china. *Science*, 289:432–436, 2000.
- [50] A. Trusina, M. Rosvall, and K. Sneppen. Communication boundaries in networks. *Phys. Rev. Lett.*, 94:238701, 2004.
- [51] 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.
- [52] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:1760, 1960.
- [53] 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.
- [54] K. Christensen, R. Donangelo, B. Koiler, and K. Sneppen. Evolution of random networks. *Phys. Rev. Letters*, 81:2380, 1998.
- [55] R. D. Luce and A. D. Perry. A method of matrix analysis of group structure. *Psychometrika*, 14:95–116, 1949.
- [56] P. W. Holland and S. Leinhardt. *Transitivity in structural models of small groups*. *Comparative Group Studies*, 2:107–124, 1971.
- [57] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:409–410, 1998.
- [58] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40:35–41, 1977.
- [59] L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1:215–239, 1978/1979.
- [60] 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.
- [61] 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.
- [62] H. Yu, P. M. Kim, E. Sprecher, V. Trifanov, and M. Gerstein. The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *Plos. Comp. Biol.*, 3:713–720, 2007.
- [63] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA*, 99:7821–7826, 2002.
- [64] K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen. Modularity and extreme edges of the internet. *Phys. Rev. Lett.*, 90:148701, 2003.
- [65] K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen. Modularity and extreme edges of the internet. *Physica A*, 336:163, 2004.

- [66] 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.
- [67] H. Rieger J. D. Noh. Random walks on complex networks. *Phys. Rev. Letters*, 92:118701, 2004.
- [68] M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27:39–54, 2005.
- [69] A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 509:286, 1999.
- [70] 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.
- [71] Shang-Keng Ma. *Statistical Mechanics*. World Scientific Publishing Co Inc, 1985.
- [72] P. Minnhagen and S. Bernhardsson. The blind watchmaker network: Scale-freeness and evolution. *PLoS ONE*, 3:e1690, 2008.
- [73] R. Cohen, K. Erez, D. ben Avraham, and S Havlin. “resilience of the internet to random breakdowns.”. *Phys. Rev. Lett*, 85:4626–4628, 2000.
- [74] M. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, 2002.
- [75] M. E. J Newman. The structure and function of complex networks. *SIAM Rev*, 45:167–256., 2002.
- [76] L. D. Valdez, C. Buono, P. A. Macri, and L. A. Braunstein. Social distancing strategies against disease spreading. *arXiv:*, 1308, 2009.
- [77] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296:910, 2002.
- [78] 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.
- [79] R. Milo et al. Superfamilies of evolved and designed networks. *Science*, 303:1538–1542, 2004.
- [80] 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.
- [81] 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.
- [82] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA*, 100:11980–11985, 2003.
- [83] 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.

- [84] J. B Axelsen, S. Bernhardsson, M. Rosvall, K. Sneppen, and A. Trusina. Degree landscapes in scale-free networks. *Phys. Rev. E*, 74:036119, 2006.
- [85] D Hughes, M Paczuski, RO Dendy, P Helander, and KG McClements. Solar flares as cascades of reconnecting magnetic loops. *Physical review letters*, 90(13):131101, 2003.
- [86] 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.
- [87] T. C. Schelling. *Models of Segregation. The American Economic Review*,, 59:488–493, 1969.
- [88] S. Wolfram. Statistical mechanics of cellular automata. *Reviews of Modern Physics*, 55:601–644, 1983.
- [89] D. Helbing I. Farkas and T. Vicsek. Simulating dynamical features of escape panic. ”. *Nature*, 407:487–490, 2000.
- [90] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. USA*, 99:7280–7287, 2002.
- [91] E. Bonabeau, M. Dorigo, and G. Theraulaz. Inspiration for optimization from social insect behaviour. *Nature*, 406:39–42, 2000.
- [92] 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.
- [93] 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.
- [94] R. Axelrod. The dissemination of culture - a model with local convergence and global polarization. *Journal of Conflict Resolution*, 41:203–226, 1997.
- [95] K. Klemm, V. M. Eguiluz, R. Toral, and M. San Miguel. *Phys. Rev. E*, 67:045101, 2003.
- [96] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social phenomena. *Rev. Mod. Phys.*, 81:591–646, 2009.
- [97] P. Clifford and A. Sudbury. *Biometrika*, 60:581–588, 1973.
- [98] R. A. Holley and T. M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probab.*, 3:643–663, 1975.
- [99] S. Galam. Minority opinion spreading in random geometry. *Eur. Phys. J. B*, 25:403–406, 2002.
- [100] K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *Int. J. Mod. Phys. C*, 11:1157–1165, 2000.

- [101] P. Chen and S. Redner. Majority rule dynamics in finite dimensions. *Phys. Rev. E*, 71:036101, 2005.
- [102] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3:87–98, 2000.
- [103] S. Huet, G. Deffuant, and W. Jager. *Advances in Complex Systems*, 11:529, 2008.
- [104] M. Rosvall and K. Sneppen. Modeling self-organization of communication and topology in social networks. *Phys. Rev. E*, 74:16108, 2006.
- [105] M. Rosvall and K. Sneppen. Reinforced communication and social navigation generate groups in model networks. *Phys. Rev. E*, 79:026111, 2009.
- [106] 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.
- [107] F. Vazquez, P.L.Krapivsky, and S. Redner. Constrained opinion dynamics: freezing and slow evolution. *Journal of Physics A*, 36:L61–L68, 2003.
- [108] I. B. Dodd, M. A. Micheelsen, K. Sneppen, and G. Thon. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell*, 129:813–22, 2007.
- [109] M. A. Micheelsen, N. Mitarai, K. Sneppen, and I. B. Dodd. Theory for the stability and regulation of epigenetic landscapes. *Phys Biol*, 7:026010, 2010.
- [110] Kim Sneppen and Namiko Mitarai. Multistability with a metastable mixed state. *Physical review letters*, 109(10):100602, 2012.
- [111] K. Yanagita. *Kagyuko*. Toko Shoin, Tokyo, 1930.
- [112] K. Sneppen, A. Trusina, M.H. Jensen, and S. Bornholdt. A minimal model for multiple epidemics and immunity spreading. *Plos One*, 5:e13326, 2010.
- [113] F. Uekermann and K. Sneppen. Spreading of multiple epidemics with cross immunization phys. *Rev. E*, 86:036108, 2012.
- [114] 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.
- [115] 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.
- [116] Nathan D Wolfe, Claire Panosian Dunavan, and Jared Diamond. Origins of major human infectious diseases. *Nature*, 447(7142):279–283, 2007.
- [117] Danielle Miller, Michael A Martin, Noam Harel, Talia Kustin, Omer Tirosh, Moran Meir, Nadav Sorek, Shiraz Gefen-Halevi, Sharon Amit, Olesya Vorontsov, et al. Full genome viral sequences inform patterns of sars-cov-2 spread into and within israel. *medRxiv*, 2020.

- [118] Akira Endo, Sam Abbott, Adam J Kucharski, Sebastian Funk, et al. Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china. *Wellcome Open Research*, 5(67):67, 2020.
- [119] Julius B Kirkegaard and Kim Sneppen. Variability of individual infectiousness derived from aggregate statistics of covid-19. *medRxiv*, 2021.
- [120] Bjarke Frost Nielsen, Lone Simonsen, and Kim Sneppen. Covid-19 superspreading suggests mitigation by social network modulation. *Physical Review Letters*, 126(11):118301, 2021.
- [121] Janet Raboud, Altynay Shigayeva, Allison McGeer, Erika Bontovics, Martin Chapman, Denise Gravel, Bonnie Henry, Stephen Lapinsky, Mark Loeb, L Clifford McDonald, et al. Risk factors for sars transmission from patients requiring intubation: a multicentre investigation in toronto, canada. *PLoS One*, 5(5):e10717, 2010.
- [122] Zhuang Shen, Fang Ning, Weigong Zhou, Xiong He, Changying Lin, Daniel P Chin, Zonghan Zhu, and Anne Schuchat. Superspreading SARS events, Beijing, 2003. *Emerging infectious diseases*, 10(2):256, 2004.
- [123] AJ Kucharski and Christian L Althaus. The role of superspreading in middle east respiratory syndrome coronavirus (mers-cov) transmission. *Eurosurveillance*, 20(25):21167, 2015.
- [124] James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.
- [125] Christian L Althaus. Ebola superspreading. *The Lancet Infectious Diseases*, 15(5):507–508, 2015.
- [126] Max SY Lau, Benjamin Douglas Dalziel, Sebastian Funk, Amanda McClelland, Amanda Tiffany, Steven Riley, C Jessica E Metcalf, and Bryan T Grenfell. Spatial and temporal dynamics of superspreading events in the 2014–2015 west africa ebola epidemic. *Proceedings of the National Academy of Sciences*, 114(9):2337–2342, 2017.
- [127] Christophe Fraser, Derek AT Cummings, Don Klinkenberg, Donald S Burke, and Neil M Ferguson. Influenza transmission in households during the 1918 pandemic. *American journal of epidemiology*, 174(5):505–514, 2011.
- [128] Max SY Lau, Bryan Grenfell, Michael Thomas, Michael Bryan, Kristin Nelson, and Ben Lopman. Characterizing superspreading events and age-specific infectiousness of sars-cov-2 transmission in georgia, usa. *Proceedings of the National Academy of Sciences*, 2020.
- [129] Kim Sneppen, Bjarke Frost Nielsen, Robert J Taylor, and Lone Simonsen. Overdispersion in covid-19 increases the effectiveness of limiting nonrepetitive contacts for transmission control. *Proceedings of the National Academy of Sciences*, 118(14), 2021.

- [130] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440–442, 1998.
- [131] Jörn Davidsen, Holger Ebel, and Stefan Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks. *Physical review letters*, 88(12):128701, 2002.
- [132] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.
- [133] Andreas Eilersen and Kim Sneppen. Estimating cost-benefit of quarantine length for covid-19 mitigation. *medRxiv*, 2020.
- [134] Louis Bachelier. *Théorie de la spéculation*. Gauthier-Villars, 1900.
- [135] Ingve Simonsen and Kim Sneppen. Profit profiles in correlated markets. *Physica A: Statistical Mechanics and its Applications*, 316(1):561–567, 2002.
- [136] Jens Feder. *Fractals*. Springer Science & Business Media, 2013.
- [137] Rafal Weron. *Modeling and forecasting electricity loads and prices: a statistical approach*, volume 403. John Wiley & Sons, 2007.
- [138] Ingve Simonsen. Volatility of power markets. *Physica A: Statistical Mechanics and its Applications*, 355(1):10–20, 2005.
- [139] 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.
- [140] 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.
- [141] 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.
- [142] 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.
- [143] Ch Baek and M Elbeck. Bitcoins as an investment or speculative vehicle? a first look. *Applied Economics Letters*, 22(1):30–34, 2015.
- [144] W Brian Arthur. Competing technologies, increasing returns, and lock-in by historical events. *The economic journal*, 99(394):116–131, 1989.
- [145] Jay R Ritter. Behavioral finance. *Pacific-Basin finance journal*, 11(4):429–437, 2003.
- [146] Jianjun Miao. Introduction to economic theory of bubbles. *Journal of Mathematical Economics*, 53:130–136, 2014.

- [147] Irving Fisher. The debt-deflation theory of great depressions. *Econometrica: Journal of the Econometric Society*, pages 337–357, 1933.
- [148] Ayumu Yasutomi. The emergence and collapse of money. *Physica D: Nonlinear Phenomena*, 82(1-2):180–194, 1995.
- [149] Raul Donangelo and Kim Sneppen. Self-organization of value and demand. *Physica A: Statistical Mechanics and its Applications*, 276(3-4):572–580, 2000.
- [150] 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.
- [151] Jean-Philippe Bouchaud. Crises and collective socio-economic phenomena: simple models and challenges. *Journal of Statistical Physics*, 151(3-4):567–606, 2013.
- [152] Herbert Spencer. *Railway Morals & Railway Policy*, volume 65. Longman, Brown, Green & Longmans, 1855.
- [153] Garrett Hardin. The tragedy of the commons. *science*, 162(3859):1243–1248, 1968.
- [154] J.L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35:917–926, 1956.
- [155] C. T. Bergström and M. Lachman. Shannon information and biological fitness. *Information Theory Workshop, IEEE*, 0-7803-8720-1:50–54, 2004.
- [156] E. Kussell, R. Kishony, N-Q. Balaban, and S. Leibler. Bacterial persistence a model of survival in changing environments. *Genetics*, 169:1807–1814, 2005.
- [157] E. Kussell and S. Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309:2075–2078, 2005.
- [158] S. Maslov and K. Sneppen. Well temperate phage. *preprint, arXiv:1308.1646*, 2013.
- [159] Maslov S and Zhang Y-C. Optimal investment strategy for risky assets. *International Journal of Theoretical and Applied Finance*, 1:377–387, 1998.
- [160] Mehran Kardar, Giorgio Parisi, and Yi-Cheng Zhang. Dynamic scaling of growing interfaces. *Physical Review Letters*, 56(9):889, 1986.
- [161] K. Sneppen and N. Mitarai. Multistability with a mixed metastable state. *Phys. Rev. Lett.*, 109:100602, 2012.
- [162] I. B. Dodd and K. Sneppen. Barriers and silencers: A theoretical toolkit for control and containment of nucleosome-based epigenetic states. *J. Mol. Biol.*, 414:624–637, 2011.
- [163] D. Abrams and S. Strogatz. Linguistics: Modeling the dynamics of language death. *Nature*, 424:900, 2003.

- [164] G. C. M. George, A. Ehrhardt, M. Marsili, and Fernando Vega-Redondo. Phenomenological models of socioeconomic network dynamics. *Phys. Rev. E*, 74:036106, 2006.
- [165] R. May. Will a large complex system be stable. *Nature*, 238:413–414, 1972.
- [166] M. C. Boerlijst and P. Hogeweg. Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites. *Physica D*, 48:17–28, 1991.
- [167] J. Bascompte and R.V. Solé. Rethinking complexity: modelling spatiotemporal dynamics in ecology. *Trends in Ecology & Evolution*, 10(9):361–366, 1995.
- [168] B. Kerr, M. A. Riley, M. W. Feldman, and B. J. M. Bohannan. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature*, 418:171–174, 2002.
- [169] T. Reichenbach, M. Mobilia, and E. Frey. Coexistence versus extinction in the stochastic cyclic lotka-volterra model. *Phys. Rev.*, 74:E051907, 2006.
- [170] J. Mathiesen, N. Mitarai, K. Sneppen, and A. Trusina. Ecosystems with mutually exclusive interactions self organize into a state of high diversity. *Phys. Rev. Lett.*, 107:188101, 2011.
- [171] N. Mitarai, J. Mathiesen, and Kim Sneppen. Emergence of diversity in a model ecosystem. *Phys. Rev. E*, 2012.
- [172] Kim Sneppen. *Models of life*. Cambridge University Press, 2014.
- [173] Kim Sneppen. Models of life: epigenetics, diversity and cycles. *Reports on Progress in Physics*, 80(4):042601, 2017.
- [174] ARD Stebbing. Competition for space between the epiphytes of fucus serratus l. *Journal of the Marine Biological Association of the United Kingdom*, 53(2):247–261, 1973.
- [175] JBC Jackson and LEO Buss. Alleopathy and spatial competition among coral reef invertebrates. *Proceedings of the National Academy of Sciences*, 72(12):5160–5163, 1975.
- [176] Patricia M Harris. Competitive equivalence in a community of lichens on rock. *Oecologia*, 108(4):663–668, 1996.
- [177] Sergei Maslov and Kim Sneppen. Diversity waves in collapse-driven population dynamics. *PLoS computational biology*, 11(9):e1004440, 2015.
- [178] Pierre-François Verhulst. Notice sur la loi que la population suit dans son accroissement. correspondance mathématique et physique publiée par a. Quetelet, 10:113–121, 1838.
- [179] Orr H Shapiro, Ariel Kushmaro, and Asher Brenner. Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. *The ISME journal*, 4(3):327–336, 2010.

- [180] T Castberg, A Larsen, RA Sandaa, CPD Brussaard, JK Egge, M Heldal, R Thyrhaug, EJ Van Hannen, and G Bratbak. Microbial population dynamics and diversity during a bloom of the marine coccolithophorid *emiliania huxleyi* (haptophyta). *Marine Ecology Progress Series*, 221:39–46, 2001.
- [181] Barbara J Campbell, Liying Yu, John F Heidelberg, and David L Kirchman. Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Sciences*, 108(31):12776–12781, 2011.