UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master's Degree in Physics

Final Dissertation

Scaling and Renormalization Group for models of neural activity

Thesis supervisor:

Prof. Amos Maritan

Thesis co-supervisor:

Prof. Samir Suweis

Candidate:

Giorgio Nicoletti

Academic year 2018/2019

More is different.
– PHILIP W. ANDERSON

ACKNOWLEDGMENTS

I thank Prof. Maritan and Prof. Suweis for their invaluable help and support, without which this work would have never been possible.

I would also like to thank my parents, my brother and my family for making it all possible in the first place and for believing in me from the very first day. And thank you, Ilaria, for keeping up with me for all this time.

Lastly, I thank Gea for diligently watching over me whenever I am at my desk. I am quite sure that, after all these years, you understand Physics better than anyone else.

CONTENTS

PREFACE

This thesis is inspired by a fascinating idea that has started its development in recent years: that the emerging phenomena of the natural world, such as of living and biological systems, can be described by the Statistical Mechanics of phase transitions. Criticality seems to be indeed a simple and elegant paradigm that is able to provide complex and rich features, whilst starting from underlying elements that are relatively simpler. Hence, it might be the answer to the question of how the complexity of the world we live in might emerge from interacting parts that, otherwise, hardly suggest the presence of such diversity at large scale.

On the other hand, criticality also seems to have been exploited by living systems. Being at the edge between order and disorder provides a tradeoff between resilience to external changes and adaptability to the environment. Moreover, topical features of criticality such as the existence of correlations at all scales and the divergence of the susceptibility could provide living systems with coherent, system wide behaviors and high sensitivity to external stimuli. But criticality is also tempting from a theoretical point of view. Indeed, the macroscopic behavior that emerges at a critical point is largely independent on most of the details of the system, giving rise to universality, which implies that we would not need to know them all to describe the large scale properties we are interested in. In recent years many arguments have been proposed that explore why criticality could provide a number of evolutionary advantages in biological systems, together with a proliferation of experimental evidence of the presence of criticality in very diverse systems, from flock of birds to neural activity. However, given both the profound implications of this idea and the difficulties in analyzing data from these kind of systems, a definitive answer is far from being reached.

This is why this thesis deals with the possibility of adapting one of the most powerful tool of the Statistical Mechanics of phase transitions, the Renormalization Group, to understand if a system is indeed at its critical point even when we do not have an explicit model to write down. This is far from being a trivial point, because it is often the case that we can only rely on data to seek signatures of criticality. Let us think about the brain as an archetypal example: it is a system so complex that it is hard to imagine if we will ever be able to grasp its full structure, let alone describe it down to the last detail. However, experimental breakthroughs now allow to record simultaneously the activity of more than a thousand neurons in their full dynamical evolution. Can we use such data to understand if brains really operate at a critical

point, even if we have no hope to write down an explicit model? The first time such a question was tackled was a few months ago with the introduction of a possible realization of a phenomenological Renormalization Group [34], and this thesis moves from this approach. In particular, we want to test this procedure to probe its ability to predict the presence of criticality and to quantify how precise it is in doing so.

The first chapter provides a brief introduction to the concept of criticality and its relation with scale invariance, self-similarity and power laws, features that are often looked for in data from experiments. The Ising model, in particular, is one of the most famous and simplest equilibrium model that undergoes a second order phase transition and proves to be very useful in understanding many properties of critical systems. In this chapter we also give a review of some of the experimental evidences that have been proposed in favor of criticality in living systems, trying to cover a number of very diverse settings to give a glimpse of how broad this concept is. Of course, if natural systems really are critical we should also try to explain why and how they poise themselves at criticality. Self-organized criticality is one of the most interesting possibilities, where the control parameter itself becomes a dynamical variable. We give an explicit example of SOC in terms of the branching process, and we also review a possible information theoretic approach as to why being critical should be an optimal strategy for a living system.

The second chapter is devoted to the contact process, a non equilibrium model that displays an absorbing phase transition between a silent state and a state where activity proliferates. We are particularly interested in the contact process because it is the simplest archetypal model for neural activity, hence it should be an excellent framework to test a phenomenological RG and make some parallels with the results obtained for real neurons. We provide a broad introduction to the concepts of dynamical phase transitions, and we solve explicitly the contact process in the mean field approximation since the one-dimensional case is already too hard to be dealt with analytically. We also derive most of the scaling laws and the relations among the critical exponents, and we simulate the model both in mean field and in a $2D$ lattice. In both cases, we obtain good agreements with the exact exponents for the mean field approximation and with the numerical values in the literature for the two-dimensional case.

The third chapter reviews in depth the phenomenological Renormalization Group proposed in [34]. In particular we show its connections with both the idea that a RG transformation can be though of as acting on probability distributions, which proved to be an extremely powerful mindset in the context of generalized limit theorems in probability theory, and with a well known unsupervised learning algorithm, principal component analysis. More details about them both are given in the appendices. Finally, we apply the phenomenological procedure to simulation data from the Ising model and from the contact process. For the latter, we also change the underlying topology to the one of a Watts-Strogatz network, where long-range connections are present, so we are able to check the stability of the results in the presence of long-range interactions. In all these cases we look at different regions of the control parameter, in order to fully understand how sensible this procedure is with respect to deviation from criticality.

CHAPTER 1

A HINT OF CRITICALITY

In this chapter we want to review the main ideas of criticality and give a glimpse of how they can be applied to a very vast landscape of fields and problems. Statistical Mechanics seems to be well suited to capture the emergent phenomena of the natural world: it was built to deal with a large number of degrees of freedom and collective behaviors in the first place. But, fascinatingly enough, biological and living systems appear to behave, as if they were poised at the critical point of a phase transition where collective phenomena and universality overcome the very microscopic details of the system.

While far from being a well-founded approach, the ideas born in the Statistical Mechanics of critical systems - from scaling to the Renormalization Group - have proven to be powerful tools in the study of interacting living systems, or at least a good starting point. Of these ideas, the spontaneous emergence of critical-like behavior in so many and so different systems is perhaps one of the most fascinating fact, albeit still speculative, of the natural world.

1.1 FROM SCALE INVARIANCE TO PHASE TRANSITIONS

Scale invariance is at the heart of critical phenomena. A system is said to be scale invariant if its physical properties remain unchanged when we describe it at different spatial scales. Self-similar geometrical objects like fractals are a good example of scale invariance: no matter how magnified, a fractal structure remains the same at all scales. They are effectively scale-free.

Suppose that we describe a physical property of a scale invariant system with a simple one-dimensional scalar function $f(x)$. If the relative variation of the function

depends only on the relative value of the variables, namely

$$\frac{f(y)}{f(x)} = \phi\left(\frac{y}{x}\right) \quad \forall x, y, \tag{1.1}$$

then the function f varies in the same way at all scales. The class of functions that satisfy this condition can be easily found. Assuming that f is differentiable and differentiating with respect to y at the point $y = x$ yields

$$\left.\frac{f'(y)}{f(x)}\right|_{y=x} = \left.\frac{1}{x}\phi'\left(\frac{y}{x}\right)\right|_{y=x},$$

that is the differential equation $f'(x) = \phi'(1)f(x)x^{-1}$. Thus the only functions that satisfy (1.1) are power laws

$$f(x) = cx^a \tag{1.2}$$

where $a = \phi'(1) = \text{const}$.

In fact, we can describe a scale change as a dilation

$$x \to \lambda x,$$

and a power law scales trivially under this transformation

$$f(\lambda x) = c(\lambda x)^a = \lambda^a f(x).$$

Therefore it is natural to expect that the properties of a scale-free system are described by power laws. A function $g(x)$ with a given characteristic scale ξ, say $g(x) = e^{-x/\xi}$, would transform differently at different scales. This is the reason for which power laws are regarded as the signature of scale invariance.

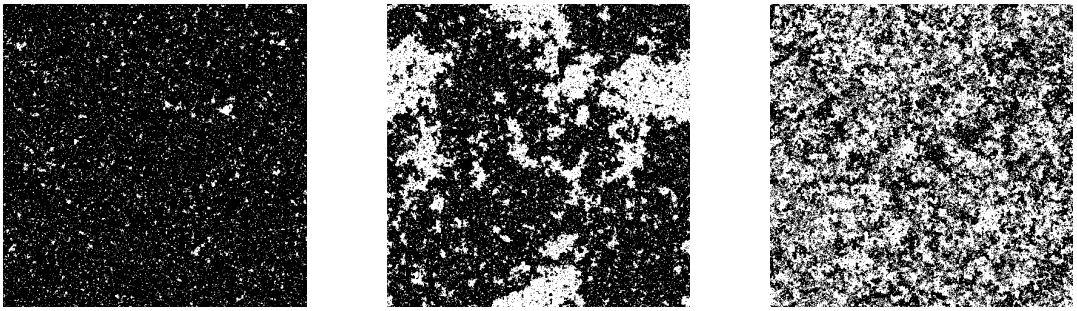
1.1.1 CONTINUOUS PHASE TRANSITIONS

One of the most striking feature of scale invariance is how it arises in the context of continuous phase transitions, where qualitatively different states are separated by a critical point [6, 25]. The most known example is probably the Ising model on a hypercubic lattice, defined by the Hamiltonian

$$H_{\text{Ising}} = -J\sum_{\langle ij\rangle} S_i S_j, \quad S_i = \pm 1 \,\forall i \tag{1.3}$$

where the sum runs over the lattice's nearest neighbors. It is well known that in dimension $d > 1$ the Ising model displays a second order phase transition at a non-zero temperature $T = T_c$ [38]. In order to describe such transition we introduce a local order parameter, namely a physical quantity that vanishes in one phase and is non vanishing in the other. In the case of the Ising model, we consider the magnetization

$$m = \frac{1}{N}\left\langle \sum_{i=1}^{N} S_i \right\rangle$$



(A) $T = 0.95T_c$ (B) $T = T_c$ (C) $T = 1.05T_c$

FIGURE 1.1: Configurations of the two dimensional Ising model in a 500×500 lattice below, at and above criticality. These and all the other configurations in this chapter are obtained from a Monte Carlo simulation we performed using the Wolff algorithm.

which is in fact zero if the spins are randomly oriented, as it happens at high T. It is around the critical temperature that the physical properties of the model are the most interesting.

The interaction described by the Hamiltonian (1.3) is clearly a short range interaction, and indeed at high temperature $T \gg T_c$ the model displays short range correlations and disorder. The magnetization is vanishing and spins are randomly oriented. On the other side of the phase transition, for $T \ll T_c$ correlations are still short range, but long range order appears: the magnetization is different from zero because the spins tend to be aligned[1]. In Figure 1.1 we show the results of three simulations of the Ising model at different temperature and how even a small deviation from $T = T_c$ can lead to a very different configuration.

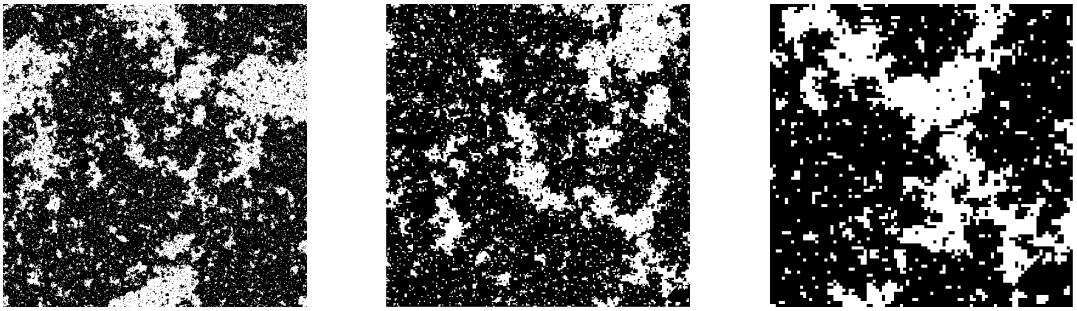
But as we approach T_c something remarkable happens: the correlations between the spins become more and more long range, leading to the formation of island of aligned spins. Exactly at criticality, these islands exist at all scales and the configuration of the Ising model becomes fractal-like. Figure 1.2 shows how sublattices at different scale resemble the full lattice.

This is perhaps one of the most striking features of criticality: at the critical point the correlations between the microscopic degrees of freedom become relevant at all scales, regardless of the nature of the starting interaction. Since no relevant scale can be defined, the physical properties of the model turn out to be scale-free as well and therefore it makes sense to describe them via power laws. For instance, the order parameter scales as

$$m \sim |t|^{\beta} \quad (t < 0) \tag{1.4}$$

where $t = (T - T_c)/T_c$ and β is called a critical exponent.

[1]The intuitive reason for the presence of phase transitions is probably best understood in terms of the Peierls argument as a competition between minimizing the energy, favoring an ordered phase, and maximizing the entropy, favoring disorder.



(A) $N_{\text{spins}} = 250000$ (B) $N_{\text{spins}} = 62500$ (C) $N_{\text{spins}} = 10000$

FIGURE 1.2: A configuration of the 2D Ising model at criticality. (A) A 500×500 lattice. (B) A 250×250 sublattice. (C) A 100×100 sublattice. Notice how islands of spins exit at all scales.

A number of other exponents can be introduced. For instance, the zero field susceptibility is divergent

$$\chi = \left.\frac{\partial m}{\partial h}\right|_{h=0} \sim |t|^{-\gamma} \tag{1.5}$$

where h is an external pinning field with a corresponding Hamiltonian $H_h = -h \sum_i S_i$ to be added to 1.3. We briefly remind that one should be careful that singularities like (1.5) only appear in the thermodynamic limit $N \to \infty$, where ergodicity is broken. The usefulness of the pinning field h is to select one of the (two, in the case of the symmetry group $\mathbb{Z}^2$ of the Ising model) ergodic regions of the phase space that become disjointed in the thermodynamic limit [20]. Therefore the limit $h \to 0$ should be taken only after $N \to \infty$. In any finite system the susceptibility, and every other response function, will peak at the critical point but will not diverge, that is they are analytic function of both T and h.

The case of the (connected) correlation function

$$G(r) = \langle S_r S_0 \rangle - \langle S_r \rangle \langle S_0 \rangle$$

is particularly interesting. Away from criticality, due to the short range nature of the microscopic interaction, we expect correlations to decay exponentially

$$G(r) \sim e^{-r/\xi}$$

apart from a possible amplitude with power law behavior. ξ is called correlation length and defines the typical scale of the system. However, close to the critical point the correlation length diverges as

$$\xi \sim |t|^{-\nu} \tag{1.6}$$

and exactly at the critical point the decay of the correlations becomes algebraic

$$G(r) \sim r^{-(d-2+\eta)}. \tag{1.7}$$

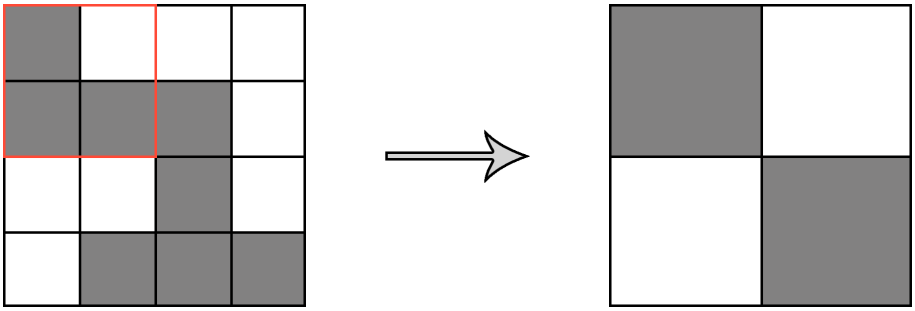


FIGURE 1.3: A block-spin transformation. A grey square represent a spin up, and a white square a spin down. We group together 2×2 blocks and for each of them we define a super spin with a majority rule (the spin is randomly assigned if there are both 2 spins up and 2 spins down..

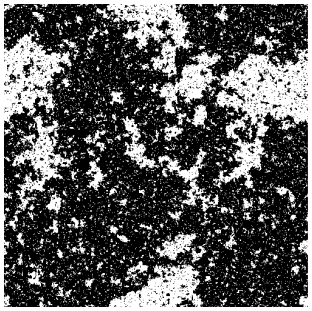
This is precisely what we mean when we say that at criticality the system is scale-free: (1.6) together with (1.7) implies that $\xi \to \infty$ as we approach the critical point, so all scales are equally relevant. The fact that the collective behavior becomes independent of the microscopic interaction is the reason of another feature of critical systems: universality. Indeed the critical behavior depends only on few key characteristics like the symmetry of the Hamiltonian, the dimensionality of the system and the range of interactions.

1.1.2 COARSE GRAINING AND UNIVERSALITY

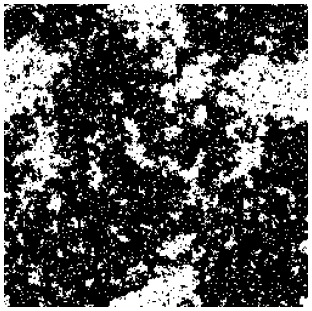
It is an experimental fact that at criticality many different systems share the same set of exponents, i.e. they all fall under the same universality class. From a theoretical point of view this is well understood in the context of the Renormalization Group, but we can get an intuition already at the level of coarse-graining.

Loosely speaking, a coarse-graining transformation amounts to changing the scale at which we look at the system. As an example, in Figure 1.3 we show a possible coarse-graining step: the spins on a 2×2 block are grouped into a single "super spin" defined by a majority rule. This is usually called a block-spin transformation. In the case of the Ising model, clearly enough for $T < T_c$ after enough coarse-graining steps the system ends up in one of its two ground states with all the spins aligned. On the other hand, for $T > T_c$ the system ends up in a fully disordered configuration. What happens is that the coarse-graining procedure smooths out the short range fluctuations of the system, unraveling its long range properties.

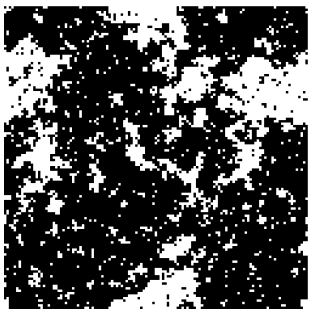
The behavior of the critical case under a coarse-graining transformation is the interesting point. In fact, at criticality the microscopic details do not play a role in the physics of the system and fluctuations exist at all scale: because of this the coarse-graining yields no change at all at the critical point. Figure 1.4 shows the effects on a 2D critical Ising model of repeating the block-spin transformation introduced before. As we can see, the structure remains indeed the same because at criticality the long range physics is the only one that matters. Qualitatively, this is where universality comes to be: at the critical point the microscopic details are irrelevant.



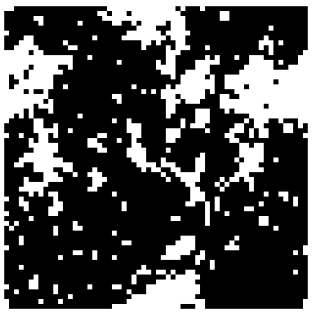
(A) Original lattice.



(B) First generation of spins.



(C) Second generation of spins.



(D) Third generation of spins.

FIGURE 1.4: A critical configuration of the 2D Ising model is left invariant under a coarse-graining transformation due to its self-similarity. The coarse-graining procedure we used here is the one depicted in Figure 1.3.

Power laws can be understood in term of the coarse-graining of a self-similar system as well[2]. Take an Ising model at $T \lesssim T_c$ and a small coarse-grain step $x' = x/b$ with $b = 1 + \epsilon$ and $\epsilon \ll 1$ such that the system remains close to criticality [40]. Still, the coarse-grained system is a bit deeper into the ordered phase and the magnetization will be higher, thus we write for some constant f

$$M'(T) = (1 + f\epsilon)M(T).$$

On the other hand, the coarse-grained system is equivalent to the original system a bit further away from the critical temperature

$$T' = T_c - t' = T_c - (1 + d\epsilon)t$$

where $t = (T_c - T)$ and d is another constant. Thus for sufficiently small ϵ, i.e. around the critical point, we have

$$M'(t) = M(t') = (1 + f\epsilon)M(t) = M(t) + d\epsilon t\frac{\mathrm{d}M}{\mathrm{d}t} + O(\epsilon^2)$$

[2]To be precise, the scaling behavior is understood in terms of the control by the fixed point of the Renormalization Group flow. Now we only aim to review the deep connection between scale invariance, power laws and criticality, albeit in a qualitative fashion.

which implies a power law behavior for the magnetization

$$M(t) \propto t^{f/d}.$$

In full generality, a coarse-graining step changes the correlation length as well

$$\xi' = \frac{\xi}{b}$$

and since ξ is the only relevant scale of the system we shall assume the scaling hypothesis in the following form [29]: at criticality the physical properties $P(t)$ depend on the temperature of the system only through $\xi = \xi(t)$, that is

$$P(t) = \Pi(\xi(t)).$$

Then as we did with the magnetization we ask

$$P'(t) = P(t') = \Pi'(\xi(t)) = \Pi(\xi'(t))$$

that is, the coarse-grained system is equivalent to the original system at a different temperature. Expanding at first order in ϵ this relation,

$$\Pi'(\xi) = (1 + a\epsilon)\Pi(\xi) = \Pi\left(\frac{\xi}{1+\epsilon}\right),$$

we find as before a power law behavior $\Pi(\xi) \propto \xi^z$ for some exponent z. Now we need to assume the divergence of the correlation length $\xi \sim t^{-\nu}$ at the critical point to find

$$P(t) \sim t^X$$

for $X = -z\nu$, and this is exactly the power law behavior in critical systems.

In general, we may also expect that scale invariance would imply a trivial scaling of the free energy density as well. In fact, the precise scaling hypothesis rests on the assumption that the singular part of the free energy is a generalized homogeneous function, that is a function $f_{\text{sing}}(x_1, x_2, \dots)$ such that

$$f_{\text{sing}}(\lambda^{s_1} x_1, \lambda^{s_2} x_2, \dots) = \lambda f_{\text{sing}}(x_1, x_2, \dots)$$

which implies that the form of the free energy doesn't change. Since Legendre transforms and partial derivatives of a generalized homogeneous function are still generalized homogeneous functions, this is true for all the physical observables that are expressible as derivatives of the free energy or of a thermodynamic potential in general. Moreover, since λ is arbitrary to begin with we could choose $\lambda = |x_1|^{-1/s_1}$ to get

$$f_{\text{sing}}(x_1, x_2, \dots) = |x_1|^{1/s_1} \tilde{f}_{\text{sing}}(1, |x_1|^{-s_2/s_1} x_2, \dots)$$

thus reducing the number of the arguments. Among the other things, the scaling approach is extremely useful since it allows one to derive a number of relations between the exponents, the so-called scaling relations, by choosing a suitable λ and looking at the

critical point $t = 0$. Once again, the precise form of the scaling function is determined by the control by the fixed point of the Renormalization Group flow. However, the scaling hypothesis that ξ is the only relevant scale and the power law behaviors previously introduced are often enough to build a phenomenological scaling theory, as we will see in the dynamical case.

Albeit simple and qualitative, these arguments hint at how criticality, scaling and power laws are deeply related to one another. However, it may still seem that criticality is fundamentally a problem of fine tuning of some external parameters to the neighborhood of the critical point. The first question one should ask is how (and why) natural systems should be poised at criticality: we present a possible way to tackle this problem in the next section.

1.2 SELF-ORGANIZED CRITICALITY

Many natural systems, and in particular biological ones, appear to bring the signature of criticality, even if there is no apparent reason for such a parameter fine tuning. Thus we should think of criticality as an emergent property of these systems, rather than an a priori condition on them: it is their dynamical evolution that spontaneously leads them to a self-organized critical state. This property is usually referred to as self-organized criticality (SOC) and it was firstly introduced by Per Bak and his collaborators [1, 2].

The SOC mechanism is essentially a feedback loop where the control parameter itself becomes a dynamical variable, such that it is always pushed towards its critical value. This is well understood in terms of absorbing phase transitions, which we will extensively review in the next chapter, where the critical point separates an absorbing quiescent phase and an active one. When the control parameter lies within the active phase dissipation drives it towards lower values, whereas in the absorbing phase a driving force is responsible for its increase. An example of a model that displays SOC as a global dynamical feature is the self-organized branching process, which we will introduce at the end of this chapter.

On the other hand, one may also wonder why self-organized criticality should emerge in the first place. As we have seen response functions diverge at the critical point leading to a very high sensitivity to environmental changes and functional advantages for signal detection. Thus criticality implies a maximal dynamical range, and large correlations may be exploited as well to induce coordinated behavior of individual parts. In general, criticality allows a system to explore a wider spectrum of possible responses and leads to complexity. Therefore we should regard it as an optimal strategy for adaptive systems.

In particular, we now describe a general mathematical model that, rather than tackling the dynamical aspect of SOC, shows why criticality may emerge spontaneously as a consequence of the interaction of a complex adaptive system with environmental conditions [23]. When a living system needs to cope with its environment the most efficient state appears to be a critical state.

1.2.1 AN INFORMATION-THEORETIC APPROACH

For the sake of simplicity, let us introduce an environmental source as a set of N variables $\mathbf{s} = (s_1, \ldots, s_N)$ with $s_i = \pm 1$, $\forall i$. Sources are distributed with a probability distribution $P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})$ that depends on some environmental parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_E)$. The environmental variability is captured by a probability density $\rho_{\text{src}}(\boldsymbol{\alpha})$ that accounts for variable external sources. We may regard ρ_{src} as the probability to encounter a source with a particular set of parameters $\boldsymbol{\alpha}$.

We shall quantify the information loss by means of the Kullback-Leibler divergence of a probability distribution $Q(\mathbf{s})$ from another probability distribution $P(\mathbf{s})$,

$$D(P(\cdot) \,||\, Q(\cdot)) := \sum_{\mathbf{s}} P(\mathbf{s}) \log \frac{P(\mathbf{s})}{Q(\mathbf{s})}. \tag{1.8}$$

The KL divergence is clearly non-negative and vanishes if and only if the two distributions are equal, but it is not symmetric since it properly quantifies the loss of information when Q is used to approximate P.

We are also interested in the amount of information encoded in the state $\mathbf{s}$ about the parameters $\boldsymbol{\alpha}$. This is measured by the Fisher information

$$\chi^{\mu\nu}(\boldsymbol{\alpha}) = \left\langle \frac{\partial P(\cdot|\boldsymbol{\alpha})}{\partial \alpha_\mu} \frac{\partial P(\cdot|\boldsymbol{\alpha})}{\partial \alpha_\nu} \right\rangle_{\boldsymbol{\alpha}} \tag{1.9}$$

where the average $\langle \cdot \rangle_{\boldsymbol{\alpha}}$ is performed with respect to $P(\cdot|\boldsymbol{\alpha})$.

For convenience, we parametrize the distribution of the source distribution as

$$P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha}) = \frac{e^{-H_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})}}{Z_{\text{src}}(\boldsymbol{\alpha})}$$

with

$$H_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha}) = \sum_{\mu=1}^{E} \alpha_\mu \phi^\mu_{\text{src}}(\mathbf{s})$$

for some suitable functions $\phi^\mu_{\text{src}}(\mathbf{s})$. Note that $Z_{\text{src}}(\boldsymbol{\alpha}) = \sum_{\mathbf{s}} e^{-H_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})}$ is just the proper normalization.

Now we need to introduce an individual living system or "agent" whose scope is to adapt to respond to the signals from an environmental source. It does so by changing its internal state, described by a different probability distribution P_{int} that depends on a different parameter set $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_I)$. We choose the same parametrization

$$P_{\text{int}}(\mathbf{s}|\boldsymbol{\beta}) = \frac{e^{-H_{\text{int}}(\mathbf{s}|\boldsymbol{\beta})}}{Z_{\text{int}}(\boldsymbol{\beta})}$$

$$H_{\text{int}}(\mathbf{s}|\boldsymbol{\beta}) = \sum_{\mu=1}^{I} \beta_\mu \phi^\mu_{\text{int}}(\mathbf{s})$$

for some suitable functions $\phi^{\mu}_{\text{int}}(\mathbf{s})$.

What we look for is the optimal choice of $\boldsymbol{\beta}$ that an agent can choose for a given $\boldsymbol{\alpha}$, and assume that this translates to the fact that an agent wants to minimize its KL divergence in adapting to the environment. The Kullback-Leibler divergence of $P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})$ from $P_{\text{int}}(\mathbf{s}|\boldsymbol{\beta})$ is defined as

$$D(\boldsymbol{\alpha}||\boldsymbol{\beta}) := D(P_{\text{src}}(\cdot|\boldsymbol{\alpha}) \,||\, P_{\text{int}}(\cdot|\boldsymbol{\beta})) = \sum_{\mathbf{s}} P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha}) \log \frac{P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})}{P_{\text{int}}(\mathbf{s}|\boldsymbol{\beta})}$$

and it measures the information loss when P_{int} is used to approximate P_{src}. Using the parametrization introduced above, we find

$$\begin{aligned} D(\boldsymbol{\alpha}||\boldsymbol{\beta}) &= \sum_{\mathbf{s}} P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha}) \left[\log P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha}) - \log P_{\text{int}}(\mathbf{s}|\boldsymbol{\beta})\right] \\ &= \sum_{\mathbf{s}} P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha}) \log P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha}) + \log Z_{\text{int}}(\boldsymbol{\beta}) + \sum_{\mathbf{s}} P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha}) H_{\text{int}}(\mathbf{s}|\boldsymbol{\beta}). \end{aligned}$$

If we introduce

$$\left\langle \phi^{\mu}_{\text{int}} \right\rangle_{\boldsymbol{\alpha}} := \sum_{\mathbf{s}} \phi^{\mu}_{\text{int}}(\mathbf{s}) P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})$$

$$S_{\text{src}}(\boldsymbol{\alpha}) := -\sum_{\mathbf{s}} P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha}) \log P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})$$

then the KL divergence is simply

$$D(\boldsymbol{\alpha}||\boldsymbol{\beta}) = -S_{\text{src}}(\boldsymbol{\alpha}) + \log Z_{\text{int}}(\boldsymbol{\beta}) + \sum_{\mu=1}^{I} \beta_{\mu} \left\langle \phi^{\mu}_{\text{int}} \right\rangle_{\boldsymbol{\alpha}}.$$

Similarly, the Fisher information may be computed by means of

$$\begin{aligned} \frac{\partial \log P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})}{\partial \alpha_{\mu}} &= \frac{1}{P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})} \frac{\partial}{\partial \alpha_{\mu}} \left[\frac{e^{-\sum_{\gamma} \alpha_{\gamma} \phi^{\gamma}_{\text{src}}(\mathbf{s})}}{Z_{\text{src}}(\boldsymbol{\alpha})} \right] \\ &= -\phi^{\mu}_{\text{src}}(\mathbf{s}) + \frac{1}{Z_{\text{src}}(\boldsymbol{\alpha})} \sum_{\mathbf{s}'} e^{-\sum_{\gamma} \alpha_{\gamma} \phi^{\gamma}_{\text{src}}(\mathbf{s}')} \phi^{\mu}_{\text{src}}(\mathbf{s}') \\ &= \langle \phi^{\mu}_{\text{src}} \rangle_{\boldsymbol{\alpha}} - \phi^{\mu}_{\text{src}}(\mathbf{s}) \end{aligned}$$

which gives

$$\chi^{\mu\nu}_{\text{src}}(\boldsymbol{\alpha}) = -\frac{\partial}{\partial \alpha \mu} \langle \phi^{\nu}_{\text{src}} \rangle_{\boldsymbol{\alpha}} = \langle \phi^{\mu}_{\text{src}} \phi^{\nu}_{\text{src}} \rangle_{\boldsymbol{\alpha}} - \langle \phi^{\mu}_{\text{src}} \rangle_{\boldsymbol{\alpha}} \langle \phi^{\nu}_{\text{src}} \rangle_{\boldsymbol{\alpha}}.$$

This is nothing but the generalized susceptibility of the sources, which measures their response to variations of the parameters $\boldsymbol{\alpha}$. Of course, we can introduce the susceptibility of the system $\chi^{\mu\nu}_{\text{src}}(\boldsymbol{\beta})$ as well, which is a far more interesting quantity since its the system that adapts to the environment.

1.2.2 ADAPTING TO THE ENVIRONMENT

We assume that the optimal choice of parameters $\boldsymbol{\beta}$ is given by the lowest average KL divergence from the parameters $\boldsymbol{\alpha}$, that is

$$\boldsymbol{\beta}_{\text{opt}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \left[\int d\boldsymbol{\alpha} D(\boldsymbol{\alpha}||\boldsymbol{\beta}) \rho_{\text{src}}(\boldsymbol{\alpha}) \right] := \underset{\boldsymbol{\beta}}{\text{argmin}} \left[d(\rho_{\text{src}}|\boldsymbol{\beta}) \right]$$

so that the agent adopts an internal state that better describes the varying environment on average. To be more precise, we may describe the average environment by means of

$$\bar{P}_{\text{src}}(\mathbf{s}|\rho_{\text{src}}) := \int d\boldsymbol{\alpha}\, \rho_{\text{src}}(\boldsymbol{\alpha}) P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})$$

and write the KL divergence

$$\begin{aligned} D(\bar{P}(\cdot|\rho_{\text{src}}) \,||\, P_{\text{int}}(\cdot|\boldsymbol{\beta})) &= \sum_{\mathbf{s}} \bar{P}_{\text{src}}(\mathbf{s}|\rho_{\text{src}}) \left[\log \bar{P}_{\text{src}}(\mathbf{s}|\rho_{\text{src}}) - \log P_{\text{int}}(\mathbf{s}|\boldsymbol{\beta}) \right] \\ &= \sum_{\mathbf{s}} \bar{P}_{\text{src}}(\mathbf{s}|\rho_{\text{src}}) \left[\log \bar{P}_{\text{src}}(\mathbf{s}|\rho_{\text{src}}) + \log \frac{P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})}{P_{\text{int}}(\mathbf{s}|\boldsymbol{\beta})} - \log P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha}) \right] \\ &= \int d\boldsymbol{\alpha}\, \rho_{\text{src}}(\boldsymbol{\alpha}) \left[D(\boldsymbol{\alpha}||\boldsymbol{\beta}) - D(P_{\text{src}}(\cdot|\boldsymbol{\alpha}) \,||\, \bar{P}(\cdot|\rho_{\text{src}})) \right] \\ &= d(\rho_{\text{src}}|\boldsymbol{\beta}) - \int d\boldsymbol{\alpha} \rho_{\text{src}}(\boldsymbol{\alpha}) D(P_{\text{src}}(\cdot|\boldsymbol{\alpha}) \,||\, \bar{P}(\cdot|\rho_{\text{src}})). \end{aligned}$$

Notice that the last term on the right hand side is independent on $\boldsymbol{\beta}$.

We can now find the stationary points of

$$d(\rho_{\text{src}}|\boldsymbol{\beta}) = D(\bar{P}(\cdot|\rho_{\text{src}}) \,||\, P_{\text{int}}(\cdot|\boldsymbol{\beta})) + \int d\boldsymbol{\alpha}\, \rho_{\text{src}}(\boldsymbol{\alpha}) D(P_{\text{src}}(\cdot|\boldsymbol{\alpha}) \,||\, \bar{P}(\cdot|\rho_{\text{src}}))$$

which has a more clear interpretation: the first contribution is from the KL divergence of the internal distribution from the average environment; the second contribution is from the average KL divergence of the average environment from the real source distribution. Thus we need to solve

$$\frac{\partial}{\partial \beta_\nu} d(\rho_{\text{src}}|\boldsymbol{\beta}) = 0 = \int d\boldsymbol{\alpha}\, \rho_{\text{src}}(\boldsymbol{\alpha}) \frac{\partial}{\partial \beta_\nu} \left[\beta_\mu \left\langle \phi^\mu_{\text{int}} \right\rangle_{\boldsymbol{\alpha}} + \log Z_{\text{int}}(\boldsymbol{\beta}) \right]$$

where for the sake of brevity we started using the Einstein convention for repeated indexes. Since

$$\frac{\partial}{\partial \beta_\nu} Z_{\text{int}}(\boldsymbol{\beta}) = -\sum_{\mathbf{s}} \phi^\nu_{\text{int}}(\mathbf{s}) e^{-H_{\text{int}}(\mathbf{s}|\boldsymbol{\beta})}$$

the solution yields immediately

$$\left\langle \phi^\mu_{\text{int}} \right\rangle_{\boldsymbol{\beta}} = \int d\boldsymbol{\alpha} \rho_{\text{src}}(\boldsymbol{\alpha}) \left\langle \phi^\mu_{\text{int}} \right\rangle_{\boldsymbol{\alpha}}. \tag{1.10}$$

Is it a minimum? The Hessian matrix is given by

$$\begin{aligned}\frac{\partial^2}{\partial\beta_\mu\partial\beta_\nu}d(\rho_{\text{src}}|\boldsymbol{\beta}) &= -\frac{\partial}{\partial\beta_\mu}\left\langle\phi^\nu_{\text{int}}\right\rangle_{\boldsymbol{\beta}} = -\frac{\partial}{\partial\beta_\mu}\left[\frac{1}{Z_{\text{int}}(\boldsymbol{\beta})}\sum_{\mathbf{s}}e^{-\beta_\gamma\phi^\gamma_{\text{int}}(\mathbf{s})}\phi^\nu_{\text{int}}(\mathbf{s})\right]\\ &= \left\langle\phi^\mu_{\text{int}}\phi^\nu_{\text{int}}\right\rangle_{\boldsymbol{\beta}} - \left\langle\phi^\mu_{\text{int}}\right\rangle_{\boldsymbol{\beta}}\left\langle\phi^\nu_{\text{int}}\right\rangle_{\boldsymbol{\beta}}\\ &= \chi^{\mu\nu}_{\text{int}}(\boldsymbol{\beta})\end{aligned}$$

which is a positive definite matrix, so the solution (1.10) is indeed a minimum.

Equation (1.10) does have some interesting implication regarding criticality. Recall that we expect the generalized susceptibility of the internal system

$$\chi^{\mu\nu}_{\text{int}}(\boldsymbol{\beta}) = \left\langle\phi^\mu_{\text{int}}\phi^\nu_{\text{int}}\right\rangle_{\boldsymbol{\beta}} - \left\langle\phi^\mu_{\text{int}}\right\rangle_{\boldsymbol{\beta}}\left\langle\phi^\nu_{\text{int}}\right\rangle_{\boldsymbol{\beta}}$$

to diverge at the critical point in the thermodynamic limit $N\to\infty$. In a finite system, criticality therefore corresponds to the peak of χ. However, so far it may not be clear why (1.10) should lead to choose an internal state that is near criticality. To show this, we choose a narrow distribution $\rho_{\text{src}}(\boldsymbol{\alpha})$ that is non zero only in a region U and whose average value $\bar{\boldsymbol{\alpha}}$ is at some distance from criticality, and ask ourselves how this distance changes when the optimal internal representation is built: is it anywhere closer to the critical point, i.e. to the maximum of the susceptibility?

For the sake of brevity, we set $I=E$ so that $\phi^\mu_{\text{src}}=\phi^\mu_{\text{int}}=\phi^\mu$. Consider the expansion around $\bar{\boldsymbol{\alpha}}$

$$\left\langle\phi^\mu\right\rangle_{\boldsymbol{\alpha}} = \left\langle\phi^\mu\right\rangle_{\bar{\boldsymbol{\alpha}}} + (\alpha_\nu-\bar\alpha_\nu)\chi^{\nu\mu}(\bar{\boldsymbol{\alpha}}) + \frac{1}{2}(\alpha_\gamma-\bar\alpha_\gamma)(\alpha_\sigma-\bar\alpha_\sigma)\chi^{\gamma\sigma\mu}(\bar{\boldsymbol{\alpha}})$$

where we introduced the shorthand notation $\chi^{\mu\nu\sigma}:=\partial_\mu\chi^{\nu\sigma}=\partial_\mu\partial_\nu\left\langle\phi^\sigma\right\rangle_{\boldsymbol{\alpha}}$. Note that $\chi^{\mu\nu\sigma}$ is symmetric under permutations of all of its indexes. Similarly, we may write

$$\left\langle\phi^\mu\right\rangle_{\boldsymbol{\beta}} = \left\langle\phi^\mu\right\rangle_{\bar{\boldsymbol{\alpha}}} + (\beta_\nu-\bar\alpha_\nu)\chi^{\nu\mu}(\bar{\boldsymbol{\alpha}}).$$

Inserting these expansion in (1.10) we find

$$\begin{aligned}(\beta_\nu-\bar\alpha_\nu)\chi^{\nu\mu}(\bar{\boldsymbol{\alpha}}) &= \int_U d\boldsymbol{\alpha}\,\rho_{\text{src}}(\boldsymbol{\alpha})\left[(\alpha_\nu-\bar\alpha_\nu)\chi^{\nu\mu}(\bar{\boldsymbol{\alpha}}) + \frac{1}{2}(\alpha_\gamma-\bar\alpha_\gamma)(\alpha_\sigma-\bar\alpha_\sigma)\chi^{\gamma\sigma\mu}(\bar{\boldsymbol{\alpha}})\right]\\ &= \frac{1}{2}\int_U d\boldsymbol{\alpha}\,\rho_{\text{src}}(\boldsymbol{\alpha})(\alpha_\gamma-\bar\alpha_\gamma)(\alpha_\sigma-\bar\alpha_\sigma)\chi^{\gamma\sigma\mu}(\bar{\boldsymbol{\alpha}})\end{aligned}$$

which yields the distance of the optimal parameters from $\bar{\boldsymbol{\alpha}}$

$$\beta_\mu-\bar\alpha_\mu = (\chi^{-1})^{\mu\nu}(\bar{\boldsymbol{\alpha}})\,\chi^{\gamma\sigma\nu}(\bar{\boldsymbol{\alpha}})\int_U d\boldsymbol{\alpha}\rho_{\text{src}}(\boldsymbol{\alpha})(\alpha_\gamma-\bar\alpha_\gamma)(\alpha_\sigma-\bar\alpha_\sigma).$$

To write it in a vectorial form we introduce the scalar field

$$\Omega(\boldsymbol{\alpha}) := \frac{1}{2}\int_U d\boldsymbol{\alpha}'\,\rho_{\text{src}}\,(\boldsymbol{\alpha}')(\boldsymbol{\alpha}'-\boldsymbol{\alpha})^T\chi(\boldsymbol{\alpha}')(\boldsymbol{\alpha}'-\boldsymbol{\alpha})$$

so that

$$\boldsymbol{\beta} - \bar{\boldsymbol{\alpha}} = \chi^{-1}(\bar{\boldsymbol{\alpha}})\nabla\Omega(\bar{\boldsymbol{\alpha}}). \tag{1.11}$$

What can we say about the gradient of $\Omega(\boldsymbol{\alpha})$? Since $\chi(\boldsymbol{\alpha})$ is a positive matrix, $\Omega(\boldsymbol{\alpha})$ is always positive and therefore it has a maximum at the critical point, which means that its gradient points to the critical point as well. But then the projection of $(\boldsymbol{\beta} - \bar{\boldsymbol{\alpha}})$ over $\nabla\Omega(\bar{\boldsymbol{\alpha}})$, namely

$$(\Omega(\bar{\boldsymbol{\alpha}}))^T \cdot (\boldsymbol{\beta} - \bar{\boldsymbol{\alpha}}) = (\Omega(\bar{\boldsymbol{\alpha}}))^T \cdot \chi^{-1}(\bar{\boldsymbol{\alpha}})\nabla\Omega(\bar{\boldsymbol{\alpha}}),$$

is positive as well, which implies that the internal parameters $\boldsymbol{\beta}$ are closer to the critical point than $\bar{\boldsymbol{\alpha}}$.

This argument shows that under the assumption that a natural systems changes its own internal state $P_{\text{int}}(\mathbf{s}|\boldsymbol{\beta})$ to resemble as closely as possible the environmental signal $P_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha})$ the optimal choice of $\boldsymbol{\beta}$ is at criticality. In other words, the fittest possible agent is the one whose internal susceptibility χ is maximum. Intuitively this is not so unreasonable: in an ordered phase a system hardly changes in response to external stimuli, whereas in a disordered phase noise destroys long range correlations. Criticality represents a trade-off that allows flexibility and plasticity together with long-range correlations and extreme sensitivity to external stimuli.

1.3 ARE THERE SYSTEMS POISED AT CRITICALITY?

In recent years growing evidences of signatures of criticality have been found in systems of very different nature, with particular developments in the investigation of biological systems and biological networks. In this section we want to characterize some of these signatures and review some of the searches for criticality, to give a glimpse of the vast landscape of fields that touches upon.

1.3.1 ZIPF'S LAW AS A DISTRIBUTION AT A CRITICAL POINT

Historically, among the first power laws to be observed was Zipf's law in the human language [44]: it is a statement about the distribution over the states of the system, a state being a single word $\boldsymbol{\sigma}$. We shall show that Zipf's law carries some interesting properties with it and may be considered as a possible signature of criticality [4].

Zipf's law states that the frequency of words $P(\boldsymbol{\sigma})$ decays as the inverse of their rank $r(\boldsymbol{\sigma})$,

$$P(\boldsymbol{\sigma}) \propto \frac{1}{r(\boldsymbol{\sigma})} \tag{1.12}$$

where the rank $r(\boldsymbol{\sigma})$ is defined as the position in the frequency table, namely the more a word is frequent the higher is its rank. This distribution is not normalizable, so for further analysis we consider a slightly modified one

$$P(\boldsymbol{\sigma}) = \frac{r(\boldsymbol{\sigma})^{-\alpha}}{\zeta(\alpha)}, \quad \alpha > 1$$

where $\zeta(\alpha) = \sum_{n=1}^{\infty} n^{-\alpha}$ is the Riemann zeta function. Being a power law, from what we previously said it is worth to explore the relation between Zipf's law and criticality, if any.

We wish to look at the energy function $E(\boldsymbol{\sigma})$ that generates Zipf's law as en equilibrium distribution, that is

$$P(\boldsymbol{\sigma}) = \frac{1}{Z} e^{-E(\boldsymbol{\sigma})/k_B T}.$$

Without loss of generality, we can set $k_B T = 1$ and $Z = 1$ and simply write

$$E(\boldsymbol{\sigma}) = -\log P(\boldsymbol{\sigma}) = \alpha \log r(\boldsymbol{\sigma}) + \log \zeta(\alpha). \tag{1.13}$$

Since states that are more frequent are also of lower energy, in this setting the rank of $\boldsymbol{\sigma}$ is exactly the cumulative density of states at the corresponding energy $E(\sigma)$, namely

$$r(\boldsymbol{\sigma}) = \mathcal{N}[E = E(\boldsymbol{\sigma})] = \sum_{\boldsymbol{\sigma}'} \mathbb{I}[E(\boldsymbol{\sigma}') \leq E(\boldsymbol{\sigma})]$$

where $\mathbb{I}(\cdot)$ is the indicator function. To say it in words, the rank of a given state $\boldsymbol{\sigma}$ with energy $E(\boldsymbol{\sigma})$ is given by how many states there are with lower energy. For instance, the most probable word has the lowest energy, hence it is of rank 1. The second most probable word is of rank 2, and so on and so forth.

In general, we can introduce the density of states over a small energy shell

$$\rho_{\delta E}(E) = \frac{1}{\delta E} \sum_{\boldsymbol{\sigma}} \mathbb{I}[E < E(\boldsymbol{\sigma}) < E + \delta E]$$

so that we can write

$$\begin{aligned} \mathcal{N}(E) &= \int_{-\infty}^{E} dE' \rho_{\delta E=0}(E') \\ &= \int_{-\infty}^{E} dE' e^{Ns(E'/N)+s_1} \end{aligned}$$

where $Ns(E/N)$ is the entropy s_1 is sub-extensive, that is $\lim_{N\to\infty} s_1/N = 0$. This integral can be evaluated by saddle-point as $\mathcal{N}(E) \sim \exp[Ns(\epsilon)]$ with $\epsilon = E/N$ for large systems, yielding

$$\log r = \log \mathcal{N}(E) \sim Ns(E/N) = S(E).$$

Inserting this relation into (1.13) we find a quite peculiar result

$$S(E) = \frac{E}{\alpha} + \dots \tag{1.14}$$

where $\dots$ is sub-extensive.

A linear dependence of the entropy on the energy implies $S''(E) = 0$ at all energies, so the specific heat

$$C(T) = \frac{N}{T^2}\left[-\frac{\mathrm{d}^2 S(E)}{\mathrm{d}E^2}\right]^{-1}$$

is always divergent. Therefore we shall regard Zipf's law as a possible signature of criticality in the sense that it may describe a system whose statistical properties resemble critical ones. However, let us stress that this should be regarded more as a first step in the search for criticality rather than a definitive one [39].

1.3.2 MODELS FOR NATURAL SYSTEMS

Maximum entropy models can be a good starting point to build a statistical model to reproduce the data, especially for living systems for which finding a good estimate of $P(\boldsymbol{\sigma})$ is challenging. The idea is to introduce a model distribution $P_m(\boldsymbol{\sigma})$ such that the entropy

$$S[P] = -\sum_{\boldsymbol{\sigma}} P_m(\boldsymbol{\sigma}) \log P_m(\boldsymbol{\sigma})$$

is maximized with the constraints

$$\langle \mathcal{O}_i(\boldsymbol{\sigma}) \rangle_m = \langle \mathcal{O}_i(\boldsymbol{\sigma}) \rangle_r$$

where $\langle \cdot \rangle_m$ and $\langle \cdot \rangle_r$ are the averages with respect to the model and the real distribution $P_r(\boldsymbol{\sigma})$ respectively. That is, we wish to build the minimal distribution that agrees with some features of the data.

Maximizing the entropy yields

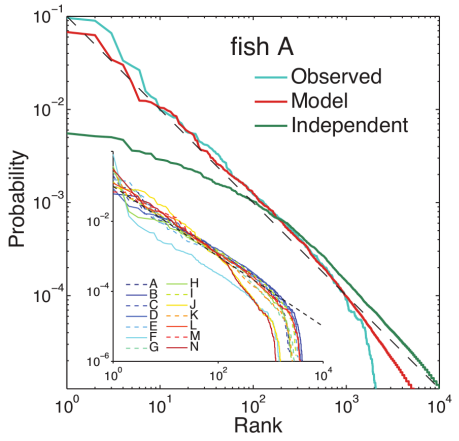
$$P_m(\boldsymbol{\sigma}) = \frac{1}{Z}\exp\left[\sum_i \beta^i \mathcal{O}_i(\boldsymbol{\sigma})\right]$$

where β^i is the Lagrange multiplier associated to the i-th constraint. In fact, if we write the KL divergence of P_m from P_r

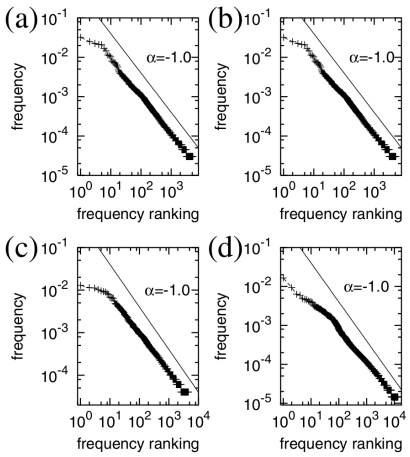
$$D(P_r||P_m) = \sum_{\boldsymbol{\sigma}} P_r(\boldsymbol{\sigma}) \log \frac{P_r(\boldsymbol{\sigma})}{P_m(\boldsymbol{\sigma})}$$

and we minimize with respect to β_a we find

$$\begin{aligned}
\frac{\partial D(P_r||P_m)}{\partial \beta_a} &= \sum_{\boldsymbol{\sigma}} P_r(\boldsymbol{\sigma}) \left[\frac{1}{Z}\frac{\partial Z}{\partial \beta_a} - \frac{\partial}{\partial \beta_a}\sum_i \beta^i \mathcal{O}_i(\boldsymbol{\sigma})\right] \\
&= \sum_{\boldsymbol{\sigma}} P_r(\boldsymbol{\sigma}) \left[\langle \mathcal{O}_a \rangle_m - \mathcal{O}_a(\boldsymbol{\sigma})\right] \\
&= \langle \mathcal{O}_a \rangle_m - \langle \mathcal{O}_a \rangle_r = 0.
\end{aligned}$$



(A) Results from [35].



(B) Results from [17].

FIGURE 1.5: (A) The maximum entropy model with pairwise correlations for the expression of the antibody in B cells predicts a Zipf's distribution which is verified in experimental data from different samples (the different letters represent different zebrafish used in the experiment). Note how pairwise correlations are needed to obtain Zipf's law (the dotted line) and correctly fit the data. (B) Zipf's law is also found in rank ordered frequency distributions of expressed genes, from [17]. In particular: (a) human liver, (b) kidney, (c) human colorectal cancer, (d) mouse embryonic stem cell.

Hence our choice is the one that minimizes the KL divergence between the real and the model distributions, and in the framework of maximum entropy models it is indeed possible to explore statistical properties starting from experimental data.

It is worth noting that this method makes sense if the observables $\mathcal{O}_i$ are chosen in such a way that their averages $\langle \mathcal{O}_i \rangle_r$ can be estimated accurately from the data. Thus a natural choice is given by the averages $\langle \sigma_i \rangle$ or $\langle \sigma_i \sigma_j \rangle$, which means requiring consistence up to pairwise correlations.

One example is given by Zipf's law, found in the gene regulatory network that expresses antibody in cells of the immune system [35]. The state of the system $\boldsymbol{\sigma}$ is defined as $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_L)$ where L is the length of the genetic sequence and σ_i is one of the possible twenty amino acids. The maximum entropy distribution resembles the form of a disordered Potts model

$$P(\boldsymbol{\sigma}) = \frac{1}{Z} \exp \left[\sum_i h_i(\sigma_i) + \sum_{ij} J_{ij}(\sigma_i, \sigma_j) \right]$$

that is consistent with two point correlation functions. In Figure 1.5 the probability density over state obtained in [35] is reported. The model correctly predicts that $P(\boldsymbol{\sigma})$ follows a Zipf's distribution once plotted in terms of the rank of each state. Interestingly enough, similar results were found in data from the most disparate gene expression without building an explicit statistical model, and a possible explanation for the emergence of Zipf's law was proposed in terms of a critical point in underlying the diffusion process of environmental nutrients [17]. In general, Zipf's law emerges in

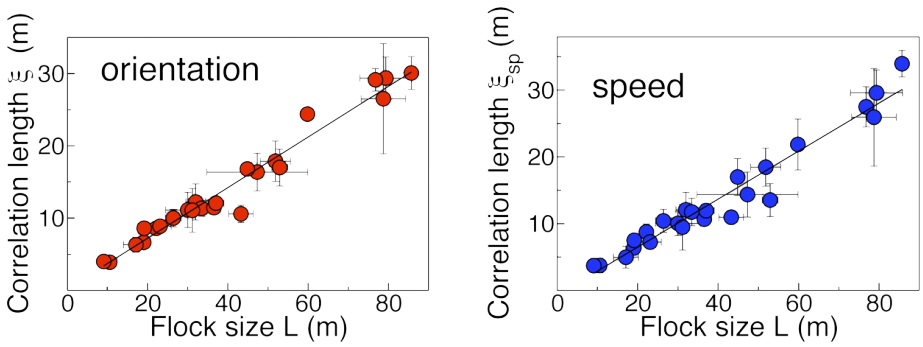


FIGURE 1.6: Results from [10]. On the left the dependence of the correlation length of the orientation of the velocities on the flock size. On the right the correlation length of the modulus of the velocities with respect to the flock size. In both cases $\xi \propto L$.

very different contexts, such as in the distribution of the size of cities around the world [32].

A vast variety of more explicit examples of criticality exists [36], like in models of cells growth [16] or in the collective motion of individuals, where even if the interactions are short-range the group is able to react as a whole. The archetypal example is the one of flocks of birds [10]. It is indeed tempting to relate the highly coherent motion and the fast collective response to the statistical properties of a critical point, such as long range correlations and divergent susceptibility. Data from flocks of birds were analyzed through the correlation function of the velocity fluctuations

$$C(r) = \frac{1}{c_0} \frac{\sum_{ij} \mathbf{u}_i \cdot \mathbf{u}_j \, \delta(r - r_{ij})}{\sum_{ij} \delta(r - r_{ij})}$$

where $\mathbf{u}_i = \mathbf{v}_i - \langle \mathbf{v} \rangle$ is the fluctuation around the mean flock's velocity and r_{ij} is the distance between bird i and j. To be precise, this is the correlation between the orientation of the velocities, but one can and should consider the correlation between the modulus of the bird's velocities as well. Without entering into the details, this is due to the fact that the system is naturally invariant under the continuous symmetry group $O(3)$. This symmetry is spontaneously broken along the direction of the order parameter, the overall orientation of the flock. Hence the ground state is invariant under rotations transverse to the order parameter, and these are Goldstone modes that cost an arbitrarily small amount of energy. Like the spin wave approximation of the XY model, this leads to an algebraic decay of correlations between the orientation of the velocities instead of an exponential one. By studying the modulus of the velocities as well, which does not correspond to any particular symmetry of the model, one can distinguish between a scale-free behavior resulting from intrinsic Goldstone modes and one that arises from criticality instead.

The results are reported in Figure 1.6. In both cases the correlation length ξ is found to scale linearly with the system size L, and this has an important implication. In fact,

near a critical point we expect a scaling form for the correlation function

$$C(r) = \frac{1}{r^{\gamma}} f\left(\frac{r}{\xi}\right)$$

and if $\xi = bL$ then in the thermodynamic limit we indeed find

$$C(r) = r^{-\gamma}$$

which is a clear signature of criticality. Correlations span the entire system giving rise to an emergent collective behavior from simple short range interactions. Interestingly enough, in [5] a maximum entropy model that takes into account the mean velocity, the mean square velocity and the pairwise correlation was built. The model correctly reproduces the data and the linear dependence of ξ on the system size if the parameter that controls the fluctuations of the modulus of the velocities (i.e. the mean square velocity) is indeed tuned at criticality.

Among this and other many examples, the possibility that networks of neurons might operate at criticality is one of the most fascinating. The following section is devoted to introduce the main aspects of such hypothesis.

1.4 THE CRITICAL BRAIN HYPOTHESIS

The brain is possibly the most impressively complex system we are able to study. The idea that the collective behavior of networks of neurons might emerge from a self-organized critical state is tempting and comes from different approaches. To briefly make a point, large scale correlations would allow a coherent global response to external stimuli as well as an optimal information transmission, and at the same time the divergent susceptibility of the system would translate into a high sensitivity to sensory stimuli. From a computational point of view, the advantage is once again the trade-off between a disordered phase where perturbations and noise can propagate unboundedly (high information transmission, but low storage capacity) and an ordered phase where changes are rapidly erased (high storage capacity, but low information transmission). However, the possibility that the brain operates a criticality is still widely debated and it is a very much open question.

It is an experimental evidence that the cerebral cortex of mammalians is never silent but is rife with spontaneous activity instead (see [36] for a comprehensive list of references). The main approach to criticality in the brain is in fact the study of the spatiotemporal organization of these outbursts of spontaneous activity, called neuronal avalanches. Neural avalanches do appear to be scale-free, i.e. power law distributed [3], both in their size and in their duration, so the underlying process from which these avalanches emerge seems to be scale invariant and therefore critical. In particular, the exponents corresponds to the ones of a mean-field branching process, which we will introduce further in this section. Criticality would therefore be found in a phase transition between a completely silent phase and a fully active phase in the resting state of the brain. Note that now we are properly referring to dynamical criticality, in

the sense that this is a non-equilibrium phase transition in the underlying dynamical process and not simply in the statistics of equilibrium states, as we will extensively see further in this thesis. On the other hand, one should note that a number of other dynamical processes have been proposed whose avalanches still show the same power law exponents but are not critical, such as in the framework of neutral theories [33].

There are approaches that focuses on statistical criticality as well, where the state of a network of neurons is defined by its spiking pattern, that is a set of binary variables that identifies whether each single neuron has spiked or remained silent. This would be the case for a standard thermodynamic phase transition such as the one of the Ising model. Of course, we shall remind that critical systems are remarkable because their large scale, collective behavior does not depend on the microscopic details of the system. Therefore the ultimate hope is that if the brain really is poised near a critical point this would allow us to describe its global properties without necessarily knowing the fine structure of its $\approx 10^{11}$ neurons and $\approx 10^{15}$ synapses, which is an almost impossible task.

Before introducing examples from these two approaches, let us note that most interestingly the signatures of criticality in experimental data appear to fade away when the brain is not in a wakeful state, such as during deep sleep and unconsciousness, where long-range temporal correlations vanish, or during epileptic seizures or anomalously large periods without sleep (once again, see [36] for a comprehensive list of references). While it is far from being a settled subject, the idea that the brain might work at a critical point is indeed tantalizing and worth exploring.

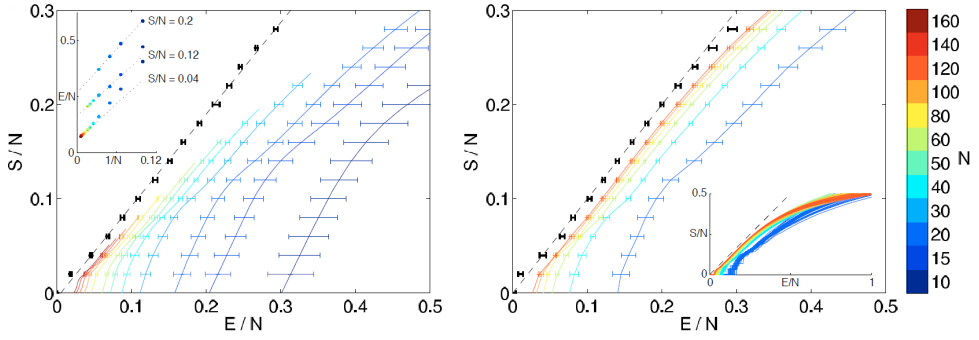
1.4.1 STATISTICAL CRITICALITY IN NETWORKS OF NEURONS

The simplest model of the activity of a neuron is essentially binary: in a sufficiently small window of time, a neuron i either spikes and we assign $\sigma_i = 1$ or does not, $\sigma_i = -1$. The spiking activity of N neurons is therefore described by the state $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)$ and we are interested in the probability distribution $P(\boldsymbol{\sigma})$ over all possible states. This is the equilibrium Statistical Mechanic approach: we are not interested in the dynamical process that leads to a particular state $\boldsymbol{\sigma}$, but rather in the statistics of such patterns.

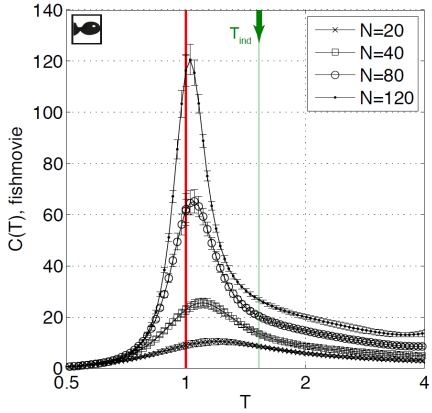
Starting from spiking patterns of a real network of neurons, a salamander retina patch, a maximum entropy model was build in [41]. If we only require consistence with the spiking average $\langle \sigma_i \rangle$, we find a distribution

$$P_1(\boldsymbol{\sigma}) = Z^{-1} \exp\left[\sum_i h_i \sigma_i\right]$$

which is a model of independent firing cells. Perhaps not surprisingly, the independent model fails to capture the global properties of the network even if correlations between any two neurons are typically weak. If we require consistence with pairwise



(A) Results from [41].



(B) Results from [41].

FIGURE 1.7: The pairwise maximum entropy model's predictions for a network of neurons. (A) On the left the dependence on the energy ofthe entropy estimated from the data. On the right the same plot predicted from the maximum entropy model. In both cases the dashed line corresponds to $S = E$. (B) The the heat capacity diverges for $T = 1$, which corresponds to the parameters of the real network.

correlations $\langle \sigma_i \sigma_j \rangle$ the maximum entropy model is a disordered Ising model,

$$P_2(\boldsymbol{\sigma}) = Z^{-1} \exp\left[-\sum_i h_i \sigma_i - \sum_{i<j} J_{ij} \sigma_i \sigma_j \right].$$

This model seems to be a good enough approximation since it is already a good predictor of the experimental three-point correlation function $\langle \sigma_i \sigma_j \sigma_k \rangle$ [4]. If we introduce

$$E(\boldsymbol{\sigma}) = -\log P(\boldsymbol{\sigma})$$

$$S(E) = \log n(E),$$

where $n(E)$ is the number of states at a given energy, the entropy dependence on the energy is linear both from direct computation and from prediction by the model. The results obtained in [41] are reported in Figure 1.7a. Therefore the distribution seems to

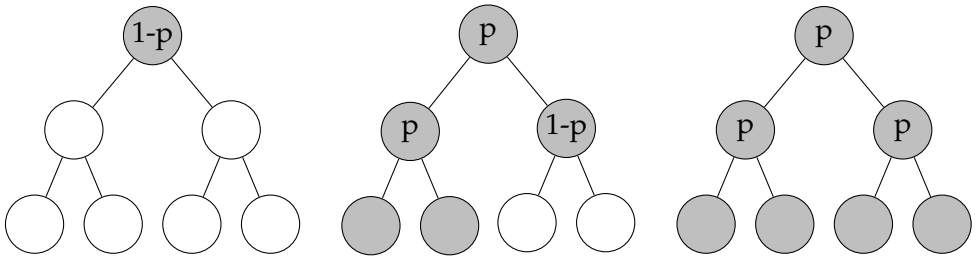


FIGURE 1.8: Some of the possible configurations of a branching process with $n = 2$, with avalanches of size $s = 1, 5, 7$ and boundary active sites $\sigma_2 = 0, 2, 4$.

be critical in the same sense as discussed with Zipf's law, that is

$$\frac{\mathrm{d}^n S}{\mathrm{d}E^n} = 0 \quad \forall n \geq 2$$

which implies a divergent heat capacity.

In a perhaps more convincing approach, it is possible to introduce a fictitious temperature and consider the one parameter family of models

$$P(\boldsymbol{\sigma}; T) = \frac{1}{Z(T)} \exp\left[-\frac{1}{T} E(\boldsymbol{\sigma})\right]$$

which means in particular rescaling the couplings of the maximum entropy model,

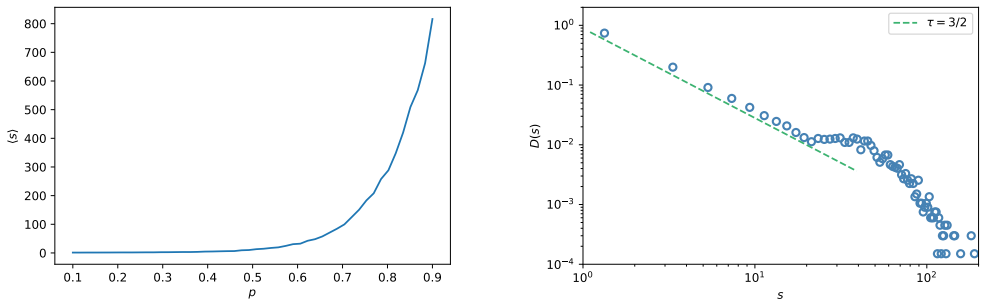
$$h_i \to \frac{h_i}{T}, \quad J_{ij} \to \frac{J_{ij}}{T}.$$

In practice, this means that changing the temperature implies changing the parameters, and it is found that only $T = 1$ describes the ones of the real network. A more clear signature of criticality is then the predicted heat capacity, which does indeed peak as $T \to 1$ as shown in Figure 1.7b. Moreover, the peak becomes more pronounced as N increases.

1.4.2 AVALANCHES IN BRANCHING PROCESSES AND DYNAMICAL CRITICALITY

Beyond statistical criticality, the most interesting features are indeed the ones of the dynamical process that underlies the spiking patterns, hence considering it as a non-equilibrium, time dependent system. This is the case for the spontaneous activity and the spatiotemporal organization of neuronal avalanches that appear to be scale-free.

A simple toy model for spontaneous activity and scale-free avalanches is the branching process. An initial active site either relaxes with probability p and generates two new active sites, or it does not relax with probability $1 - p$ and leads to two inactive sites. This is called the first generation of the process. We iterate this procedure for each active site, that can either generate two new active sites with probability p or none with probability $1 - p$, until no active site remains. We define the size of an avalanche s as the number of active sites when the process dies or reaches the boundary, see Figure 1.8.



(A) Phase transition in the branching process.

(B) Scaling of the avalanche distribution.

FIGURE 1.9: Simulation of a branching process with $n = 10$ generations. (A) Evolution of the order parameter $\langle s \rangle$ in terms of the control parameter p. (B) Scaling of the avalanche distribution $D(s)$.

It turns out that in the thermodynamic limit $n \to \infty$ it exists a critical probability p_c above which the probability to have an infinite avalanche is non vanishing, that is the system displays a second order phase transition with the average size of an avalanche $\langle s \rangle$ as the order parameter. In Figure 1.9a we show the results of a simulation for $n = 10$ and different value of p. Even if the number of generations is small and therefore the transition is not sharp, we can see how around $p_c = 1/2$ the behavior of the system changes.

The fact that we are at a critical point becomes more evident when we consider the avalanche distribution $D(s)$, which is evaluated in a simulation with $n = 10$ and reported in Figure 1.9b. We observe in fact a scaling region where $D(s)$ displays a power law behavior,

$$D(s) \sim s^{-\tau}$$

with $\tau = 3/2$, and exponential tails due to the finite size of the system.

We can be more quantitative. We introduce the probability $P_n(s, p)$ of having an avalanche of size s in a system of size n and the corresponding generating function

$$f_n(x, p) = \sum_s P_n(s, p) x^s.$$

Clearly, we have (see Figure 1.8 for reference)

$$f_0(x, p) = x, \quad f_1(x, p) = x(1 - p) + x^3 p,$$

$$f_2(x, p) = x(1 - p) + x^3 p(1 - p)^2 + x^5 2p^2(1 - p) + x^7 p^3$$

and in general the following recursion relation

$$f_{n+1}(x, p) = x \left[(1 - p) + p f_n^2(x, p) \right]$$

holds. In the large n limit $n \gg 1$ we expect $f_{n+1}(x,p) = f_n(x,p) = f(x,p)$ so that we can solve it to find

$$f(x,p) = \frac{1 - \sqrt{1 - 4x^2 p(1-p)}}{2xp}. \tag{1.15}$$

In order to determine the corresponding probability $P(s,p)$ we need to expand the generating function (1.15) as a power series in x. Recall that the binomial theorem implies

$$\sqrt{1+x} = \sum_{k=0}^{\infty} \binom{1/2}{k} x^k = 1 - \sum_{k=0}^{\infty} \frac{2}{k+1} \binom{2k}{k} \left(-\frac{x}{4}\right)^{k+1}$$

where we used the identity

$$\binom{1/2}{k} = \frac{(-1)^{k-1}}{2^k k!} \binom{2k-2}{k-1}.$$

Hence

$$f(x,p) = \sum_{k=0}^{\infty} \frac{1}{k+1} \binom{2k}{k} p^k (1-p)^{k+1} x^{2k+1}.$$

Identifying $s = 2k+1$ we can write

$$\begin{aligned} P(s,p) &= \frac{2}{s+1} \binom{s-1}{(s-1)/2} p^{(s-1)/2} (1-p)^{(s+1)/2} \\ &= \frac{2}{s+1} \binom{s-1}{(s-1)/2} \sqrt{\frac{1-p}{p}} \, [p(1-p)]^{s/2} \end{aligned}$$

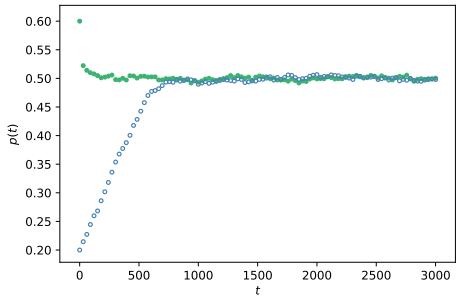
and now we shall consider the large s limit $1 \ll s \lesssim n$, that is we consider avalanches big enough but still neglect the finite size of the system. In such case the Stirling approximation gives

$$\binom{s-1}{(s-1)/2} \approx \frac{\sqrt{2\pi(s-1)}}{\pi(s-1)} \frac{[(s-1)/e]^{s-1}}{[(s-1)/2e]^{s-1}} = \frac{2^{s-1}}{\sqrt{\pi(s-1)/2}}$$

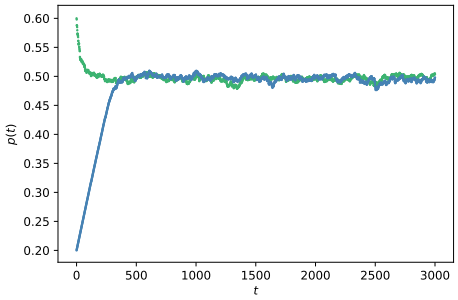
so that in this limit we find

$$\begin{aligned} P(s,p) &\approx \frac{1}{s} \sqrt{\frac{2(1-p)}{s\pi p}} \, [4p(1-p)]^{s/2} \\ &= s^{-3/2} \sqrt{\frac{2(1-p)}{\pi p}} \exp\left[-\frac{s}{s_c(p)}\right] \end{aligned}$$

where the cutoff $s_c(p)$ is given by $s_c(p) = -2/\log[4p(1-p)]$.



(A) Simulation of Equation 1.17.



(B) Simulation of Equation 1.16.

FIGURE 1.10: Simulation of a SOBP with $n = 10$ generations and 3000 timesteps. (A) The continuous time dynamics. (B) The discrete time SOBP. The system ends up in the stable fixed point $p = p_c$ and fluctuates around it, poising itself at criticality.

Notice that $s_c(p) \to \infty$ as $p \to p_c = 1/2$, and this kind of divergence is indeed characteristic of a second order phase transition. At criticality we find the power law behavior

$$D(s) = P(s, p_c) = \sqrt{\frac{2}{\pi}} s^{-\tau}$$

with $\tau = 3/2$, which is indeed what we find in the simulation in Figure 1.9b.

To better understand the finite size case, we introduce the random variable σ_n that describes the number of active sites in the n-th generation and its generating function

$$\varphi_n(x, p) = \sum_k \mathbb{P}[\sigma_n = k] x^k$$

and in particular

$$\varphi_1(x, p) = (1 - p) + x^2 p.$$

Due to the iterative nature of the process, we also have

$$\varphi_{n+1}(x, p) = \sum_k \mathbb{P}[\sigma_n = k][\varphi_1(x, p)]^k = \varphi_n[\varphi_1(x, p)]$$

thus the expectation value

$$\langle \sigma_{n+1} \rangle = \left. \frac{\partial \varphi_{n+1}(x, p)}{\partial x} \right|_{x=1} = \varphi_n'(1, p) \varphi_1'(1, p)$$

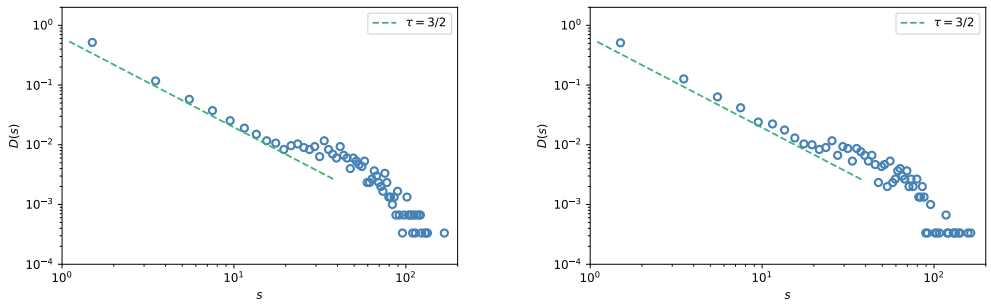
since $\varphi(1, p) = 1$. Therefore we have

$$\langle \sigma_1 \rangle = 2p$$

$$\langle \sigma_2 \rangle = (2p)^2$$

$$\vdots$$

$$\langle \sigma_n \rangle = (2p)^n.$$



(A) Avalanches for $p(0) = 0.6$. (B) Avalanches for $p(0) = 0.2$.

FIGURE 1.11: Simulation of a SOBP with $n = 10$ generations and 3000 timesteps. (A) The initial branching process is supercritical. (B) The initial branching process is subcritical. In both cases $D(s)$ is distributed as a critical branching process.

We expect the number of active sites at the boundary of the tree to be $\langle \sigma_n \rangle < 1$ in the subcritical case and $\langle \sigma_n \rangle > 1$ in the supercritical case. The critical case then is $\langle \sigma_n \rangle = 1$ for $p_c = 1/2$.

This simple model of a branching process shows how scale-free avalanches can emerge if the underlying dynamics is indeed critical. But we can do more: in fact, we can introduce the self-organized branching process (SOBP) as a toy model for self-organized criticality [43]. In general, in a subcritical or supercritical branching process at fixed p the avalanche distribution does not display a power law region thus it is not scale-free. Therefore we need a global dynamic for the control parameter $p(t)$ that spontaneously leads the branching process to criticality, regardless of the initial condition $p(0)$, in such a way that there is no need for fine tuning to its critical value p_c.

We consider once again the dynamics at fixed n, so that it exists a boundary for the system and boundary conditions must be specified. In particular, we say that if an avalanche reaches the boundary of the tree we have dissipation and therefore p should decrease in the next avalanche. On the other hand, if the avalanche stops before the boundary p will increase. If each timestep corresponds to a standard branching process, the dynamical equation for the SOBP reads

$$p(t+1) = p(t) + \frac{1 - \sigma_n(p,t)}{N} \tag{1.16}$$

where $N = 2^{n+1} - 1$ is the size of the tree and $\sigma_n(p,t)$ is the number of active sites at the boundary. In fact, if $\sigma_n = 0$ the probability increases and if $\sigma_n > 0$ the probability decreases, so the last term acts as a stochastic driving force. The corresponding continuous time process can be determined if we assume

$$\sigma_n(p,t) = \langle \sigma_n(p,t) \rangle + \eta(p,t)$$

where $\eta(p,t)$ is the white noise that describes the fluctuations around the average, hence

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{1 - (2p)^n}{N} + \frac{\eta(p,t)}{N}. \tag{1.17}$$

Without the noise term, $p_c = 1/2$ is a fixed point, which is exactly what we want from a self-organization to criticality. Hence we can linearize the dynamics around it by means of $p = p_c + \epsilon$ for $\epsilon \ll 1$,

$$\frac{\mathrm{d}\epsilon}{\mathrm{d}t} = \frac{1}{N}\left[-2\epsilon + \eta + O(\epsilon^2)\right]$$

which implies that the fixed point is indeed attractive. Thus Equation (1.17) spontaneously leads to criticality.

In Figure 1.10 the results of a simulation of both the discrete time and the continuous time SOBP are reported. We can see that after an initial transient the control parameter starts to fluctuate around the critical value $p_c = 1/2$ without any need for fine tuning. If we look at the avalanche distribution $D(s)$ in Figure 1.11, the SOPB displays scale-free avalanches regardless on the initial probability $p(0)$: this is an example of self-organized criticality.

CHAPTER 2

THE CONTACT PROCESS

Among the many, the contact process is probably one of the simplest non equilibrium models that can be used to describe the spreading of activity in a network of neurons. It was first introduced by Harris as a toy model for epidemic spreading without immunization [21]. Much like the Ising model for equilibrium systems, the contact process is also an archetypal example of a non equilibrium phase transition that arises as a consequence to its local dynamic, which is described as follows [22, 30, 31].

Consider for instance a d-dimensional lattice with N sites, or in general a collection of $i = 1, \ldots, N$ nodes of a given network. Each site can be either active or inactive, and we identify its state by means of a binary variable $\sigma_i(t) = 1,\ 0$ respectively[1]. The activity spreads via a nearest neighbors interaction, and it depends on the number of active neighbors,

$$n_i(t) = \sum_{j \in \langle i \rangle} \sigma_j(t),$$

whereas each active site is emptied at a given rate μ, regardless of its neighbors. Thus the rates $w[\sigma_i(t) \to \sigma_i(t + dt), n_i(t)]$ that define the process in a lattice of coordination number q are given by

$$w[0 \to 1, n_i] = \frac{\lambda n_i}{q} \tag{2.1}$$

$$w[1 \to 0, n_i] = \mu \tag{2.2}$$

where λ is the control parameter that represents the spreading rate, and usually we take $\mu = 1$, so that a site is emptied at a unitary rate. In Figure 2.1 the possible transition rates w are showed for the one-dimensional case.

[1] If we think of disease spreading, $\sigma_i(t) = 1$ represents an infected organism and $\sigma_i(t) = 0$ a healthy one. In general we will also refer to occupied sites ($\sigma_i = 1$) and empty sites $\sigma_i = 0$.

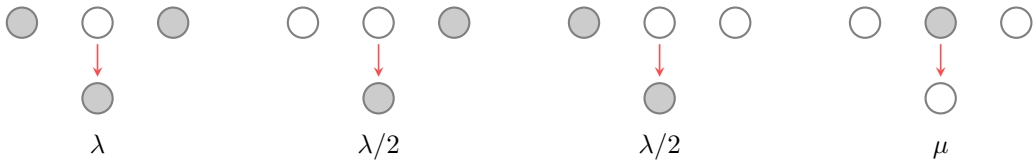


FIGURE 2.1: Transition rates in the one-dimensional contact process.

The nature of the phase transition is quite different from the equilibrium one, much like the case of the branching process introduced at the end of the previous chapter. In fact, we intuitively expect that it may exist a critical value of the control parameter $\lambda = \lambda_c$ that separates two qualitatively different global behaviors. In fact, for different values of λ what changes is the typical configuration in which the contact process finds itself at large times. If $\lambda > \lambda_c$ the spreading rate is high and a site does not recover fast enough, so at large times an active fluctuating phase permeates the system. On the other hand, if $\lambda < \lambda_c$ the activity is not able to spread fast enough and the system eventually finds itself in a state where all the sites are empty. From its definition, the local dynamical rules (2.1)-(2.2) do not allow to escape from such a configuration since $w[0 \to 1, 0] = 0$ for all sites: it is an absorbing configuration.

These two phases appear only in the long time limit, that is they are are stationary solutions of the model, and contrary to the equilibrium case the transition to the absorbing state is irreversible. Hence we say that the contact process displays an absorbing phase transition. It is at criticality that in the thermodynamic limit an infinite cluster of active sites appears, and the system becomes scale-free.

2.1 AN ABSORBING PHASE TRANSITION

In which sense do we say that the transition to an absorbing state is in fact a phase transition? Is it possible to translate some of what we now of criticality in equilibrium system in this time dependent framework?

If we are to identify a phase transition, we need to introduce an order parameter first. One possible choice is the density of active site $\rho(t)$ defined as

$$\rho(t) = \lim_{N\to\infty} \frac{1}{N} \left\langle \sum_{i=1}^{N} \sigma_i(t) \right\rangle \tag{2.3}$$

where the average $\langle\cdot\rangle$ is performed over many realizations of the stochastic process. This seems to be a good order parameter, since it is clearly zero in the absorbing state and is different from zero if activity is proliferating. Note that another possible choice exists [22], namely using the survival probability

$$P(t) = \left\langle 1 - \prod_{i=1}^{N} (1 - \sigma_i(t)) \right\rangle .$$

In fact, the quantity inside the average is equal to 1 as long as $\sigma_i(t) = 1$ for at least one site i, whereas it is vanishing if none of the sites is active. For the time being, we shall consider (2.3) as the order parameter, since it can be shown [22] that they are equivalent.

We know that in the equilibrium case correlations play an important role in the emergence of critical behavior. Since we are now interested in the time evolution of the system, we introduce both an equal time correlation function

$$c_\perp(r,t) = \langle \sigma_i(t)\sigma_{i+r}(t)\rangle - \langle \sigma_i(t)\rangle \langle \sigma_{i+r}(t)\rangle \tag{2.4}$$

and a time autocorrelation function

$$c_\parallel(t) = \langle \sigma_i(t)\sigma_i(0)\rangle - \langle \sigma_i(t)\rangle \langle \sigma_i(0)\rangle \tag{2.5}$$

so there will be a correlation length in space $\xi_\perp$ and a correlation length in time $\xi_\parallel$, and they are independent. Away from criticality, we expect that for large values of their arguments both these correlation functions decay exponentially

$$c_\perp(r,t) \sim e^{-r/\xi_\perp}, \quad c_\parallel(t) \sim e^{-t/\xi_\parallel}.$$

In analogy with the Ising model, for $\lambda < \lambda_c$ the former measures the typical extension in space of clusters of active sites, whereas the latter their typical decay time. For $\lambda > \lambda_c$ we expect that an infinite cluster exists, so $\xi_\perp$ and $\xi_\parallel$ measure the typical size and the typical duration of inactive islands, respectively.

An analogous of a pinning field exists as well. Suppose we introduce to the rates (2.1)-(2.2) the transition

$$w_h[0 \to 1] = h$$

so that we introduce the possibility of spontaneous activation, having in mind the case of the Ising model where adding a magnetic field is useful to determine the response of the system to an external perturbation. Of course, for $h \neq 0$ an absorbing configuration does not exist anymore, nor the phase transition. The susceptibility that corresponds to this rate is

$$\chi = \frac{\partial \rho(t)}{\partial h} \tag{2.6}$$

and we are interested in the limit $h \to 0$.

As we expect in a phase transition, power laws appear at criticality because the infinite cluster of active sites is scale-free. However, we should stop for a moment and consider that what we are really interested in is the asymptotic behavior of the system, that is the steady state at $t \to \infty$. In such a way that we ignore the initial transient that depends on the initial configuration and on the microscopic details of the lattice. In fact, it is in this limit that we distinguish between an absorbing phase and an active phase and we can effectively introduce critical exponents. Implicitly, we will consider this limit from now on.

In particular, we shall verify the stationary density of active sites to scale as

$$\rho \sim |\lambda - \lambda_c|^{\beta} \tag{2.7}$$

much like the order parameter of an equilibrium system. Two different exponents exist for the correlation lengths,

$$\begin{aligned} \xi_{\perp} &\sim |\lambda - \lambda_c|^{-\nu_{\perp}} \\ \xi_{\parallel} &\sim |\lambda - \lambda_c|^{-\nu_{\parallel}} \end{aligned} \tag{2.8}$$

so that both diverge at the critical point. The susceptibility is divergent as well,

$$\chi \sim |\lambda - \lambda_c|^{-\gamma} \tag{2.9}$$

and the systems becomes infinitely sensible to external perturbations. In the presence of an external field, exactly at the critical point $\lambda = \lambda_c$ the order parameter varies with h as a power law that we write as

$$\rho \sim h^{\beta/\sigma} \tag{2.10}$$

where σ is a new critical exponent. Up to now, nothing is deeply different from the time independent case.

What is really new with respect to equilibrium phase transitions is that in this model the asymptotic dynamical evolution itself is power law distributed in time. Suppose the initial configuration of the system is of a single active site. Then, one finds that the number of active sites scales asymptotically as

$$N(t) = \left\langle \sum_{i=1}^{N} \sigma_i(t) \right\rangle \sim t^{\theta} \tag{2.11}$$

and properties such as this one are perhaps the main peculiarity of non equilibrium phenomena. Among the others, we can introduce an exponent for the survival probability

$$P(t) \sim t^{-\delta} \tag{2.12}$$

and for the mean square spreading

$$R^2(t) = \frac{1}{N(t)} \left\langle \sum_{r} r^2 \sigma_r(t) \right\rangle \sim t^{\tilde{z}} \tag{2.13}$$

which represents the spread of activity from an original seed. These quantities show how criticality emerges as a dynamical feature in the time evolution from a given initial condition.

One more interesting feature of absorbing phase transitions is the so-called critical slowing-down. We are usually able to write an equation of motion for the order parameter $\rho(t)$, and we expect that at large times there exist different steady states ρ_{st}, which exchange stability for some value of the control parameter. These are, as we will

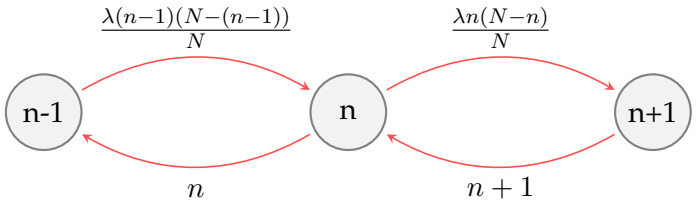


FIGURE 2.2: The transition scheme for the mean field contact process, where n is the number of active sites, $N-n$ the number of empty sites, and the spreading rate is rescaled $\lambda \to \lambda/N$. The spontaneous emptying rate is unitary.

see in a mean field approximation, the inactive absorbing phase and the active phase. Suppose now that the initial configuration is a system with all sites active, and not a single seed as for the previous power laws. If we look at the time it takes for the system to relax to the steady state, it turns out that at criticality the relaxation time diverges, leading to the power law behavior

$$\rho(t) \sim t^{-\alpha} \tag{2.14}$$

which implies that it takes an infinite time to reach ρ_{st} and defines a new critical exponent.

Finally, it is important to note that once again the phase transition exists only in the thermodynamic limit, i.e. for $N \to \infty$. A finite system will always end up in the absorbing state after a sufficiently long time.

2.1.1 MEAN FIELD APPROACH

In principle, to solve the dynamics of the contact process is to solve the equation of motion for the local order parameter $\rho(x,t)$, defined as

$$\rho(x,t) = \mathbb{P}[\sigma_x(t) = 1].$$

From the definition of the rates (2.1) we write the master equation

$$\mathbb{P}[\sigma_x(t+\Delta t) = 1] = \mathbb{P}[\sigma_x(t) = 1][1 - \mu\Delta t] + \frac{\lambda\Delta t}{q} \sum_{y \in \langle x \rangle} \mathbb{P}[\sigma_x(t) = 0, \sigma_y(t) = 1]$$

and the continuous time limit yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(x,t) = -\rho(x,t) + \frac{\lambda}{q} \sum_{y \in \langle x \rangle} \mathbb{P}[\sigma_x(t) = 0, \sigma_y(t) = 1]$$

which is still an exact relation. However, the last term in the r.h.s. makes it impossible to solve it exactly.

In order to better understand the nature of the phase diagram of the contact process, we shall consider a mean field approach. The fully-connected contact process obeys

the continuous time equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(x,t) = -\rho(x,t) + \frac{\lambda}{Nq}\sum_{y=1}^{N}\mathbb{P}[\sigma_x(t)=0]\mathbb{P}[\sigma_y(t)=1]$$

where we rescaled the control parameter to restore extensivity. Hence in the assumption of spatial homogeneity $\rho(x) = \rho$ we find

$$\begin{aligned}\frac{\mathrm{d}\rho}{\mathrm{d}t} &= -\rho + \frac{\lambda}{Nq}\left[Nq(1-\rho)\rho\right] \\ &= \rho(\lambda-1) - \lambda\rho^2.\end{aligned} \tag{2.15}$$

In general, we could obtain the same result starting directly from the master equation for the mean field contact process. The rates of the process are outlined in Figure 2.2, where n is the number of active sites and $N-n$ the number of empty sites. Hence the master equation is

$$\dot{p}_n = -p_n\left[n + \frac{\lambda n}{N}(N-n)\right] + p_{n+1}(n+1) + p_{n-1}\frac{\lambda(n-1)}{N}\left[N-(n-1)\right] \tag{2.16}$$

where $p_n(t)$ is the probability of having n active sites at time t.

We are interested in the average value of active sites $\langle n\rangle$, whose time evolution is obtained from (2.16) as

$$\langle\dot{n}(t)\rangle = \sum_n n\,\dot{p}_n(t)$$

which yields

$$\begin{aligned}\langle\dot{n}(t)\rangle &= \left\langle -n^2 + \frac{\lambda n^2(N-n)}{N} + n(n-1) + \frac{n(n+1)(N-n)}{N}\right\rangle \\ &= \left\langle n(\lambda-1) - \frac{\lambda n^2}{N}\right\rangle.\end{aligned}$$

We define the density of active sites as

$$\rho(t) = \lim_{N\to\infty}\frac{\langle n(t)\rangle}{N} \tag{2.17}$$

and one should note that in this limit the fluctuations are assumed to be arising from the central limit theorem, allowing us to write

$$\left\langle\left(\frac{n(t)}{N} - \rho(t)\right)^2\right\rangle = \frac{\langle n^2(t)\rangle}{N^2} - \rho^2(t) \sim \frac{1}{N} \to 0.$$

Hence in the large N limit $\langle n\rangle^2$ coincides with $\langle n^2\rangle$ and $\rho^2 = \langle n^2\rangle/N^2$. Thus the dynamical equation for $\langle n\rangle$ from (2.16) yields

$$\dot{\rho}(t) = \rho(t)(\lambda-1) - \lambda\rho^2(t)$$

which is exactly (2.15). However, in this framework one could in principle look at next-to-leading order solutions performing a system-size expansion of (2.16) [18].

Equation (2.15) has two possible stationary solutions $\dot{\rho} = 0$ given by

$$\rho_{\text{st}}^{v} = 0, \qquad \rho_{\text{st}}^{a} = \frac{\lambda - 1}{\lambda}$$

namely the vacuum solution and the active solution. Notice that if we ask for ρ_{st} to be positive, then ρ_{st}^{a} exists only if $\lambda > 1$. If we let $\rho(t) = \rho_{\text{st}} + \epsilon(t)$ for $\epsilon \ll 1$ we find

$$\frac{\mathrm{d}\epsilon}{\mathrm{d}t} = (\lambda - 1)\epsilon - 2\lambda\rho_{\text{st}}\epsilon + \rho_{\text{st}}(\lambda - 1) - \lambda\rho_{\text{st}}^2 + O(\epsilon^2)$$

so that $\dot{\epsilon} = (\lambda - 1)\epsilon$ in the vacuum case and $\dot{\epsilon} = -(\lambda - 1)\epsilon$ in the active one. Thus this linear stability analysis shows that there is a bifurcation point at a value

$$\lambda_c = 1 \tag{2.18}$$

which is the mean field critical point for the absorbing phase transition of the contact process. If $\lambda > \lambda_c$ the stable solution is the active one whereas for $\lambda < \lambda_c$ the vacuum is stable, plus there is no discontinuity in the order parameter at the transition, similarly to the case of second order phase transitions in equilibrium Statistical Mechanics.

Let $\Delta = \lambda - \lambda_c$. Suppose that the system is near the critical point so $\Delta \ll 1$, and that it is in its active phase $\Delta > 0$. The stationary condition reads

$$\Delta - \rho_{\text{st}}^{a}(\Delta + \lambda_c) = 0$$

so we find a power law

$$\rho_{\text{st}}^{a} = \frac{\Delta}{\Delta + \lambda_c} = \frac{\Delta}{\lambda_c}\left(1 - \frac{\Delta}{\lambda_c} + O(\Delta^2)\right) \sim \Delta + O(\Delta^2).$$

Comparing this result with (2.7) implies that $\beta_{\text{MF}} = 1$.

In order to find the susceptibility we need to add a spontaneous activation rate h, as said before. The mean field equation of motion thus become

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \rho(\lambda - 1) - \lambda\rho^2 + h(1 - \rho)$$

with stationary solution

$$\rho_{\text{st}} = \frac{\lambda - h - 1 \pm \sqrt{(\lambda - h - 1)^2 + 4\lambda h}}{2\lambda}.$$

Only the solution with the positive sign is the physical one, since the other gives a negative density. Then if the system is critical $\lambda = 1$ and $h \to 0$ we find

$$\rho_{\text{st}}(\lambda = 1, h) = \frac{\sqrt{h^2 + 4h} - h}{2} \approx \frac{\sqrt{4h}(1 + O(h) - h)}{2} \sim \sqrt{h}$$

so that 2.10 and the mean field value of the β exponent yield $\sigma_{\rm MF} = 2$. The susceptibility is instead given by

$$\chi = \left.\frac{\partial \rho_{\rm st}}{\partial h}\right|_{h=0} = \frac{1}{\lambda(\lambda-1)} \sim \Delta^{-1}$$

therefore $\gamma_{\rm MF} = 1$ from (2.9).

What about critical slowing-down? We can explicitly solve the differential equation $\dot\rho = \Delta\rho - \lambda\rho^2$ for $\rho(0) = \rho_0$, finding the solution

$$\rho(t) = \frac{\rho_0 \Delta e^{\Delta t}}{\Delta + \lambda\rho_0(e^{\Delta t} - 1)}$$

which holds for $\Delta \neq 0$. If we expand it at leading order for large times we find

$$\rho(t) \simeq \begin{cases} \Delta \left[\frac{\Delta}{\rho_0} - \lambda\right]^{-1} e^{\Delta t} & \Delta < 0 \\ \frac{\Delta}{\lambda} + \frac{\Delta}{\lambda^2}\left[\lambda - \frac{\Delta}{\rho_0}\right] e^{-\Delta t} & \Delta > 0 \end{cases}$$

which implies

$$\rho(t) - \rho_{\rm st} \sim e^{-|\Delta| t}. \tag{2.19}$$

Therefore away from criticality, the system approaches the stationary solution exponentially with a relaxation time $\tau = |\Delta|^{-1}$. Clearly, as we approach the critical point the relaxation time τ diverges leading to critical-slowing down. In fact, the differential equation $\dot\rho = -\rho^2$ yields

$$\rho(t) = \frac{\rho_0}{1 + \rho_0 t} \qquad (\Delta = 0).$$

In a fully occupied system at $t = 0$ we have $\rho_0 = 1$, so the density decays as

$$\rho(t) \sim \frac{1}{t}$$

which is the expected algebraic decay. Thus $\alpha_{\rm MF} = 1$. If instead we consider a finite system with N sites and a single occupied site as an initial condition, we have $\rho_0 = 1/N$ and

$$N(t) = \frac{1}{1 + t/N}$$

with $\rho(t) = N(t)/N$. If we take the thermodynamic limit $N \to \infty$ before the large time limit we end up with $N(t) = 1$, which implies that $\theta_{\rm MF} = 0$.

2.1.2 A COARSE-GRAINED APPROACH

In order evaluate correlations we need to refine Equation (2.15) in a coarse-grained fashion. A first improvement is in fact to let go of spatial homogeneity while still considering independent sites, hence

$$\begin{aligned}\frac{\partial}{\partial t}\rho(\mathbf{x},t) &= -\rho(\mathbf{x},t) + \frac{\lambda}{q}\sum_{\mathbf{y}}\mathbb{P}[\sigma_{\mathbf{x}}(t)=0]\mathbb{P}[\sigma_{\mathbf{y}}(t)=1] \\ &= -\rho(\mathbf{x},t) + \frac{\lambda}{q}\sum_{\mathbf{y}}\rho(\mathbf{y},t)[1-\rho(\mathbf{x},t)]\end{aligned}$$

where $\mathbf{x}$ is a d-dimensional vector. Having in mind a continuum space limit, we expand $\rho(\mathbf{y},t)$ about the point $\mathbf{x}$

$$\rho(\mathbf{y},t) \approx \rho(\mathbf{x},t) + (y_i - x_i)\partial_i\rho(\mathbf{x},t) + \frac{1}{2}(y_i-x_i)(y_j-x_j)\partial_i\partial_j\rho(\mathbf{x},t) + \dots$$

where we use the Einstein convention. Notice that the second term of the r.h.s. vanishes upon integrating over $\mathbf{y}$, being an odd function, and similarly the only non vanishing part of the third term is for $i=j$. We also consider a finite volume V for the time being, so we approximate the sum with an integral $\sum_{\mathbf{y}} \approx q/V\int_V d^d y$. Taking these assumptions into account, inserting this expansion into the equation of motion yields

$$\begin{aligned}\frac{\partial}{\partial t}\rho(\mathbf{x},t) = -\rho(\mathbf{x},t) + \frac{\lambda}{V}\int d^d y\Big[&\rho(\mathbf{x},t) + \frac{1}{2}|\mathbf{y}-\mathbf{x}|^2\nabla^2\rho(\mathbf{x},t) \\ &- \rho(\mathbf{x},t)^2 - \frac{1}{2}|\mathbf{y}-\mathbf{x}|^2\rho(\mathbf{x},t)\nabla^2\rho(\mathbf{x},t) + \dots\Big].\end{aligned} \tag{2.20}$$

In principle, $\dots$ contains all possible couplings between the order parameter and its derivatives. The simplest choice is to stop at the lowest order and consider a diffusive coupling only, that is

$$\frac{\partial}{\partial t}\rho(\mathbf{x},t) = \Delta\rho(\mathbf{x},t) - \lambda\rho^2(\mathbf{x},t) + D\nabla^2\rho(\mathbf{x},t) \tag{2.21}$$

where D is the constant factor appearing in (2.20).

We can understand Equation (2.21) very much like in a static field theory. In fact, finding the stationary solution for the coarse-grained order parameter amounts to extremizing the functional

$$\mathcal{F}[\rho] = \int d^d z\left[-\frac{\Delta}{2}\rho^2(\mathbf{z},t) + \frac{\lambda}{3}\rho^3(\mathbf{z},t) + D\partial_i\partial_i\rho(\mathbf{z},t)\right]$$

which resembles indeed the effective action in the static Landau theory. For instance, adding a rate for spontaneous activation amounts to set

$$\frac{\delta\mathcal{F}[\rho]}{\delta\rho(\mathbf{x},t)} = h.$$

The simplest dynamics in this framework is a simple relaxation towards the extreme [24], that is

$$\frac{\partial \rho(\mathbf{x},t)}{\partial t} = -\frac{\delta \mathcal{F}[\rho]}{\delta \rho(\mathbf{x},t)}$$

so we find (2.21). In principle, we can also include fluctuations by means of

$$\frac{\partial \rho(\mathbf{x},t)}{\partial t} = -\frac{\delta \mathcal{F}[\rho]}{\delta \rho(\mathbf{x},t)} + \tilde{\eta}(\mathbf{x},t)$$

where $\tilde{\eta}(\mathbf{x},t)$ is a random function defined by its probability distribution, which we would like to choose to be simple white noise. However, one should be careful that the noise must vanish in the vacuum for it to be an absorbing state, and that $\tilde{\eta}(\mathbf{x},t)$ should be interpreted as the sum of the individual noise contributions $\tilde{\eta} = \sum_{i=1}^{N} \eta_i$. In the large N limit, we expect that the fluctuations grow as $\sqrt{\rho}$ by means of the central limit theorem. Thus we should write

$$\frac{\partial \rho(\mathbf{x},t)}{\partial t} = -\frac{\delta \mathcal{F}[\rho]}{\delta \rho(\mathbf{x},t)} + \sqrt{\rho(\mathbf{x},t)}\eta(\mathbf{x},t) \tag{2.22}$$

where η is specified by

$$\langle \eta(\mathbf{x},t) \rangle = 0$$

$$\left\langle \eta(\mathbf{x},t)\eta(\mathbf{x}',t') \right\rangle = \Gamma \delta(t-t')\delta^{(d)}(\mathbf{x}-\mathbf{x}')$$

so it is independent on ρ.

We can understand the origin of the noise already from the master equation (2.16), by means of the Kramers-Moyal expansion [18]

$$\frac{\partial P(n)}{\partial t} = \sum_{i=1}^{\infty} \frac{(-1)^i}{i!} \frac{\partial^i}{\partial n^i} [\alpha_i(n) P(n)] \tag{2.23}$$

where

$$\alpha_i(n) = \int dn' \, (n'-n)^i w(n'|n)$$

Since the contact process is a jump process, the coefficients $\alpha_i(n)$ depend on the transition rates $w[n \pm 1|n]$ only, that is

$$\alpha_i(n) = w[n+1|n] + (-1)^i w[n-1|n].$$

We are particularly interested in the first two terms, since they give rise to a Fokker-Planck equation. For the mean field contact process, they are

$$\alpha_1(n) = \lambda \frac{n(N-n)}{N} - n$$

$$\alpha_2(n) = \lambda \frac{n(N-n)}{N} + n$$

hence

$$\frac{\partial P(n)}{\partial t} = -\frac{\partial}{\partial n}\left[\alpha_1(n)P(n)\right] + \frac{1}{2}\frac{\partial^2}{\partial n^2}\left[\alpha_2(n)P(n)\right].$$

A simple change of variable to $\rho = n/N$ gives

$$\frac{\partial P(\rho)}{\partial t} = -\frac{\partial}{\partial \rho}\left[\left(\rho(\lambda-1) - \lambda\rho^2\right)P(\rho)\right] + \frac{1}{2N}\frac{\partial^2}{\partial \rho^2}\left[\rho\left(\lambda(1-\rho)+1\right)P(\rho)\right]. \tag{2.24}$$

The corresponding Langevin equation is

$$\dot{\rho} = \rho(\lambda-1) - \lambda\rho^2 + \sqrt{\frac{1}{N}\rho\left[\lambda(1-\rho)+1\right]}\eta \tag{2.25}$$

where η represents the zero-mean Gaussian noise we were looking for.

Notice that, much like the static Ginzburg-Landau theory, this framework still gives the mean field exponents, but it accounts for spatial variation so we may compute correlations. In order to do so, let us consider (2.21) first. Since for $\Delta < 0$ the stationary value is $\rho_{\text{st}} = 0$, we expect $\rho(\mathbf{x}, t)$ to be small at large times and it is tempting to neglect the non-linear term. However, this is unjustified in the case $\Delta > 0$ unless we rather consider deviations from the homogeneous steady state,

$$\psi(\mathbf{x}, t) = \begin{cases} \rho(\mathbf{x}, t) & \Delta < 0 \\ \rho(\mathbf{x}, t) - \frac{\Delta}{\lambda} & \Delta > 0. \end{cases}$$

For the field $\psi(\mathbf{x}, t)$ the equation of motion reads

$$\frac{\partial \psi}{\partial t} = -|\Delta|\psi - \lambda\psi^2 + D\nabla^2\psi \approx -|\Delta|\psi + D\nabla^2\psi$$

assuming that at large times the deviations from the steady state are small and the quadratic part is therefore negligible. For the sake of simplicity, let us solve this equation in one dimension only. A simple Fourier transform

$$\psi(x, t) = \frac{1}{2\pi}\int dk e^{ikx} f(k, t)$$

gives

$$f(k, t) = A(k)\exp\left[-(|\Delta| + Dk^2)t\right].$$

The solution is exactly the Green's function $G_0(x, t)$ if we impose $\psi(x, 0) = \delta(x)$, which implies $A(k) = 1$. Hence

$$G_0(x, t) = \frac{1}{\sqrt{4\pi D t}} e^{-|\Delta|t - \frac{x^2}{4Dt}}. \tag{2.26}$$

Before going any further, we shall include back the noise. Suppose we are very close to the active steady state, so that $\langle\tilde{\eta}(x, t)\tilde{\eta}(x', t')\rangle \approx \Gamma\rho_{\text{st}}^a\delta(t - t')\delta(x - x')$. In this

approximation the equal time correlation function $c_\perp(x_1,x_2)=\langle\psi(x_1,t)\psi(x_2,t)\rangle$ can be written as

$$c_\perp(x_1,x_2)=\int_{-\infty}^{t}dt_1dt_2\int_{-\infty}^{+\infty}dy_1dy_2\,G_0(x_1-y_1,t-t_1)G_0(x_2-y_2,t-t_2)\cdot\langle\tilde\eta(y_1,t_1)\tilde\eta(y_2,t_2)\rangle$$

so that

$$\begin{aligned}c_\perp(x_1,x_2)&=\frac{\Gamma\rho_{\text{st}}^a}{4\pi D}\int_0^\infty\frac{d\tau}{\tau}e^{-2\Delta\tau}\int dy\exp\left[-\frac{(x_1-y)^2+(x_2-y)^2}{4D\tau}\right]\\&=\frac{\Gamma\rho_{\text{st}}^a}{\sqrt{8\pi D}}\int_0^\infty\frac{d\tau}{\sqrt{\tau}}e^{-2\Delta\tau}\exp\left[-\frac{(x_1-x_2)^2}{8D\tau}\right]\\&=\frac{\Gamma\rho_{\text{st}}^a}{4\sqrt{D\Delta}}\exp\left[-\frac{|x_1-x_2|}{\sqrt{D/\Delta}}\right].\end{aligned}$$

This implies that

$$\xi_\perp\sim\Delta^{-1/2}$$

so we find the exponent $\nu_\perp^{\text{MF}}=1/2$.

Note that the Green's function $G_0(x-x',t-t')$ describes how a perturbation at time t' and position x' propagates trough the system to a point x at time t. Thus from (2.26) we see immediately that the lifetime of such fluctuations is $\propto\Delta^{-1}$, which implies the exponent $\nu_\parallel^{\text{MF}}=1$. The previous result can be understood in this heuristic way as well, since the propagation at time t spreads within a distance $\approx\sqrt{Dt}\sim\Delta^{-1/2}$.

A more systematic way to account for correlation functions is by means of a path integral approach to the dynamics described by Equation (2.22) [42]. In full generality, the partition function for the stochastic process can be written as

$$Z=\int\mathcal{D}\eta\mathcal{D}\rho\,P(\eta)\,\delta\left[\frac{\partial\rho(\mathbf{x},t)}{\partial t}-\frac{\delta\mathcal{F}[\rho]}{\delta\rho(\mathbf{x},t)}-\eta(\mathbf{x},t)\right]$$

where η is a Gaussian random variable defined by

$$P(\eta)\sim e^{-\int d^dxdt\,\frac{\eta^2}{2\Gamma\rho}},\quad\left\langle\eta(\mathbf{x},t)\eta(\mathbf{x}',t')\right\rangle=\Gamma\rho(\mathbf{x},t)\delta(t-t')\delta^{(d)}(\mathbf{x}-\mathbf{x}').$$

Let us call $\frac{\delta\mathcal{F}}{\delta\rho}:=F[\rho]$. Inserting the identity

$$\delta(y)=\frac{1}{2\pi}\int d\tilde\rho\,e^{i\tilde\rho y}$$

at each spacetime point $(\mathbf{x},t)$ and performing a Wick rotation yields

$$Z\sim\int\mathcal{D}\eta\mathcal{D}\rho\mathcal{D}\tilde\rho\,e^{-\int d^dxdt\,\tilde\rho[\partial_t\rho-F[\rho]-\eta]}\,e^{-\int d^dxdt\,\frac{\eta^2}{2\Gamma\rho}}.$$

By performing the functional Gaussian integration over η one finds the partition function

$$Z \sim \int \mathcal{D}\rho\mathcal{D}\tilde{\rho}\, e^{-\mathcal{J}[\rho,\tilde{\rho}]} \tag{2.27}$$

where the action $\mathcal{J}[\rho,\tilde{\rho}]$ is the so-called Janssen-De Dominicis functional,

$$\mathcal{J}[\rho,\tilde{\rho}] = \int d^d x dt \, \left[\partial_t \rho - F[\rho] - \frac{\Gamma\rho\tilde{\rho}}{2}\right]\tilde{\rho}. \tag{2.28}$$

If we write explicitly the action for the contact process we readily identify a Gaussian part and two interaction terms, namely

$$\begin{aligned}\mathcal{J}[\rho,\tilde{\rho}] &= \int d^d x dt \, \left[\tilde{\rho}(\partial_t - (\lambda-1) - D\nabla^2)\rho\right] + \left[\lambda\tilde{\rho}\rho^2 - \frac{\Gamma\rho\tilde{\rho}^2}{2}\right] \\ &= \mathcal{J}_0[\rho,\tilde{\rho}] + \mathcal{J}_{\text{int}}[\rho,\tilde{\rho}]\end{aligned}$$

and we shall consider the free part only.

Let us introduce the doublet $\Phi = (\tilde{\rho},\rho)^T$ and $H = (\tilde{h},h)^T$, where H represents an external field coupled with Φ. The free partition function in the presence of such external field reads

$$Z[H] \sim \int \mathcal{D}\rho\mathcal{D}\tilde{\rho}\, e^{-\mathcal{J}_0[\Phi]}\, e^{\int d^d x dt\, H^T\Phi}.$$

It is convenient to write the free action in Fourier space (ω, q),

$$\mathcal{J}_0[\Phi] = \frac{1}{2}\Phi^T \begin{pmatrix} 0 & i\omega - (\lambda-1) + Dq^2 \\ -i\omega - (\lambda-1) + Dq^2 & 0 \end{pmatrix}\Phi$$

hence the partition function is simply another Gaussian integration. We shall note that the two-point correlation function for the doublet Φ reads

$$C^{(2)}(\mathbf{x}-\mathbf{x}', t-t') = \begin{pmatrix} \frac{\delta^2 Z[H]}{\delta h(\mathbf{x},t)\delta h(\mathbf{x}',t')} & \frac{\delta^2 Z[H]}{\delta h(\mathbf{x},t)\delta \tilde{h}(\mathbf{x}',t')} \\ \frac{\delta^2 Z[H]}{\delta \tilde{h}(\mathbf{x},t)\delta h(\mathbf{x}',t')} & \frac{\delta^2 Z[H]}{\delta \tilde{h}(\mathbf{x},t)\delta \tilde{h}(\mathbf{x}',t')} \end{pmatrix}$$

$$:= \begin{pmatrix} G^{(2,0)}((\mathbf{x}-\mathbf{x}', t-t')) & G^{(1,1)}(\mathbf{x}-\mathbf{x}', t-t') \\ G^{(1,1)}(\mathbf{x}'-\mathbf{x}, t'-t) & G^{(0,2)}(\mathbf{x}-\mathbf{x}', t-t') \end{pmatrix}.$$

From the form of the free action, it is clear that the only non vanishing part of the correlation function is $G^{(1,1)}$, which is

$$G^{(1,1)}(\mathbf{q},\omega) = \frac{1}{i\omega - (\lambda-1) + Dq^2} \tag{2.29}$$

in Fourier space. In real space,

$$G^{(1,1)}(\mathbf{r},\Delta t) \sim \int d^d q \int d\omega \frac{1}{i\omega - (\lambda-1) + Dq^2}\, e^{i\omega\Delta t} e^{i\mathbf{q}\cdot\mathbf{r}}$$

where for the sake of brevity we write $r = \mathbf{x} - \mathbf{x}'$, $\Delta t = t - t'$. Upon the change of variables $\omega = \omega'(\lambda - 1)$ and $q = q'\sqrt{(\lambda-1)/D}$ [26] we find

$$G^{(1,1)}(\mathbf{r},\Delta t) \sim F\left(\frac{\Delta t}{\Delta^{-1}}, \frac{\mathbf{r}}{\Delta^{-1/2}}\right) \tag{2.30}$$

where recall that $\Delta = \lambda - \lambda_c$. Thus the exponents for the correlation length ad the correlation time around the Gaussian field theory of the contact process are indeed $\nu_\perp^{\text{MF}} = 1/2$, $\nu_\parallel^{\text{MF}} = 1$.

So far, we have found

$$\begin{gathered} \beta_{\text{MF}} = 1, \quad \gamma_{\text{MF}} = 1, \quad \sigma_{\text{MF}} = 2, \\ \theta_{\text{MF}} = 0, \quad \alpha_{\text{MF}} = 1, \quad \nu_\perp^{\text{MF}} = 1/2, \quad \nu_\parallel^{\text{MF}} = 1 \end{gathered} \tag{2.31}$$

which are already a number of exponents. Of course, we do not expect them to be the correct values below the upper critical dimension, which in the case of the contact process turns out to be $d_u = 4$. However, we know that in the equilibrium case these critical exponents are not all independent, but they are connected by scaling relations that emerge from the underlying scale invariance. The case of absorbing phase transition is not different, as we will see in the next section.

2.2 DYNAMICAL SCALING IN AN ABSORBING PHASE TRANSITION

The scaling hypothesis assumes that it is possible to characterize the critical behavior by means of $\xi_\perp$ and $\xi_\parallel$ since they diverge near the critical point, thus becoming the only relevant scales in the system. Moreover, scale invariance implies that most of the quantities of interest can be expressed as generalized homogeneous functions, as we have seen. For instance, if we scale the control parameter via a dilation

$$\Delta \to \mu\Delta \tag{2.32}$$

we assume that the stationary density of active sites changes as

$$\rho(\mu\Delta) = \mu^\beta \rho(\Delta).$$

In fact, it is sufficient to choose $\mu = \Delta^{-1}$ to find

$$\rho(\Delta) = \Delta^\beta \tilde{\rho}(1)$$

which is exactly the power law we derived.

Notice that under a transformation of the form (2.32) the correlation lengths scale as

$$\xi_\perp \to \mu^{-\nu_\perp}\xi_\perp, \quad \xi_\parallel \to \mu^{-\nu_\parallel}\xi_\parallel$$

and since they define the only relevant scales in a critical system we expect distances and time intervals to scale like

$$r \to \mu^{-\nu_\perp} r, \quad t \to \mu^{-\nu_\parallel} t$$

as well. The striking difference with the equilibrium case is that we can look at the scaling of the dynamical evolution from a certain initial condition, since the correlation length $\xi_\parallel$ tells us how time should scale. For instance we shall add the time dependence to the previous scaling relation of the order parameter to get

$$\rho(\mu^{-\nu_\parallel} t, \mu\Delta) = \mu^\beta \rho(t, \Delta)$$

and we choose $\mu = t^{1/\nu_\parallel}$, so that at the critical point

$$\rho(t, 0) \sim t^{-\beta/\nu_\parallel}$$

which compared with (2.14) gives the first scaling relation

$$\alpha = \frac{\beta}{\nu_\parallel}. \tag{2.33}$$

If we consider an external field as well, from (2.10) we know that it should transforms as $h \to \mu^\sigma h$ which yields

$$\rho(h\mu^\sigma, \mu\Delta) = \mu^\beta \rho(\Delta, h)$$

and similarly for the susceptibility

$$\chi(h\mu^\sigma, \mu\Delta) = \mu^{-\gamma}\chi(\Delta, h).$$

Since $\chi = \partial_h \rho$, these two scaling relations hold if

$$\gamma = \sigma - \beta. \tag{2.34}$$

For the mean square spreading we have from the scaling of distances

$$R^2(t, \Delta) = \mu^{2\nu_\perp} R^2(\mu^{-\nu_\parallel} t, \Delta\mu) \sim t^{\tilde{z}}$$

and once again for $\Delta = 0$ we choose $\mu = t^{1/\nu_\parallel}$ to find

$$\tilde{z} = \frac{2\nu_\perp}{\nu_\parallel}. \tag{2.35}$$

In a similar fashion, the survival probability obeys

$$P(t, \Delta) = \mu^{\delta\nu_\parallel} \tilde{P}(t, \Delta)$$

and it has been shown that as an order parameter in the contact process $P(\Delta)$ scales as $\rho(\Delta)$, that is they share the same exponent β in the $t \to \infty$ limit. In this case at criticality we find

$$\delta = \frac{\beta}{\nu_\parallel} \tag{2.36}$$

and it also implies that $\alpha = \delta$.

Insofar, none of these relations is really an hypercaling relation, that is none involves the dimension of the system. In order to find one, we shall consider that we can write the number of active sites as

$$N(t, \Delta) = \int d^d r P(t, \Delta) \rho_{\text{st}}(t, \Delta)$$

so the scaling gives

$$N(\mu^{-\nu_\parallel} t, \mu \Delta) = \mu^{-d\nu_\perp - 2\beta} N(t, \Delta)$$

and with $\Delta = 0$ and $\mu = t^{1/\nu_\parallel}$ we get

$$\theta = \frac{d\nu_\perp - 2\beta}{\nu_\parallel} \tag{2.37}$$

which is the hyperscaling relation.

Notice that we could have considered a slightly different approach to scaling. If the initial condition is a single active site at the origin, the symmetry of the problem demands that we should get a scaling function of the form $F(r^2/t^{\tilde{z}}, \dots)$, and the easiest time dependence we can build is given by $t/\xi_\parallel = t\Delta^{\nu_\parallel}$. Thus we could write the density of active states as

$$\rho(r, t) = A(t) F(r^2 t^{-\tilde{z}}, \Delta^{\nu_\parallel} t) \tag{2.38}$$

for some function $A(t)$ whose value is determined by the dynamical scaling at criticality. In fact for $\Delta = 0$

$$R^2(t) = \frac{1}{N(t)} A(t) \int d^d x x^2 f(x^2 t^{-\tilde{z}}) = t^{-\theta} A(t) t^{d\tilde{z}/2} t^{\tilde{z}} \int d^d y f(y^2) y^2 \sim t^{\tilde{z}}$$

which implies $A(t) = t^{\theta - d\tilde{z}/2}$. A number of previous relations can be obtained by means of (2.38): imposing $\rho(r, t) \sim e^{-r/\xi}$ at large t gives for instance (2.35), and the hypercaling relation is obtained in a similar fashion [31].

2.2.1 AVALANCHES IN THE CONTACT PROCESS

As with the branching process, we shall think of an avalanche in the contact process as the spread of activity due to an initial small perturbation that causes a cascade of activity in an otherwise absorbing state [37]. In the system described by the contact

process, a single active site at $t = 0$ may generate this kind of activity that spreads until the absorbing state is reached once more.

For a single seed initial configuration, it makes sense to introduce the exponents defined in (2.11) - (2.13) so at criticality

$$P(t) \sim t^{-\delta}, \quad N(t) \sim t^{\theta}.$$

However, now we wish to look at the distributions of the avalanches size and the avalanches duration, $P_{\text{av}}(s)$ and $P_{\text{av}}(t)$ respectively. We start from the latter. The survival probability $P(t)$ can be casted in terms of avalanches by means of

$$P(t) = \int_t^{\infty} dt' P_{\text{av}}(t')$$

hence the survival probability at time t is the integrated avalanche duration. Thus we find immediately a power law behavior

$$P_{\text{av}}(t) \sim t^{-(\delta+1)} := t^{-\tau_t}$$

that is

$$\tau_t = \delta + 1 \tag{2.39}$$

which yields $\tau_t^{\text{MF}} = 2$ using $\delta_{\text{MF}} = \alpha_{\text{MF}} = 1$.

On the other hand, an avalanche of a given size s can have different durations, say in the interval $[t_1, t_2]$ so that

$$P_{\text{av}}(s) = \int_{t_1}^{t_2} dt P(s|t) D(t)$$

where $P(s|t)$ is the probability of an avalanche having size s given it has died at time t, and $D(t)$ is the probability that an avalanche dies at time t. The latter is simply given by

$$D(t) = -\frac{\mathrm{d}P(t)}{\mathrm{d}t} \sim t^{-\delta-1}.$$

One should also consider that the total number of active sites in surviving trials scales as $N_s(t) \sim t^{\theta+\delta}$, so we expect the typical size of an avalanche that dies at time t to be

$$s \sim \int_0^t dt' N_s(t') \sim t^{1+\theta+\delta}$$

hence we assume $P(s|t)$ to have a maximum at $t \sim s_c := s^{1/(1+\theta+\delta)}$. We also assume that the dependence on s comes through the rescaled variable s/s_c, writing

$$P_{\text{av}}(s) = \int dt\, t^{-(1+\theta+\delta)} F\left(\frac{s}{s_c}\right) t^{-(1+\delta)}$$

and upon changing the variable to $u = s/s_c$ the integral becomes

$$P_{\text{av}}(s) = s^{1/(1+\theta+\delta)} \int du\, u^{-1/(1+\theta+\delta)} \frac{1}{s} \left(\frac{u}{s}\right)^{(\delta+1)/(1+\theta+\delta)} F(u)$$
$$\sim s^{-(1+\theta+2\delta)/(1+\theta+\delta)}.$$

Thus the avalanche distribution is power law distributed[2] with an exponent

$$\tau = \frac{1+\theta+2\delta}{1+\theta+\delta} \tag{2.40}$$

whose value in mean field theory is $\tau_{\text{MF}} = 3/2$, exactly the same of the branching process. Note that by means of (2.37) and (2.36) we can also write

$$\tau = \frac{d\nu_\perp + \nu_\parallel}{d\nu_\perp + \nu_\parallel + \beta}$$

which is specified in terms of the exponents of the correlation length, the correlation time and the order parameter. Similarly,

$$\tau_t = \frac{\nu_\parallel + \beta}{\nu_\parallel}$$

for the time distribution.

2.2.2 FINITE-SIZE SCALING

The dynamical scaling behavior may be affected by the finiteness of the system if we are not in the thermodynamic limit $N \to \infty$. In particular, the spatial correlation length does not diverge since it is limited by the size of the system L. However, in an absorbing phase transition another complication emerges: for finite N the only stationary state is the vacuum, since there is always a non vanishing probability of reaching the absorbing state. Thus, we should look at the quasistationary state described by the surviving trials after an initial transient before they die out.

Since $\xi \sim \Delta^{-\nu_\perp}$, we assume that near the critical point

$$L \to \mu^{-\nu_\perp} L$$

is the correct transformation of the system size. If we add L as a scaling field we find

$$\rho(\Delta, L) = \mu^{-\beta} \rho(\mu\Delta, \mu^{-\nu_\perp} L)$$

and if we choose $\mu = L^{1/\nu_\perp}$ we obtain the scaling form

$$\rho(\Delta, L) = L^{-\beta/\nu_\perp} f(\Delta L^{1/\nu_\perp}). \tag{2.41}$$

[2]Notice that the same result can be obtained considering directly from the power law behavior of $P_{\text{av}}(t)$ with a change of variable $t(s) = s^{-1/(1+\theta+\delta)}$ [28].

The scaling function $f(x)$ need to be such that $f(x) \sim x^\beta$ for large x, so that we recover the correct behavior in the limit $L \gg \xi_\perp$. Thus rescaling the order parameter $\rho \to \rho L^{\beta/\nu_\perp}$ gives a universal function of $\Delta L^{1/\nu_\perp}$, independently on the system size. Exactly at the critical point the L dependence is a power law

$$\rho(0, L) \sim L^{-\beta/\nu_\perp}. \tag{2.42}$$

Similarly, if we include the time dependence the scaling is

$$\rho(t, \Delta, L) = \mu^{-\beta} \rho(\mu^{-\nu_\parallel} t, \mu\Delta, \mu^{-\nu_\perp} L)$$

and the same choice as before for μ at $\Delta = 0$ leads to

$$\rho(t, 0, L) = L^{-\beta/\nu_\perp} g(L^{-\nu_\parallel/\nu_\perp} t). \tag{2.43}$$

This form shows that finite size effects become important at a time scale $t_{\text{FSS}} \sim L^{\nu_\parallel/\nu_\perp}$, so that for $t \ll t_{\text{FSS}}$ we expect that

$$g(x) \sim x^{-\alpha}$$

so that we recover $\rho(t) \sim t^{-\alpha}$.

One could also look at the stationary variance of the density of active states, namely

$$\chi_\rho = \lim_{L\to\infty} L^d \left[\left\langle \rho_{\text{st}}^2 \right\rangle - \left\langle \rho_{\text{st}} \right\rangle^2 \right] \tag{2.44}$$

where $\rho_{\text{st}} = L^{-d} \sum_i \sigma_i$. We further assume that at the critical point this quantity scales like $\chi \sim |\Delta|^{-\gamma'}$, that is fluctuations at criticality become very large. The scaling for the variance is

$$\chi_\rho(\mu\Delta, \mu^{-\nu_\perp} L) = \mu^{\gamma'} \chi_\rho(\Delta, L)$$

hence the scaling form

$$\chi_\rho(\Delta, L) = L^{\gamma'/\nu_\perp} u(\Delta L^{1/\nu_\perp})$$

which at the critical point becomes

$$\chi_\rho(0, L) \sim L^{\gamma'/\nu_\perp} \tag{2.45}$$

Notice however that scale invariance for (2.44) implies

$$(\mu^{-\nu_\perp} L)^d \left[\left\langle \mu^{2\beta} \rho^2 \right\rangle - \left\langle \mu^\beta \rho \right\rangle^2 \right] = \mu^{-\nu_\perp d + 2\beta} \chi_\rho \sim |\mu\Delta|^{-\gamma'}$$

which holds only if the hyperscaling relation

$$\gamma' = \nu_\perp d - 2\beta \tag{2.46}$$

is valid. This result gives $\gamma'_{\text{MF}} = 0$ and it is quite interesting since we know that $\gamma_{\text{MF}} = 1$, signaling that the fluctuation-dissipation theorem of equilibrium Statistical Mechanics breaks down in non equilibrium systems.

However, we can still say something about the relation between the stationary equal time correlation function $c_\perp(\mathbf{r})$ and the fluctuations of the order parameter [13]. We normalize the correlation function by its value in $r = 0$, which is nothing but the variance

$$c'_\perp(r) = \frac{c_\perp(r)}{c_\perp(0)} = \frac{c_\perp(r)}{\langle\sigma_i\rangle^2 - \langle\sigma_i\rangle^2}.$$

Due to translational invariance, the denominator is given by $\langle\sigma_i\rangle^2 - \langle\sigma_i\rangle^2 = \rho(1-\rho)$, so close to the critical point we find

$$c'_\perp(r) \approx \rho^{-1} c_\perp(r).$$

Recall that we expect the correlation function to decay algebraically at criticality, yielding an exponent η defined as

$$c'_\perp(r) \sim r^{-(d-2+\eta)}. \tag{2.47}$$

Furthermore, in a finite system of linear size L, the variance (2.44) can be written as

$$\begin{aligned}
\chi_\rho &= L^d \left[\left\langle \frac{1}{L^{2d}} \sum_{ij} \sigma_i \sigma_j \right\rangle - \left\langle \frac{1}{L^d} \sum_i \sigma_i \right\rangle \left\langle \frac{1}{L^d} \sum_j \sigma_j \right\rangle \right] \\
&= L^{-d} \sum_{ij} \left[\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle \langle\sigma_j\rangle \right] \\
&= L^{-d} \sum_{ij} c_\perp(|i-j|).
\end{aligned}$$

If we approximate the sum with an integral, we find

$$\chi_\rho = \int d^d r \, \rho \, r^{-(d-2+\eta)} \sim \rho \int_0^L dr r^{d-1-(d-2+\eta)}$$

and considering the scaling relation for the order parameter (2.42) we end up with

$$\chi_\rho \sim L^{2-\eta-\beta/\nu_\perp}.$$

From the scaling of the variance (2.45) and the hypersclaing relation (2.46) the exponent η is determined by

$$\eta = 2 - d + \frac{\beta}{\nu_\perp} = 2 - d\frac{\theta+\delta}{\theta+2\delta} \tag{2.48}$$

This implies, for instance, $\eta_{\text{MF}} = 0$. This is expected: in fact, it is already clear from (2.29).

This is the general mechanism for finite-size scaling: we consider L as a scaling field like any other. In such a way we are able to identify scaling function and tackle the fact that in finite systems no real phase transition happens.

Algorithm 1 Gillespie algorithm

1: $r_1, r_2 = \text{uniform}[0,1]$;
2: $a_0 \leftarrow \sum_{i=1}^{M} a_i$;
3: $\tau \leftarrow -a_0^{-1} \log r_2$;
4: **if** $a_0^{-1} \sum_{i=1}^{j'-1} \leq r_2 < a_0^{-1} \sum_{i=1}^{j'}$ **then** $j \leftarrow j'$;
5: $t \leftarrow t + \tau$;
6: $R \leftarrow R_j$.

2.3 SIMULATING THE CONTACT PROCESS

Let us describe one of the mos popular algorithm to simulate stochastic processes, the Gillespie algorithm [19]. In particular, is is useful to simulate fully connected models such as the mean field contact process. We will describe a suitable choice for the $2D$ lattice later on.

Suppose we have a set of M reactions $(R_1, \dots, R_M)$ between N species $(S_1, \dots, S_N)$. At a time t, we call $\mathbf{X}(t)$ the vector whose i-th entry represents the number of the i-th species and we introduce

$$a_j(\mathbf{x})dt = \mathbb{P}\left[R_j \in [t, t+dt] \,|\, \mathbf{X}(t) = \mathbf{x}\right]$$

that is the probability that the reaction R_j happens in the time interval $[t, t+dt]$ given that at time t the state of the system is $\mathbf{x}$. In particular, we are interested in

$$p(\tau, j \,|\, \mathbf{x}, t) = \mathbb{P}\left[R_j \in [t+\tau, t+\tau+d\tau] \,|\, \mathbf{X}(t) = \mathbf{x}\right],$$

the probability that the next reaction occurs in $[t+\tau, t+\tau+d\tau]$ and that it is the j-th reaction.

Let us compute the probability that no reaction happens in $[t, t+\tau]$, with $\tau = kd\tau$ as

$$P_0(\tau \,|\, \mathbf{x}, t) = \left[1 - d\tau \sum_{j=1}^{M} a_j(\mathbf{x})\right]^k \xrightarrow[k\to\infty]{} e^{-\tau a_0(\mathbf{x})}$$

where $a_0(\mathbf{x}) = \sum_j a_j(\mathbf{x})$. Hence

$$\begin{aligned} p(\tau, j \,|\, \mathbf{x}, t) = a_j(\mathbf{x}) d\tau P_0(\tau|\mathbf{x}, t) &= \left[\frac{a_j(\mathbf{x})}{a_0(\mathbf{x})}\right] \left[a_0(\mathbf{x}) e^{-\tau a_0(\mathbf{x})}\right] \\ &= p(j \,|\, \mathbf{x}, t) p(\tau \,|\, \mathbf{x}, t) \end{aligned}$$

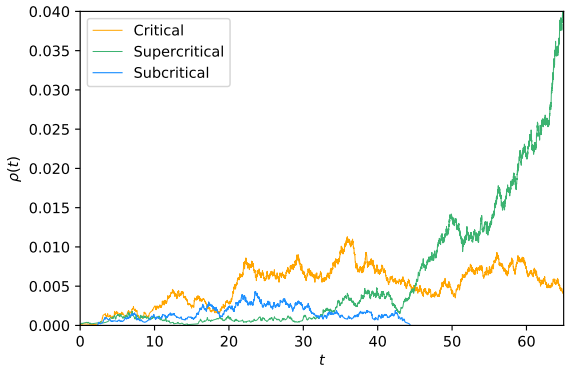


FIGURE 2.3: Typical examples of the time evolution of the density of active states in the mean field contact process, for subcritical ($\lambda = 0.9$), critical ($\lambda = 1$) and supercritical ($\lambda = 1.1$) values of the control parameter. The initial configuration is a single active site and $N = 10^4$.

so j is distributed with point probabilities a_j/a_0, whereas τ is exponentially distributed. We implement a simple tower sampling procedure to generate j and the well-known transformation method to generate τ[3].

The Gillespie algorithm is explicitly shown in Algorithm 1. Notice that the master equation and its rates, such as in Figure 2.2, fully specify the parameters a_j that are needed to perform the algorithm.

2.3.1 THE MEAN FIELD CONTACT PROCESS

In the case of the mean field contact process, the master equation gives the following reactions for the active nodes A and the inactive ones I,

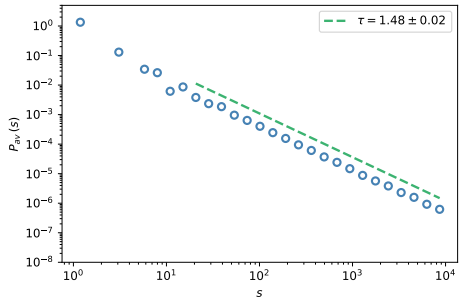
$$A + I \xrightarrow[\lambda n/N]{} A + A$$

$$A \xrightarrow[\mu=1]{} I$$

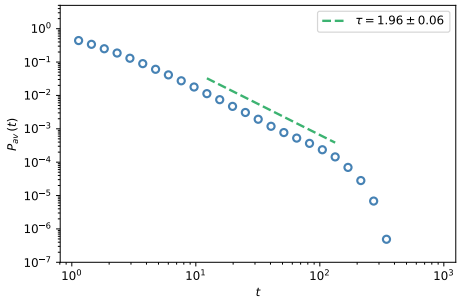
where n is the number of active site and N is the total number of sites. The results of a simulation performed with the Gillespie algorithm for $N = 10^4$ and $n(0) = 1$ are shown in Figure 2.3. As we can see, the density of active states $\rho(t)$ behaves quite differently in the subcritical, in the critical and in the supercritical regime. Activity is not sustainable if $\lambda < 1$, whereas it spread for $\lambda > 1$, limited only by the fact that the system is finite.

In the same setting, that is $n(0) = 1$, we can study the propagation of avalanches as well. We define the size of an avalanche s as the number of activations, i.e. the number of times the system undergoes a reaction $A + I \to A + A$ before going back

[3]Given a probability distribution $p_X(x)$, suppose we can find its cumulative distribution $F_X(x)$ and that we can generate a random number u uniformly distributed in $[0, 1]$. To sample $p_X(x)$ it is sufficient to look for the value x^* such that $u = F_X^{-1}(x^*)$: indeed, a trivial change of variables gives that x^* follows the distribution $p_X(x)$. For an exponential distribution $p_X(x) = ae^{-ax}$ we find immediately $x = a^{-1} \log u$.

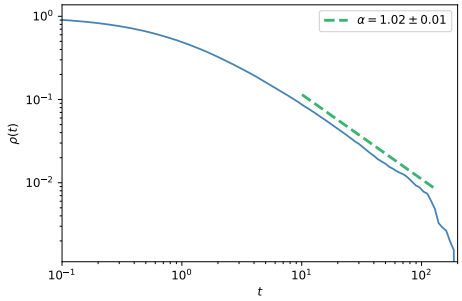


(A) Avalanche distribution in size.

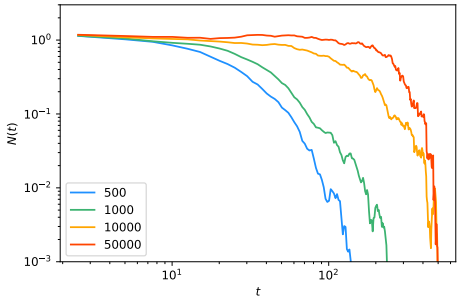


(B) Avalanche distribution in time.

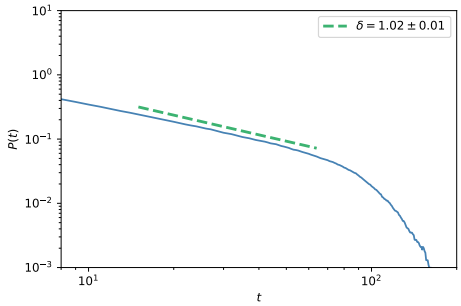
FIGURE 2.4: Avalanches in the critical mean field contact process for $N = 10^4$ nodes, starting from a single active site. (A) The size of an avalanche s in the contact process is defined as the number of activations before the system reaches the absorbing state, with $P_{\text{av}}(s) \sim s^{-\tau}$ and $\tau_{\text{MF}} = 3/2$. (B) The duration of the avalanches is power law distributed as well, with $P_{\text{av}}(t) \sim t^{-\tau_t}$ and $\tau_t^{\text{MF}} = 3/2$.



(A) Mean field α exponent.



(B) Mean field θ exponent.



(C) Mean field δ exponent.

FIGURE 2.5: The density of active site $\rho(t)$ and the survival probability $P(t)$ in the critical mean field contact process for $N = 10^4$ nodes. (A) If $n(0) = N$, after an initial transient the density decays as $\rho(t) \sim t^{-\alpha}$ before finite-size effects. (B) If $n(0) = 1$, the number of active sites should increase as $N(t) \sim t^{\theta}$. Since $\theta_{\text{MF}} = 0$, finite-size effects are particularly relevant and we increase the system size N from 500 to 50000 sites to check that the period of time in which $N(t)$ stays constant increases as well. (C) If $n(0) = 1$, the survival probability decays as $P(t) \sim t^{-\delta}$.

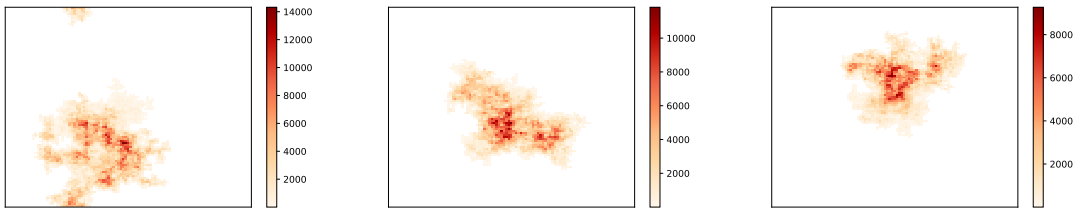


FIGURE 2.6: Three different examples of a critical $2D$ contact process simulated using Algorithm 2 for 10^5 timesteps and 101^2 sites, starting from a single active site at the center of the lattice. The colormap identifies for how many steps of the algorithm a given site has been active.

to the absorbing state. We expect from (2.40) that at criticality the distribution $P_{\text{av}}(s)$ follows a power law with critical exponent $\tau = 3/2$. We find

$$\tau_{\text{MF}} = 1.48 \pm 0.02$$

which is fully consistent with the expected result. Similarly, the Gillespie algorithm allows to study the distribution of the duration of an avalanche $P_{\text{av}}(t)$, since we correctly sample the time at which each reaction takes place. From (2.39) we expect a power law distribution with critical exponent $\tau_t^{\text{MF}} = 2$. In our simulation we find

$$\tau_t^{\text{MF}} = 1.96 \pm 0.06.$$

Both these results are reported in Figure 2.4.

The avalanche exponents depend on the value of δ and θ, that is the critical exponents of the survival probability $P(t) \sim t^{-\delta}$ and the number of active sites $N(t) \sim t^{\theta}$. These two power laws are valid if $n(0)N$. If we choose a fully occupied system instead, with $n(0) = N$, we would estimate the critical exponent α rather than θ, since the density of active sites decay as $\rho(t) \sim t^{-\alpha}$ in this case.

In Figure 2.5 the estimates for these exponents are reported. We expect $\delta_{\text{MF}} = \alpha_{\text{MF}} = 1$, and we find

$$\delta_{\text{MF}} = 1.02 \pm 0.01$$

$$\alpha_{\text{MF}} = 1.02 \pm 0.01.$$

Notice that there is a slight complication for θ, since we know that $\theta_{\text{MF}} = 0$. Hence $N(t) \sim 1$ in the mean field contact process. This implies that finite size effect are particularly relevant, as we can see from Figure 2.5b. To check this, we perform simulations for increasing system size and we verify that $N(t)$ stays in fact constant for longer periods of time.

2.3.2 THE $2D$ CONTACT PROCESS

The Gillespie algorithm as we have described is not feasible to simulate the contact process in two dimensions, since the rate depends on the lattice site[4]. Consider instead

[4]There exist variants of the Gillespie algorithm that can be adapted for topologies that are not fully connected, but they are much more complex and beyond the scope of this thesis. See, for example, [11]

Algorithm 2 $2D$ contact process

1: $i = \text{uniform}[0, N_a]$; $r = \text{uniform}[0, 1]$;
2: $p_\lambda \leftarrow \lambda/(1+\lambda)$;
3: **if** $r < p_\lambda$ **then**:
4: $\quad j = \text{uniform}[0, 4]$;
5: $\quad j$-th nn of $i \leftarrow 1$;
6: **else**
7: $\quad i$-th active site $\leftarrow 0$;
8: $t \leftarrow t + 1/N_a$.

that at each timestep a given active site either attempts to infect one among the four nearest neighbor with rate λ, or is inactivated with a unitary rate. Thus we implement the following strategy, summarized in Algorithm 2 [12]: we randomly select an active site i and one of its neighbors j, and with probability

$$p_\lambda = \frac{\lambda}{1+\lambda}$$

the site j becomes active if it was empty. With probability $1 - p_\lambda$ the site i becomes inactive instead. Since there are N_a active sites at time t to begin with, we increase the time of $\Delta t = 1/N_a$ so that a unit time interval corresponds on average to one attempt per site. In this way we avoid selecting empty sites to begin with, so the algorithm is more efficient. The simulation stops either after a given time or if it reaches the absorbing state, from which it cannot escape.

The critical value of the spreading rate in $2D$ is $\lambda_c = 1.6488$ from [31]. In Figure 2.6 three different examples of a critical contact process with periodic boundary conditions obtained with Algorithm 2 after 10^5 iterations are shown. Notice how the number of activations for each site are rather inhomogeneous: in the supercritical case each site would remain active much longer instead.

We can study the propagation of avalanches in the very same way as in the mean field case. The power law behaviors at criticality, both in size and in time, are reported in Figure 2.7a-2.7b. For the distribution in size we find

$$\tau = 1.27 \pm 0.01$$

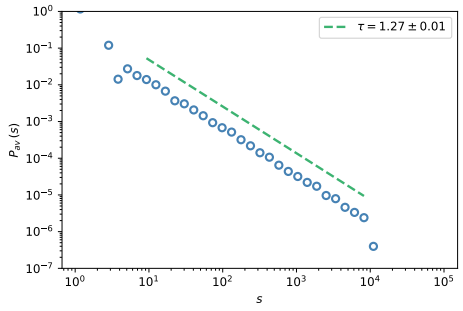
while for the distribution in time

$$\tau_t = 1.45 \pm 0.03.$$

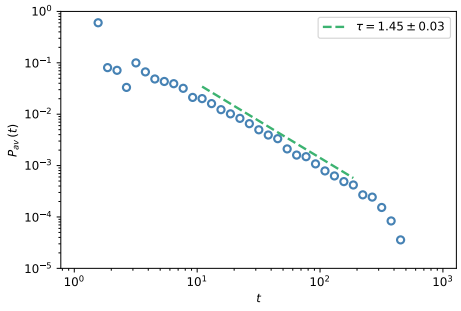
We expect $\tau \approx 1.268$ and $\tau_t \approx 1.450$ from previous works [37], hence both of our results are consistent with the known values. The exponents are lower with respect to the mean field case, which implies that the tails of the distributions are heavier and it is more probable to find larger and longer avalanches.

	Mean Field		2D	
	Simulation	Exact	Simulation	[31, 37]
τ	1.48	1.5	1.27	1.268
τ_t	1.96	2	1.45	1.4505
θ	0	0	0.228	0.2295
δ	1.02	1	0.448	0.4505

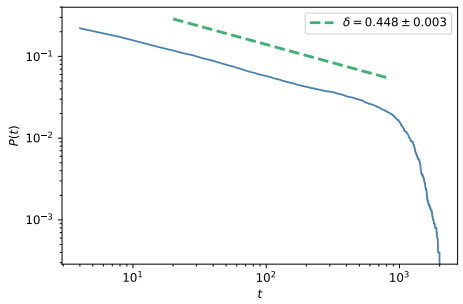
TABLE 2.1: Critical exponents of the contact process.



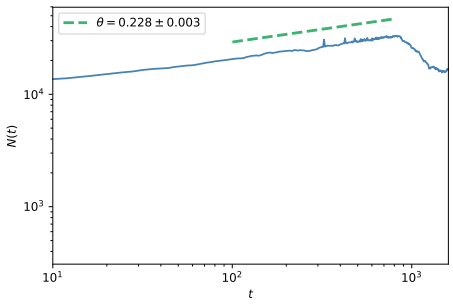
(A) Avalanche distribution in size.



(B) Avalanche distribution in time.



(C) Survival probability.



(D) Number of active sites.

FIGURE 2.7: Simulation of a critical $2D$ contact process on a 101^2 lattice, starting from a single active site. (A) Like in the mean field case, the distribution of avalanches size is power law distributed, with $P_{\text{av}}(s) \sim s^{-\tau}$ and $\tau \approx 1.27$. (B) The duration of the avalanches is power law distributed as well, with $P_{\text{av}}(t) \sim t^{-\tau_t}$ and $\tau_t \approx 1.45$. (C) The survival probability decays as $P(t) \sim t^{-\delta}$, with $\delta \approx 0.448$. (D) The number of active sites increases as $N(t) \sim t^{\theta}$, with $\theta \approx 0.228$

We study once more the survival probability and the total number of active sites, in order to evaluate the exponents δ and θ respectively. As reported in Figure 2.7c-2.7d, the results are

$$\delta = 0.448 \pm 0.003$$

and

$$\theta = 0.228 \pm 0.003.$$

In the literature these exponents in two dimensions are given by $\delta \approx 0.4505$ and $\theta \approx 0.2295$ [31], so both our results are fully consistent. With respect to the mean field case, the probability of a run to survive decays faster at short times, but has much heavier tails since the exponent in two dimensions is roughly a half of the mean field one. On the other hand, $\theta \neq 0$ implies that the average number of active sites increases at criticality, whereas it stays constant in the mean field approach. Notice also that with these values the relations (2.39) and (2.40) yield

$$\tau_t \approx 1.448$$

and

$$\tau \approx 1.267$$

once more. These various results are summarized in Table 2.1.

CHAPTER 3

A RENORMALIZATION GROUP APPROACH

In the first part of this chapter we aim to review the phenomenological coarse-graining procedure introduced by the authors in [34] to deal with the long range interactions that one reasonably expects in a network of neurons. They analyzed the data from an experiment that was able to monitor the single-cell activity of 1485 neurons in the hippocampus of a mouse moving in a virtual corridor for a long period of time. The activity is subsequently discretized so that cells are either "on", $\sigma_i = 1$, or "off", $\sigma_i = 0$, much like the case of the contact process. With enough resolution in time, this his allows to look directly at the dynamical evolution of the network of neurons and to coarse-grain it.

The results obtained by the authors seem to show signatures similar to the ones that one might expect if the system was driven to a non-trivial fixed point by the Renormalization Group (RG) flow. Hence one might argue that the data point to the possibility that neural activity is critical and essentially scale-free, giving credit to the critical brain hypothesis. Moreover, this draws a somewhat different conclusion with respect to the investigation of avalanches of spontaneous activity, since the brain of the mouse is not in a resting state. However, it is not clear if and how this procedure is really equivalent to a RG flow, and whether the behavior of a system at its critical point can be recovered.

We try to address these questions, hence in the second part of this chapter we apply the same procedure to the critical contact process. The idea is that we have full control over the control parameter, so we know for certain whether our system is or is not at the edge of a phase transition. In such a way, we try to seek validation for the relation between the scaling behaviors that arise from this coarse-graining procedure and the underlying criticality.

3.1 THE COARSE-GRAINING PROCEDURE

In the simplest models in Statistical Mechanics, interactions are usually short-range. Hence if we want to recover the macroscopic, collective behavior it makes sense to think of the coarse-graining procedure as an average over the short distance, microscopic details. For instance, this is the case of the block spin transformation described in the first chapter, where we sum neighboring spins to define the coarse-grained variables.

In general, if we start with a set of variables $\{\sigma_i\}$ a coarse-grained transformation is defined by

$$\sigma_i \to \tilde{\sigma}_{i'} = f\left(\sum_{j\in\mathcal{N}_i} \sigma_j\right) \tag{3.1}$$

where $\mathcal{N}_i$ denotes the neighborhood of the variable i and i' indexes the new coarse-grained variables. For instance, in the case of a block spin transformation, the function $f(\cdot)$ is just the identity. It is worth noting that under the coarse-graining transformation the joint probability distribution changes as well, namely

$$P(\{\sigma_i\}) \to \tilde{P}(\{\tilde{\sigma}_{i'}\}).$$

These probability distributions are distributions over a different number of degrees of freedom, since the coarse-graining naturally reduces the number of variables. Hence to compare them we should expand the system back to its original size. If we think of a system that follows a Boltzmann distribution

$$P(\{\sigma_i\}) = \frac{1}{Z} e^{-\beta H(\{\sigma_i\})}$$

we ask the RG transformation to leave the partition function Z invariant and let the Hamiltonian transform. After the coarse-graining step we find a new Hamiltonian $\tilde{H}(\{\tilde{\sigma}_i\})$ for the variables $\{\tilde{\sigma}_i\}$. Consequently, we need to properly rescale the scale of the system $i' = g(i)$ and renormalize the variables $\tilde{\sigma}_{i'} = h(\sigma_i)$ so that the new system has the same physical properties as before. Hence the coarse-graining step is properly only the first out of the three steps the RG transformation is made of, and arguably it is also the hardest to perform exactly[1].

It is important to stress out that the RG procedure does not necessarily rely on the Boltzmann distribution that is typically found in Statistical Mechanics, but it is actually more general. This is what we want to show now for the simple case of independent variables. Consider for instance a sequence of independent and identically distributed (iid) random variables $\{X_i\}$ and let $p(x)$ be the probability density of X. What is the distribution for the sum of these variables? We may think of this question as a coarse-graining step of the form

$$X_1^{(n)} = X_1^{(n-1)} + X_2^{(n-1)}, \quad X_2^{(n)} = X_3^{(n-1)} + X_4^{(n-1)}, \quad \dots$$

[1] To find the new, effective Hamiltonian that describes the coarse-grained variable exactly is generally hard and one typically has to rely on perturbation theory around a Gaussian Hamiltonian. Moreover, it is not always trivial to define a coarse-graining step that preserves exactly the features of the system, such as its symmetries.

which is the n-th step of a decimation process. We also renormalize each coarse-grained variable by a factor ζ that we are about to determine, so that after n coarse-graining step the i-th coarse-grained variable is given by

$$X_i^{(n)} = \frac{X_{2i-1}^{(n-1)} + X_{2i}^{(n-1)}}{\zeta}.$$

The pdf transforms as

$$\begin{aligned} p_{n+1}(x) &= \int dx_1 dx_2\, p_n(x_1) p_n(x_2) \delta\left(x - \frac{x_1 + x_2}{\zeta}\right) \\ &= \zeta \int dx'\, p_n(x') p_n(\zeta x - x') \\ &= (\mathcal{R}_\zeta p_n)(x) \end{aligned}$$

and $\mathcal{R}_\zeta$ is exactly the RG transformation we have described. This transformation is simpler in terms of the characteristic function

$$\varphi(z) = \int dx\, p(x) e^{ixz} = 1 + \sum_{k=1}^{\infty} \frac{i^k}{k!} \left\langle X^k \right\rangle z^k,$$

since a Fourier transform yields

$$\begin{aligned} (\mathcal{R}_\zeta p_n)(x) &= \frac{\zeta}{(2\pi)^2} \int dx' dz dz'\, \varphi_n(z) \varphi_n(z') e^{-izx'} e^{-iz'(\zeta x - x')} \\ &= \frac{1}{2\pi} \int dz \varphi_n^2\left(\frac{z}{\zeta}\right) e^{-izx}. \end{aligned}$$

Hence we find that $\mathcal{R}_\zeta$ acts in a simpler local way, albeit still non-linear, on the characteristic functions

$$(\mathcal{R}_\zeta \varphi_n)(z) = \varphi_n^2\left(\frac{z}{\zeta}\right).$$

A further simplification comes if we consider the cumulants of the distribution, defined as $\omega(z) = \log \varphi(z)$, so that

$$(\mathcal{R}_\zeta \omega_n)(z) = 2\omega_n\left(\frac{z}{\zeta}\right). \tag{3.2}$$

We are interested in the asymptotic distribution for the coarse-grained sequence of $\{X_i\}$, which we can think of as the distribution of its normalized sum. Hence we look for the fixed points ω^* of the RG transformation, which we can write in a simple form for the cumulants

$$\omega^*(z) = i\omega_{(1)} z - \frac{1}{2}\omega_{(2)} z^2 + \cdots = 2\omega^*\left(\frac{z}{\zeta}\right) = i\omega_{(1)} \frac{2z}{\zeta} - \omega_{(2)} \left(\frac{z}{\zeta}\right)^2 + \ldots$$

where $\omega_{(k)}$ is the k-th cumulant of the distribution. For this relation to hold at all orders we need

$$\zeta = 2, \quad \omega_{(k)} = 0 \qquad \forall k \geq 2$$

which implies that the fixed point is $\varphi^*(z) = e^{i\omega_1 z}$. Since $\omega_1 = \langle X \rangle$ the fixed point corresponds to the pdf

$$p^*(x) = \delta\left(x - \langle X \rangle\right) \tag{3.3}$$

which is nothing but the law of large numbers (LLN).

A linear stability analysis for $\omega(z) = \omega^*(z) + \delta\omega(z)$ for small $\delta\omega$ gives

$$(\mathcal{R}_2 \delta\omega)(z) = 2\delta\omega\left(\frac{z}{\zeta}\right) = \sum_{k=1}^{\infty} \frac{(iz)^k}{k!} 2^{1-k} \delta\omega_{(k)} = \sum_{k=1}^{\infty} 2^{1-k} u_{(k)}(z)$$

so that for each k the following relation holds

$$(\mathcal{R}_2 u_{(k)})(z) = 2^{1-k} u_{(k)}(z).$$

Thus the eigenvalues of the transformation are $\lambda_k = 2^{1-k}$ with eigenvectors $u_{(k)}(z)$. After m iterations the RG transformation yields

$$(\mathcal{R}_2^m u_{(k)})(z) = 2^{(1-k)m} u_{(k)}(z)$$

and we are interested in the $m \to \infty$ limit. Thus we find that the behavior of the perturbation along a given eigenvector depends on the eigenvalue: in fact, $k = 1$ gives a marginal perturbation which amounts to a shift in the fixed point by $\delta\omega(z) = iz\delta\omega_{(1)}$, so the form of the distribution doesn't change. This is not true for the remaining eigenvalues, namely for $k > 1$, which are characterized by the limit $2^{(1-k)m} \to 0$ for $m \to \infty$. What we find is that no relevant perturbations exist: the law of large number is a stable fixed point in the space of probability distributions.

What happens if the probability distribution is such that $\langle X \rangle = \omega_{(1)} = 0$? In such case upon expanding Equation (3.2) the first non-trivial order is the second one, which implies

$$\zeta = \sqrt{2}, \quad \omega_{(k)} = 0 \qquad \forall k \geq 3.$$

The fixed point is described by $\omega^*(z) = -1/2\,\omega_{(2)} z^2$, where $\omega_{(2)} = \mathrm{var}(X) = \sigma^2$. This is nothing but a Gaussian distribution

$$p^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

so with the proper renormalization a fixed point still exists and we recover the central limit theorem (CLT) for centered distributions. However, it is clear that the eigenvalues of the RG transformation now are

$$\lambda_k = 2^{1-k/2}$$

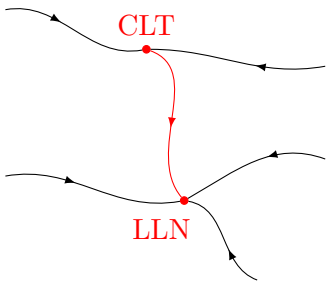


FIGURE 3.1: Renormalization Group flow in the space of probability distributions, where the central limit theorem (CLT) fixed point is unstable and the law of large number (LLN) one is stable.

so the flow is quite different from before. For $k = 2$ we find a shift in the variance of the distribution, which is a marginal perturbation, and for $k > 2$ all perturbations are irrelevant. It is for $k = 1$ that the perturbation becomes relevant, and this is not unexpected since λ_1 is the eigenvalue associated with the first cumulant: if the mean is not exactly zero we do not expect to find a Gaussian distribution in the limit, but rather the law of large number. Hence a perturbation that changes the first moment flows away from the central limit theorem fixed point, which is thus unstable as in Figure 3.1[2].

This shows how we may think of the RG flow as a flow in the space of probability distributions. In general, if the variables are weakly correlated, the coarse-graining drives the joint distribution towards a Gaussian distribution, which is a fixed point of the transformation [27]. This is the case, for instance, of the supercritical Ising model or the supercritical contact process and it is reminiscent of the flow of a block spin transformation. In fact, if $T > T_c$ eventually the configurations become random and the joint distribution is naturally described by the central limit theorem.

One should be careful because this picture is somewhat different with respect to the usual one in Statistical Mechanics, where fixed points lie in the critical surface. We usually think of the Gaussian fixed point as the fixed point of critical systems embedded in a dimension d higher than their upper critical dimension d_u, where correlations are weak and mean field theory applies. Instead, let us stress out that in this probabilistic view the Gaussian fixed point is trivial in the sense that it is a consequence of the central limit theorem, hence it appears for independent or weakly correlated variables, away from criticality.

For a critical system the variables are strongly correlated and the previous derivation does not hold anymore. We might expect that the non-trivial structure of the correlations gives rise to a non-trivial fixed point, i.e. a fixed point for which the central limit theorem does not hold. However, in general analytical calculations are hard and it is not possible to find such fixed point exactly. In Appendix A we give a brief

[2]It is worth noting that even for independent random variables the central limit theorem is not a fixed point if the first two moments of $p(x)$ do not exist, giving rise to a larger class of stable distributions and the so called generalized central limit theorems. For instance, an interesting case is the one of probability distributions that have power law tails, for which such class can be analytically found. The Levy distribution belongs to it, for instance.

account of what happens in an exactly solvable case, the one of a hierarchical model, to better understand what behavior we expect for the joint probability distribution of coarse-grained variables at criticality and away from it.

3.1.1 COARSE-GRAINING IN FOURIER SPACE AND PCA

Of course, working in real space is not the only option we have. Instead of performing the coarse-graining step via Equation (3.1), let us consider consider the Fourier representation of the variables

$$s(\mathbf{k}) = \frac{1}{\sqrt{N}} \sum_i e^{-i\mathbf{k}\cdot\mathbf{x}_i} \sigma_i$$

where $\mathbf{x}_i$ is the d-dimensional vector that describes the position of the site attached to the variable σ_i. We may think that averaging over the microscopic details of the system in real space is equivalent to averaging over the high frequency Fourier modes. Hence the coarse-graining procedure amounts to place some cutoff Λ in order to keep only the variables corresponding to modes such that $|\mathbf{k}| < \Lambda$, namely

$$\sigma_i \to \tilde{\sigma}_i = \frac{z_\Lambda}{\sqrt{N}} \sum_{|\mathbf{k}|<\Lambda} e^{i\mathbf{k}\cdot\mathbf{x}_i} s(\mathbf{k}).$$

There is a slight difference with respect to the real space coarse-graining, since now it is not the number of variables in the system that changes, but the number of Fourier modes. This is not important in the thermodynamic limit $N \to \infty$, but it is relevant for finite systems, as we will see.

It is interesting to look at what happens in this setting to the covariance matrix

$$C_{ij} = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle . \tag{3.4}$$

In a system with translational invariance each element of the covariance matrix is given by $C(\mathbf{x}_i - \mathbf{x}_j)$ for some function C, whose Fourier transform is given by

$$\begin{aligned} C(\mathbf{k}, \mathbf{q}) &= \frac{1}{N} \sum_{i,j} C(\mathbf{x}_i - \mathbf{x}_j)\, e^{-i\mathbf{x}_i\cdot\mathbf{k}} e^{-i\mathbf{x}_j\cdot\mathbf{q}} \\ &= \frac{1}{N} \sum_{n,j} C(\mathbf{x}_n)\, e^{-i\mathbf{x}_n\cdot\mathbf{k}} e^{-i\mathbf{x}_j\cdot(\mathbf{k}+\mathbf{q})} \\ &= \frac{1}{\sqrt{N}} \delta_{\mathbf{k},-\mathbf{q}} \sum_n e^{-i\mathbf{x}_n\cdot\mathbf{k}} C(\mathbf{x}_n) \\ &= \delta_{\mathbf{k},-\mathbf{q}} G(\mathbf{k}) \end{aligned}$$

where

$$G(\mathbf{k}) = N^{-1/2} \sum_n e^{-i\mathbf{x}_n\cdot\mathbf{k}} C(\mathbf{x}_n).$$

Hence we find the matrix

$$C_{ij} = \frac{1}{N} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{x}_i - \mathbf{x}_j)} G(\mathbf{k}),$$

which means that in Fourier space the correlation matrix is diagonal[3]. In fact, if we consider the eigenvalue equation for C_{ij}

$$\sum_j C_{ij} u_j = \lambda u_i$$

and we write the eigenvectors as $u_j = N^{-1/2} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}_i} u_{\mathbf{k}}$ we find

$$\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}_i} G(\mathbf{k}) u_{\mathbf{k}} = \sum_{\mathbf{k}} \lambda_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}_i} u_{\mathbf{k}}$$

which means that the eigenvalues are given by the Fourier transform of the correlation function $G(\mathbf{k})$. This has a non trivial implication for the eigenvalue spectrum of the covariance matrix in a critical system, where we expect the algebraic decay

$$G(r) \sim r^{-(d-2+\eta)}.$$

Since the eigenvalues are the Fourier transform of the correlation function, we shall write

$$\begin{aligned}
\lambda_{\mathbf{k}} &= \int d^d r \, e^{i\mathbf{k}\cdot\mathbf{r}} r^{-(2-d+\eta)} \\
&= \int dr d\Omega_d \, r^{1-\eta} \, e^{i|\mathbf{k}| r \cos\theta} \\
&\sim \frac{1}{|\mathbf{k}|^{2-\eta}}.
\end{aligned}$$

If this is a decreasing function of $|\mathbf{k}|$, that is if $\eta < 2$, then we consider a ranking of eigenvalues from small momentum to large momentum. Hence the highest eigenvalue has rank $r = 1$, which implies

$$\begin{aligned}
r[\lambda_{\mathbf{k}}] &= \sum_{\mathbf{k}'} \mathbb{I}[\lambda_{\mathbf{k}'} > \lambda_{\mathbf{k}}] = \sum_{\mathbf{k}'} \mathbb{I}[|\mathbf{k}'| < |\mathbf{k}|] \\
&\approx L^d \int d^d k' \, \theta(|\mathbf{k}'| < |\mathbf{k}|) \\
&\sim (L|\mathbf{k}|)^d.
\end{aligned}$$

[3] This is not very surprising, since in the case of translational invariance C_{ij} is the matrix representation of a convolution, which is diagonal in the Fourier basis. However, even for systems without translational invariance it is possible to employ perturbation theory and study the spectrum of the covariance matrix and its behavior under coarse graining, see [8].

This gives that the eigenvalues of the covariance matrix should decay as a power law of their rank, namely

$$\lambda_r \sim \frac{1}{r^{\mu}} \tag{3.5}$$

where $\lambda_1 \geq \lambda_2 \geq \dots$ and $\mu = (2-\eta)/d$.

Let us stress one more thing. Since the higher wavevector corresponds to the smaller eigenvalues, coarse-graining by a cutoff Λ amounts to averaging over the smaller eigenvalues. This is strikingly similar to what is usually done in principal component analysis (PCA) in the framework of unsupervised learning algorithms. If we consider the set of ranked eigenvalues $\{\lambda_l\}$ and the corresponding orthonormal eigenvectors $\{\mathbf{u}_l\}$ of the covariance matrix (3.4) we can define the family of projection operators

$$P_{ij}(K) = \sum_{l=1}^{K} u_{il} u_{jl} \tag{3.6}$$

that project the variables onto the subspace spanned by the K eigenvectors associated with the K largest eigenvalues. Then the coarse-grained variables can be defined by means of

$$\sigma_i \to \tilde{\sigma}_i = \sum_{j=1}^{N} P_{ij}(K)\sigma_j.$$

The idea is very much the same of PCA, that the most information is contained where the variance is higher, but there is a key difference. Let us denote by X_{ij} the $M \times N$ matrix of N variables sampled M times, so each column represent the observations of a single variable $\mathbf{x}_i$. Principal component analysis is most often used as a method of dimensionality reduction to a new set of D variables $\{\mathbf{y}_i\}_{i=1}^{D}$ defined as

$$Y_{ml} = \sum_{n} X_{mn} u_{nl}, \quad l = 1, \dots, D \tag{3.7}$$

where Y_{ml} is a $M \times D$ matrix and u_{nr} is the matrix whose columns are the ranked eigenvectors, which are called principal directions. The projection of the data on a principal direction is called principal component. The usual choice is $D = 2$, so that only the first two principal components are kept. Dimensionality reduction makes sense if the D modes effectively capture a large fraction of the total variance, that is

$$\frac{\sum_{i=1}^{D} \lambda_i}{\sum_{j=1}^{N} \lambda_j} \approx 1$$

which implies the presence of a spectral gap between λ_D and λ_{D+1}. Interestingly enough, as we will see further in this chapter, this is the case of the subcritical Ising model. A brief review of PCA and a more detailed account between its relation and the coarse-graining procedure is given in Appendix B.

3.1.2 COARSE-GRAINING THE CORRELATIONS SPACE?

We are now in a position to present the coarse-graining procedure introduced in [34]. We will comment on their meaningfulness further on, when we apply the procedure to various systems.

In a network of neurons, we expect the interaction to be long-range due to the synaptic connections between far away neurons: hence, coarse-graining in real space would fail because of the non-locality of the interactions. The authors proposed to build clusters variables by grouping together neurons that are most correlated. If we refer to the initial variables as $\sigma_i^{(1)}$, we search for the maximal non-diagonal element of the correlation matrix

$$c_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}$$

thus finding a pair $(i, j_*(i))$ of maximally correlated variables. We remove those indexes and search again, ending up with a set of pairs$\{i, j_*(i)\}$ and define the coarse-grained variables

$$\sigma_i^{(2)} = \sigma_i^{(1)} + \sigma_{j_*(i)}^{(1)}$$

where $i = 1, \dots, N/2$. We iterate this process, producing clusters of $K = 2, 4, \dots, 2^{k-1}$ variables. Each one defines a new variable $\sigma_i^{(k)}$ as the summed activity of cluster i, so this should be quite similar to coarse-graining in real space.

The authors of [34] analyze the behavior of the following quantities under this coarse-graining procedure, in order to make some parallels with the behavior of critical systems. The mean variance is given by

$$M_2(K) = \frac{1}{N_k} \sum_{i=1}^{N_k} \left[\left\langle \left(\sigma_i^{(k)}\right)^2 \right\rangle - \left\langle \sigma_i^{(k)} \right\rangle^2 \right] \tag{3.8}$$

where $\langle \cdot \rangle$ represents the average over the time-series of the activity and N_k is the number of variables after k steps. If the variables were independent the variance inside the square bracket would be additive, hence the mean variance should scale linearly as $M_2(K) \propto K$. In particular, the activity of real neurons shows a scaling of the form

$$M_2(K) \propto K^{\tilde{\alpha}}$$

with $\tilde{\alpha} \approx 1.4$.

More generally, we can look at the full distribution of the individual coarse-grained variables. Since a coarse-grained variable $\sigma_i^{(k)}$ is vanishing if and only if all the 2^{k-1} raw variables are zero, we can write

$$P\left(\sigma_i^{(k)}\right) = P_{\text{silence}}(K)\delta\left(\sigma_i^{(k)}, 0\right) + \left[1 - P_{\text{silence}}(K)\right] A_K\left(\sigma_i^{(k)}/K\right)$$

where $A_K\left(\sigma_i^{(k)}/K\right)$ is the probability distribution of the normalized activity. Hence, $P\left(\sigma_i^{(k)}\right)$ is the probability distribution of the renormalized variables. If we parametrize the joint distribution inside a cluster of size K as

$$P_K\left(\{\sigma_i^{(1)}\}\right) = \frac{1}{Z_K}\exp\left[-E\left(\{\sigma_i^{(1)}\}\right)\right]$$

we might choose the energy in such a way that $E\left(\{\sigma_i^{(1)}=0\}\right)=0$ so that

$$P_{\text{silence}} = \frac{1}{Z_K}.$$

Thus the authors look at an effective free energy

$$F(K) = \log P_{\text{silence}} \tag{3.9}$$

and for the network of neurons analyzed by the authors a scaling behavior $F \propto -K^{\tilde{\beta}}$ is found, with $\beta \approx 0.88$. Moreover, the bulk of the distribution A_K of the normalized activity inside a cluster appears to be invariant as K grows. More in general, we may look at a fixed form of this probability distribution.

As we have seen from Equation (3.5) at the critical point we expect a simple scaling behavior of the ranked spectrum of the covariance matrix. If we consider the clusters that we build with this coarse-graining procedure, the highest rank is given by the number of variables K and corresponds to the maximum wavevector $k_{\text{max}} \sim a^{-1}$. Since we know that the rank is given by $r[\mathbf{k}] \sim (L|\mathbf{k}|)^d$, we have

$$r[\mathbf{k}_{\text{max}}] = K \sim \left(\frac{L}{a}\right)^d$$

which yields

$$r[\lambda_{\mathbf{k}}] \sim |\mathbf{k}|^d K.$$

Thus we expect the spectrum of the covariance matrix of a cluster of K variables to show the same scaling at different K, that is the collapse

$$\lambda_r \propto \left(\frac{K}{r}\right)^{\mu} \tag{3.10}$$

with $\mu = (2-\eta)/d$. For the network of neurons analyzed by the authors, the exponent is given by $\mu \approx 0.71$.

Finally, we can look at the time evolution of the variables. In particular, we evaluate the mean autocorrelation function of the system

$$C^{(k)}(t) = \frac{1}{N_k}\sum_i C_i^{(k)}(t) \tag{3.11}$$

where

$$C_i^{(k)}(t) = \frac{\langle \sigma_i^{(k)}(t_0)\sigma_i^{(k)}(t_0+t)\rangle - \langle \sigma_i^{(k)}\rangle^2}{\langle (\sigma_i^{(k)})^2\rangle - \langle \sigma_i^{(k)}\rangle^2}.$$

In a critical system we might expect dynamical scaling, which would imply that the behavior at different levels of coarse-graining would collapse onto a single curve,

$$C^{(k)}(t) = C\left[\frac{t}{\tau_c(K)}\right] \tag{3.12}$$

with $\tau_c(K)$ the correlation time. The scaling of the correlation time found by the authors is $\tau_c \propto K^{\tilde{z}}$, with an exponent $\tilde{z} \approx 0.16$.

A different test can be performed by exploiting the similarity between PCA and coarse-graining in momentum space. As we have said, coarse-graining in momentum space amounts to averaging over the high frequency Fourier modes, which in turns amounts to averaging over low variance contributions in the covariance matrix. Hence, if we consider the projectors (3.6)

$$P_{ij}(K) = \sum_{l=1}^{K} u_{il}u_{jl}$$

we can define N coarse-grained variables by means of

$$\phi_i(\hat{K}) = z_i(\hat{K}) \sum_j P_{ij}(\hat{K}) \left[\sigma_j^{(1)} - \langle \sigma_j^{(1)}\rangle\right]$$

where $\hat{K}$ is the cutoff, i.e. the highest rank we keep. We do so with

$$z_i(\hat{K}) = \sqrt{\frac{1}{\sum_{jm} P_{ij}P_{im}\left[\langle \sigma_i\sigma_m\rangle - \langle \sigma_i\rangle \langle \sigma_m\rangle\right]}}$$

so that the coarse-grained variables have zero mean and unitary variance, $\langle \phi_i^2(\hat{K})\rangle = 1$. In this setting it is interesting to look at the joint distribution

$$P_{\hat{K}}(\phi) = \frac{1}{N}\sum_i \mathbb{P}\left[\phi_i(\hat{K}) = \phi\right] \tag{3.13}$$

as we change the cutoff $\hat{K}$. In fact, as we have seen the RG transformation typically drives the joint probability distribution of a set of variables towards a fixed point, which for zero-mean and independent variables is exactly the CLT one. Hence, the authors propose to use this RG approach to test whether the joint distribution converges towards a critical fixed point or a Gaussian one.

3.2 TESTS OF THE PROCEDURE

We now want to analyze the quantities introduced in the previous section in the context of the critical contact process, to see if the results reported in [34] for real neurons are in fact the ones of a critical system. However, before moving the non-equilibrium system, we shall test the procedure in simpler environments such as a time series of random binary matrices and the configurations sampled from a Montecarlo simulation of the Ising model.

3.2.1 RANDOM ACTIVITY

In order to find a baseline, in Figure 3.2 we show the main results of the coarse-graining applied at a time series of random binary matrices, which are the results one would expect for independent variables. The size of each matrix, which we may think as a lattice of active and inactive sites, is 50×50.

In Figure 3.2a we can see a randomly distributed correlation matrix, as expected. The variance (3.8) is additive and indeed in Figure 3.2b it scales linearly with the cluster size. Since we are not at criticality, no peculiar scaling of the spectrum of the covariance matrix is present, as can be seen from the eigenvalues in Figure 3.2e, and there is no collapse for clusters of different size.

If we look at the free energy (3.9) defined by the silence probability, i.e. the probability that all variables in a cluster are zero, we find a behavior compatible with the trivial exponent $\tilde{\beta} = 1$,

$$F_K \propto K.$$

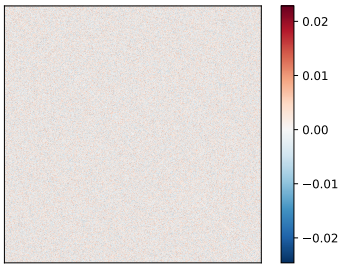
This implies that the silence probability decays exponentially as a function of K,

$$P_{\text{silence}}(K) \sim e^{-K}$$

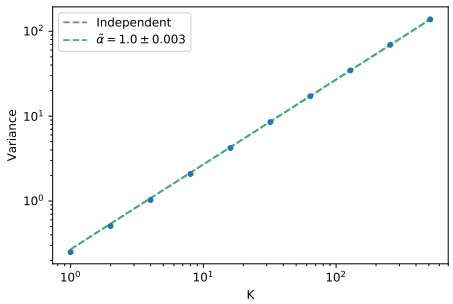
which is not surprising since the variables are independent. One should note that clusters involving $K = 32$ already have a vanishing probability of being completely silent, and this should be regarded as a signature of the randomness of the underlying activity.

Interestingly enough, in Figure 3.2d we see that the coarse-graining drives the distribution of the nonzero normalized activity towards a Gaussian centered in $1/2$ with decreasing variance. It is not hard to imagine that for $N \to \infty$ the limit distribution is going to be a delta function, which is what we would expect from independence and the LLN fixed point. The fact that the center of the joint distribution is exactly at $1/2$ is a signature of randomness, since it reflects the balance between active and inactive sites.

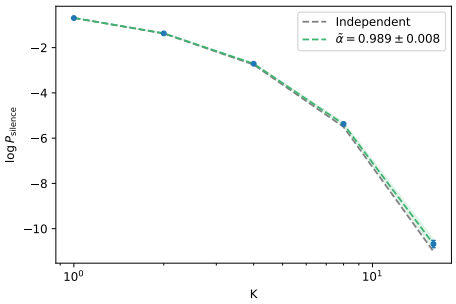
Finally, the momentum shell approach previously described gives the result in Figure 3.2f. The initial distribution, where all the modes are kept, corresponds to a simple standardization of the variables and a change of basis to the eigenvectors of the covariance matrix. Not surprisingly, this gives a superposition of two delta functions. The coarse-graining is extremely fast in driving the distribution towards a Gaussian distribution, which is exactly the CLT fixed point in the space of probability distributions.



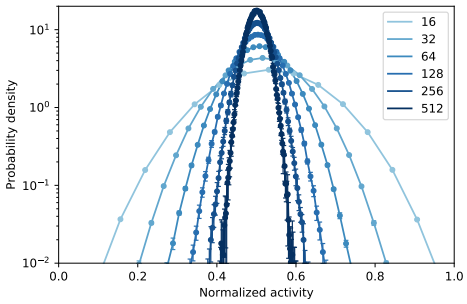
(A) Correlation matrix of the raw variables (diagonal set to zero).



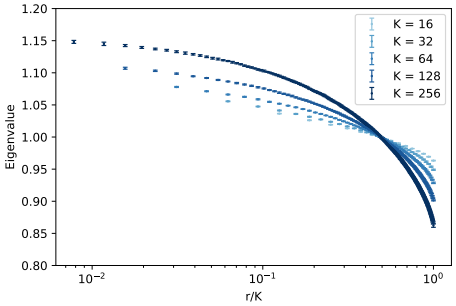
(B) Scaling of the variance.



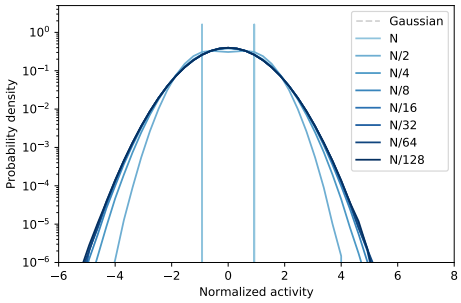
(C) Scaling of the logarithm of the silence probability.



(D) Scaling of distribution of nonzero activity.



(E) Scaling of the eigenvalues of the covariance matrix.



(F) Probability distribution of the projected variables.

FIGURE 3.2: Results of the coarse-graining for a sequence of random configurations.

3.2.2 THE ISING MODEL

The case of the Ising model is emblematic. We are not interested in the dynamics, but rather in what this RG approach can tell us about the static properties of the system.

Let us look first at the supercritical Ising model at $T \approx 4T_c$ in a lattice of 50×50 spins. To sample the Boltzmann distribution, we implement the Wolff algorithm and keep a single configuration every 100 timesteps, in order to sample rather uncorrelated states. We collect approximately 10^6 configurations. Hence for the time being the averages $\langle \cdot \rangle$ should be intended as proper ensemble averages rather than time averages. The results are reported in Figure 3.4.

In Figure 3.4a-3.4b we can see that in the correlation matrix of the raw variables there is a structure typical of nearest-neighbors correlations. However, such structure vanishes after a few steps of the coarse-graining. This suggests that the coarse-graining drives the supercritical system towards a random configuration, much like the case of a block-spin transformation. In fact, the scaling of the variance in Figure 3.4c is compatible with the simple linear scaling we found in the previous section.

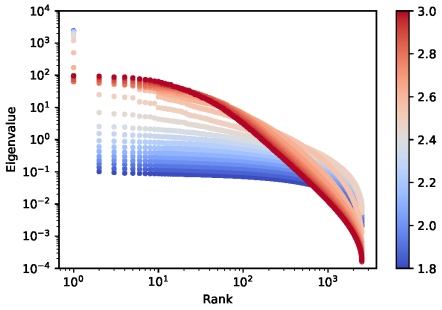
This is further confirmed by both the distribution of the normalized activity and the the distribution of the variables projected over the principal components, see Figure 3.4d and Figure 3.4f. In particular, the system seems to be well described by a trivial Gaussian fixed point. The eigenvalues of the covariance matrix of the clusters are small and show no signals of collapse or power law decay.

Next, we shall consider a subcritical Ising model at $T \approx 0.9T_c$, in the same simulation setting as before. Being in the ordered phase, the system is quite resilient so the spin orientations hardly change between different configurations. The results are reported in Figure 3.5 and in fact show a high correlation between the variables[4]. Once more, much like a block-spin transformation this coarse-graining seems to drive the system towards perfectly ordered configurations, as we can see in Figure 3.5d-3.5f. Notice in particular that averaging over the low-variance modes seems to point towards a trivial result, namely two delta functions, which albeit trivial is not the Gaussian fixed point. One should note that the convergence is much slower with respect to the supercritical regime.

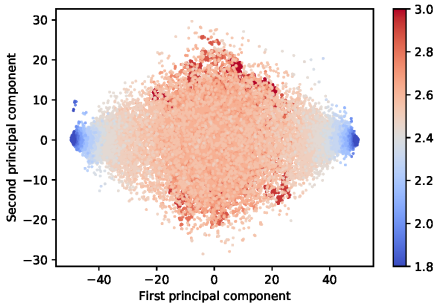
There is one result that is particularly interesting, namely the eigenvalue spectrum of the covariance matrix in Figure 3.5e. There is a clear dominance by a single eigenvalue, meaning that a single principal component contributes to most of the variance of the system. In other words, the subcritical Ising model shows clear signatures of low-dimensionality in the PCA sense previously explained, while this is not true in the supercritical regime. To get a glimpse of this transition it is enough to look at the spectrum of the covariance matrix for different temperatures in Figure 3.3a. The supercritical Ising model shows a continuous eigenvalue spectrum, hence low dimensionality disappears, but as we lower the temperature a gap in the spectrum becomes more and more evident.

If we do apply the dimensionality reduction, as in Equation (3.7), we can plot the first two principal components against each other to obtain Figure 3.3b. What we find

[4]Notice that is the correlation between subsequent configurations, i.e. it is analogous to the correlation in time. The spins in the subcritical Ising model, instead, are weakly correlated among each other.



(A) Spectrum of the covariance matrix.



(B) First two principal components.

FIGURE 3.3: Results of the PCA analysis of the Ising model in a $2D$ lattice.

is that the phase transition becomes quite evident in this low dimensional space. Below the critical temperature the system lives in two disjointed regions, which we may regard as a direct consequence of the $O(1)$ symmetry group of the Ising Hamiltonian, that allows two ground states.

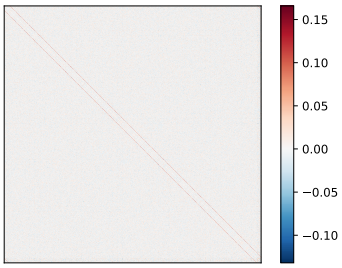
Hence, in the subcritical regime removing the low-variance modes does not affect much the averaged distribution because the overall change is rather marginal. This is the likely reason of the slower convergence towards a fixed point. Moreover, in the thermodynamic limit one might expect that only a single eigenvalue would be different from zero, so that removing the remaining $N - 1$ modes would not affect the system at all. At the same time, the subcritical Ising model shows us that one must be extremely careful when using this coarse-graining approach to look for signatures of criticality: the eigenvalue spectra might be non trivial, and there might be other fixed distributions that are not Gaussian but still are not related to a critical point.

Finally, let us consider the critical Ising model. Notice that we define the finite-size critical temperature as the temperature at which the specific heat is maximum [14], which implies a shift with respect to the real critical point $T_c \approx 2.269 J/k_B$. We still sample the configurations via the Wolff algorithm as before, since it is particularly efficient in the critical case. The results are reported in Figure 3.6.

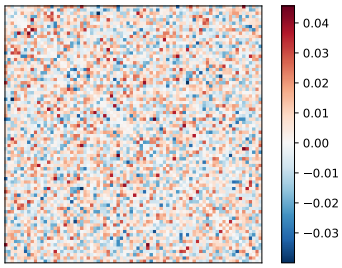
With respect to the previous cases, the correlation matrix of the raw variables displays a non trivial structure that is preserved by the coarse-graining procedure. We do see a power law scaling of the variance at different levels of coarse-graining, with an exponent that lies in between the values found for the supercritical and subcritical regimes. This is also true for the joint distribution of the coarse-grained variables, which seems to approach an intermediate fixed form.

One of the most interesting features of the ordered phase is the low dimensionality. Of course, at criticality the system is not ordered, but the long-range correlations manifest in the presence of islands of aligned spins that exist at all scales. Hence we might foresee that the eigenvalue spectrum will be affected by their presence. The spectrum is showed in Figure 3.6e.

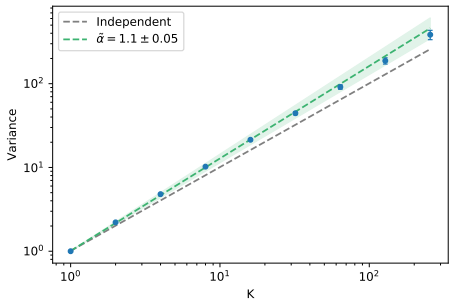
First of all, let us note that we know the exact value of the exponent we expect



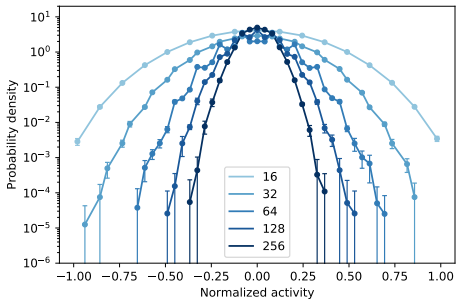
(A) Correlation matrix of the raw variables (diagonal set to zero).



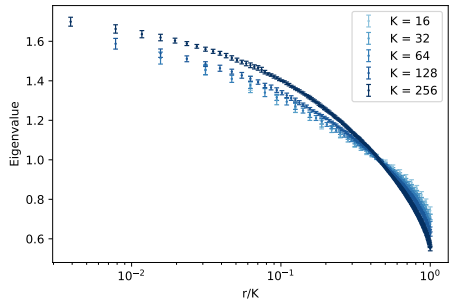
(B) Correlation matrix of clusters of 32 variables (diagonal set to zero).



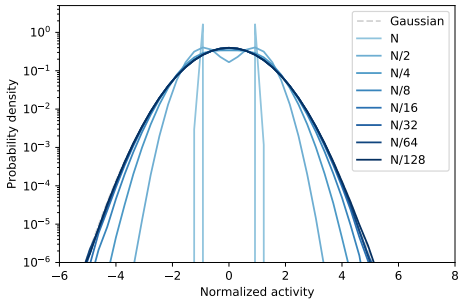
(C) Scaling of variance.



(D) Distribution of the normalized variables.

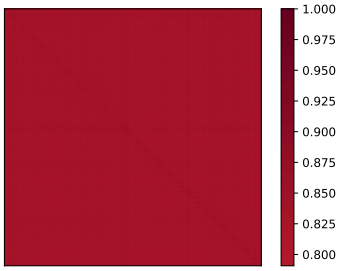


(E) Scaling of the eigenvalues of the covariance matrix.

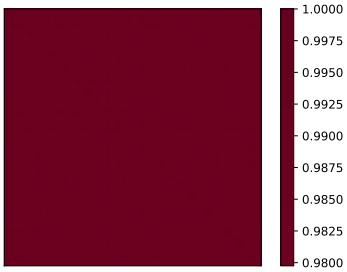


(F) Probability distribution of the projected variables.

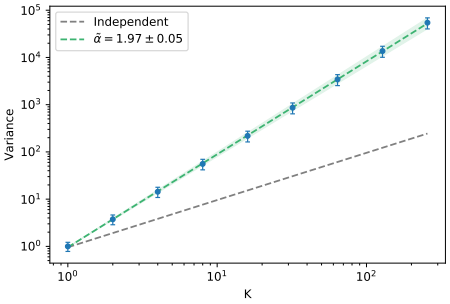
FIGURE 3.4: Results of the coarse-graining for a supercritical Ising model in a $2D$ lattice.



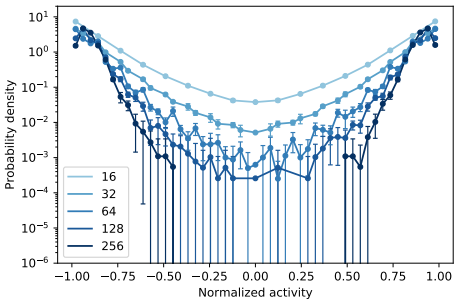
(A) Correlation matrix of the raw variables.



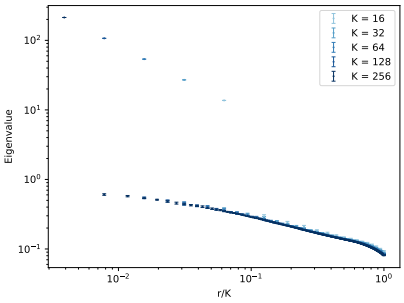
(B) Correlation matrix of clusters of 32 variables.



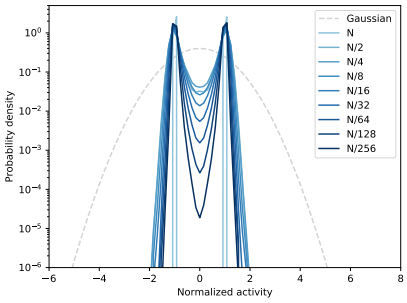
(C) Scaling of variance.



(D) Distribution of the normalized variables.

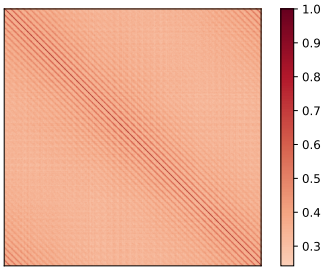


(E) Scaling of the eigenvalues of the covariance matrix.

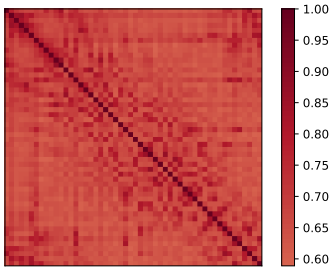


(F) Probability distribution of the projected variables.

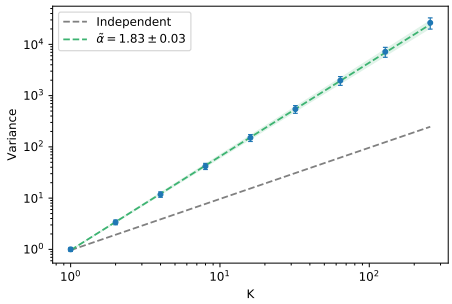
FIGURE 3.5: Results of the coarse-graining for a critical Ising model in a $2D$ lattice.



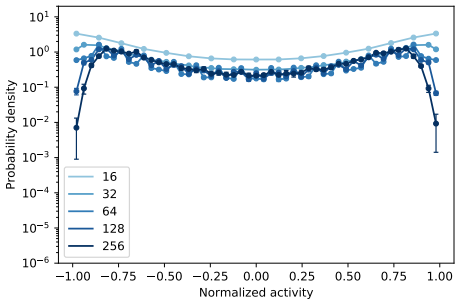
(A) Correlation matrix of the raw variables (diagonal set to zero).



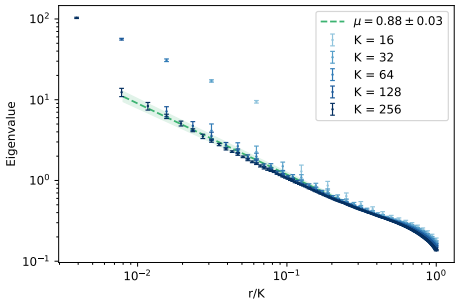
(B) Correlation matrix of clusters of 32 variables (diagonal set to zero).



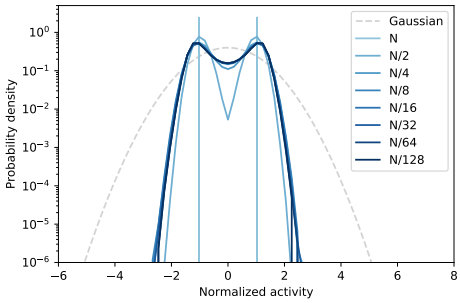
(C) Scaling of variance.



(D) Distribution of the normalized variables.



(E) Scaling of the eigenvalues of the covariance matrix.



(F) Probability distribution of the projected variables.

FIGURE 3.6: Results of the coarse-graining for a critical Ising model in a $2D$ lattice.

to see. The Onsager solution of the 2-dimensional Ising model gives $\eta = 1/4$, which implies

$$\mu = \frac{7}{8} = 0.875.$$

However, the scaling we see is not trivial. The presence of islands of aligned spins does imply a weak dominance of the highest eigenvalue, which is not enough to make the system clearly low dimensional, but makes a true collapse impossible to be seen. We argue that the scaling is indeed present if one considers that both λ_1 and $\{\lambda_i\}_{i>1}$ appear to have the same power law dependence on the fractional rank r/K, where K is the clusters size. If we do this, the critical exponent μ we find is fully compatible with the exact one.

The example of the Ising model highlights how careful we need to be in analyzing the results of this coarse-graining procedure. The Ising model has a trivial symmetry group, that allows for two degenerate ground states only, but one can imagine more complex models where this analysis would be much more complicate. Moreover, if we are to apply this method directly to the data much of the considerations we made here would be harder to come up with. On the other hand, from the results it is rather clear that the critical point separates a regime of highly correlated variables and a regime of weakly correlated ones. Overall these examples do show that the coarse-graining procedure gives at least meaningful results, so it makes sense to move to the non-equilibrium case of the contact process.

3.3 COARSE-GRAINING THE CONTACT PROCESS

We now look at the stationary configurations of the contact process in a $2D$ lattice. Of course, the nature of the transition is quite different from the one of the Ising model. The critical and supercritical systems are mostly different in how often they wander around the absorbing state rather than in how ordered they are. Hence, we might intuitively expect that it will be harder for this coarse-graining procedure to distinguish between these two phases.

We consider a lattice of 41×41 sites and roughly 10^6 configurations in time, both in the critical and the supercritical regime. However, we need to be much more careful with respect to the simulations we carried out in chapter 2. The algorithm 2 is an asynchronous algorithm that performs a continuous time update, which means that every step results in one spin update at most and a typically small time increment. Thus two neighboring configurations in time are extremely similar, if not identical, which means that they are highly correlated.

In the case of an equilibrium simulation our goal would be to sample independent configurations, and to this end we can always discard a sufficiently large number of subsequent configurations to minimize the correlation between them. This is to say that such correlation is induced by the Montecarlo algorithm and is not physical, since there are no dynamical features to begin with. For the contact process, however, it is exactly the dynamical evolution we are interested in. Hence it is not trivial to distinguish among the physical correlation between two configurations and the spurious one due

to our choice of the algorithm. For instance, one could implement a synchronous update where all the spins are updated at once. The underlying dynamics would still be the one of the contact process, but two neighboring configurations would look very differently. On the other hand, if we sample too far in time we might loose the physical correlation as well, plus the problem becomes much more computationally heavy at criticality.

For this very reason we follow the approach of [34] to evaluate the error bars of our results. Since it is not trivial to understand the real number of independent configurations, instead of evaluating the error of the mean as in the usual Montecarlo scheme we divide our sample of 10^6 configurations in smaller time windows. We evaluate all the relevant quantities in each of them, so that errors are simply the standard deviations across different time windows.

Let us call r the sampling rate, that is the number of configurations we discard. To try to understand which might be a good choice for r, we apply the coarse-graining procedure both with $r = 1$, that is keeping every configuration, and with $r = 200$. The latter choice is somewhat justified a posteriori: for the critical contact process the results display a weak dependence on the sampling rate, whereas for the supercritical regime there is a strong difference between $r = 1$ and $r = 200$. The results become more stable for $r > 200$, and this analysis is performed in Appendix C.

In this section we keep $r = 200$ for the critical regime, which is still computationally feasible, and $r = 41^2 = N$ for the supercritical one, since the supercritical contact process hardly reaches the absorbing state and therefore it can be simulated for longer periods of time. We focus on a single realization of the stochastic process, even if we know that the exponents will be slightly different across different realizations. Still, it is instrumental to do so because it is exactly what happens with the experimental data in [34]. Of course, in the case of ours simulations the question of the reproducibility is easily answered: the scaling behavior is consistent throughout different runs, meaning that we see the same features and compatible exponents.

3.3.1 THE CRITICAL CONTACT PROCESS

In Figure 3.7 the results of the coarse-graining of the critical contact process are shown. Let us point out immediately Figures 3.7a-3.7b, that show how the time evolution drives the system towards large fluctuations and often close to the absorbing state. This is the distinctive feature of criticality in systems with absorbing phase transitions. It is also the reason for which keeping the sampling rate r high is extremely heavy, since the simulations starts over every time the absorbing state is reached. The correlation matrix 3.7c-3.7d does seem to have a non trivial structure, which is magnified by the coarse-graining as it was for the Ising model.

Most of the scaling features that are seen by the authors in [34] for a network of active neurons are present for the critical contact process. The variance scales as a power law of the cluster size with the non trivial exponent

$$\tilde{\alpha} = 1.45 \pm 0.02$$

which is not so far from the exponent the authors found in neurons. The behavior of the free energy associated with the silence probability in a cluster (3.9) is compatible with a power law scaling as well, with an exponent

$$\tilde{\beta} = 0.65 \pm 0.03.$$

The decay of the silence probability with the cluster size is slightly slower, meaning that in the simpler model of critical contact process the sites tend to be more active than real neurons. It is significant that we see a non vanishing silence probability even for large clusters, since this will not be true for the supecritical case, nor it was true for random activity.

We now turn to the eigenvalue spectrum for clusters of different sizes in Figure 3.7g. We immediately notice that there are no signs of a low-dimensional structure. This is not surprising since we argued in the previous section that such low dimensionality is related to the spontaneous symmetry breaking that characterize the second order phase transition of the Ising model, and no such thing is present in the contact process. From (2.48) we expect an exponent

$$\mu = \frac{\theta + \delta}{\theta + 2\delta} \approx 0.6$$

and we find the value

$$\mu = 0.63 \pm 0.02$$

Even if the eigenvalues estimation might be quite affected by the finiteness of our sample, the exponent is compatible with the expected one and the spectrum of the covariance matrix shows a clear signature of criticality.

Let us now consider the mean autocorrelation function (3.11), which is plotted for different cluster sizes in Figure 3.7h. As expected for a critical system, the correlation in time is quite persistent and far from decaying exponentially. Such persistence is obviously magnified by the coarse-graining, since we are clustering together highly correlated variables in the first place. Notice however that this is the result most affected by the choice of the sampling rate r, of course, so we should be rather careful in its interpretation. It is also the most striking difference with respect to the network of real neurons, where the correlation in time seems to decay much faster. In fact, the exponent we find for the correlation times in Figure 3.7i is

$$\tilde{z} = 0.49 \pm 0.03$$

and it is considerably higher than what found for neurons, suggesting a faster growth with the cluster size. However, the rescaling in Figure 3.7j works for short times only, and fails in the tails of the correlation function. This effect might very well be a consequence of the sampling rate problem.

Finally, we implement the momentum space approach and we evaluate the joint probability distribution of variables projected into the subspace spanned by the largest eigenvectors of the covariance matrix in Figure 3.7l. It is interesting to compare it with

the distribution of normalized, non-zero activity in Figure 3.7k. As we can see, there seem to be an approach to a non-Gaussian fixed form, whose features such the heavily non-Gaussian tales do resemble the one found in real neurons. Still, the convergence is nowhere as fast as the one of the critical Ising model.

3.3.2 THE SUPERCRITICAL CONTACT PROCESS

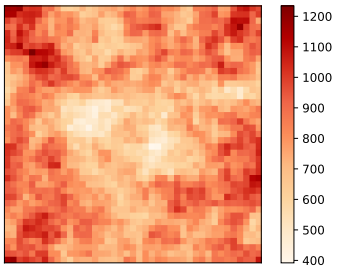
Insofar, our results seem to back the idea that the scaling behaviors found in real neurons arise from an underlying critical dynamic. Nonetheless, we have already noted how the two phases of an absorbing phase transition are not easy to distinguish. Thus it makes sense to ask whether this scaling behavior appears also in the supercritical regime, a state in which the neural activity may very well be. This one of the advantages of this "control case": we can look at the coarse-graining as we vary the control parameter.

The coarse-graining of the supercritical contact process is shown in Figure 3.8. We choose $\lambda = 3$, so we are clearly into the critical phase and activity proliferates. Notice that we also choose $r = 41^2 = N$ since it is not computationally too heavy. As we just noted before, the behavior is different from the critical case in how the number of active sites fluctuates: from Figure 3.8b we see that the system is indeed far from the absorbing state, and shows contained fluctuations around a fixed number of active sites. If we look in general at the time evolution of the system and the correlation matrix of the raw variable in Figure 3.8c there is a clear signature of the nearest neighbors interaction, which is destroyed by the coarse graining: it disappears for clusters of 32 sites already, see Figure 3.8d. Intuitively, we may compare this with thw block spin transformation that drives the supercritical Ising model to a random configuration.

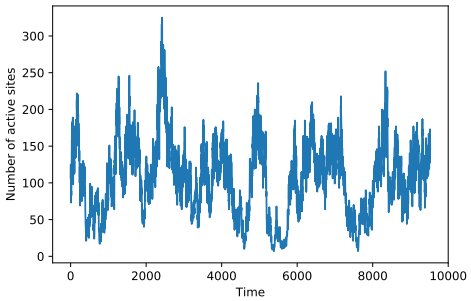
The scaling of the variance, however, is not completely gone. Nor is the power law scaling of the free energy (3.9), albeit there is a clear distinction with respect to the critical case. In fact, as for random activity in clusters of $K > 32$ variables the silence probability vanishes, meaning that no cluster is made up of silent variables only. Even if this is an obvious difference between the supercritical and the critical regime, the scaling for $K \leq 32$ does suggest some carefulness in interpreting these results for small systems. In fact, the exponent $\tilde{\beta} \approx 0.84$ is far from the independent case.

The case of the autocorrelation function is quite emblematic. The decay is evidently faster than before, as we expect, even if this depends heavily on the sampling rate r. Notably, the correlation times in Figure 3.8i hardly scale at all and even if we try to fit them with a power law the exponent $\tilde{z} \approx 0.08$ is extremely small. On the other hand, the autocorrelation functions do collapse if we rescale the time axis, but one should note the collapse by itself could also be a sign of a simple exponential decay. Arguably the fact that the correlation times do not scale as a power law is much more significant. To some extent, the behavior of the autocorrelation function of real neurons seems to be much more similar to the one of the supercritical contact process than the critical one, as pointed out before.

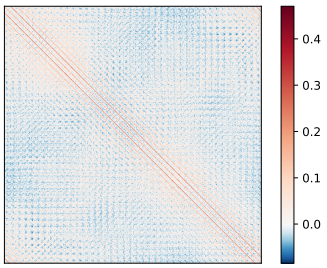
The scaling of eigenvalues is very different from the critical case as well. In particular, one should note that in Figure 3.8g the eigenvalues are extremely small, hence they could be quite affected by sampling problems. We do see a collapse, but rather than



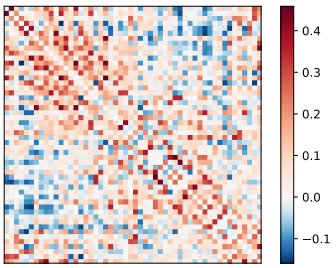
(A) Density of activated sites during the simulation.



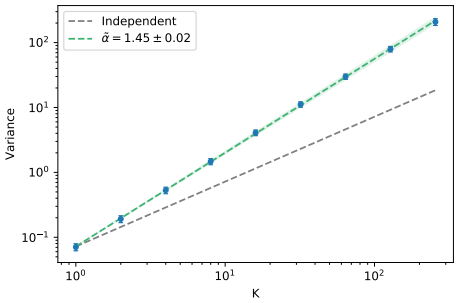
(B) Fluctuations of the numbers of active sites.



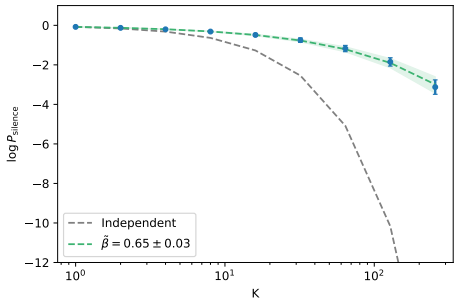
(C) Correlation matrix of the raw variables (diagonal set to zero).



(D) Correlation matrix of clusters of 32 variables (diagonal set to zero).

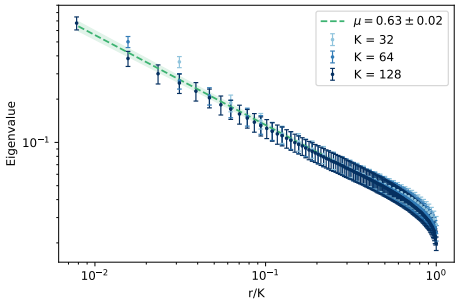


(E) Scaling of the eigenvalues of the variance.

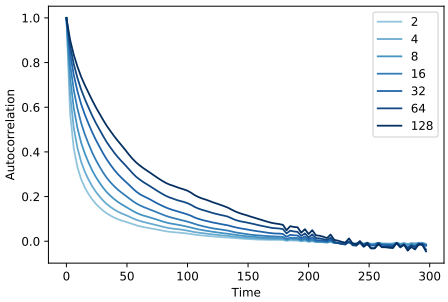


(F) Scaling of the logarithm of the silence probability.

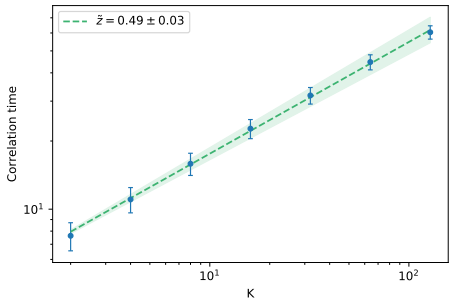
FIGURE 3.7: Results of the coarse-graining for a critical contact process in a $2D$ lattice, with $r = 200$.



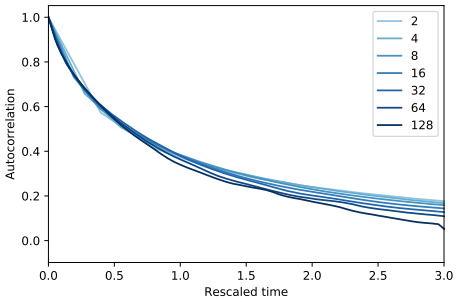
(G) Scaling of the eigenvalues of the covariance matrix.



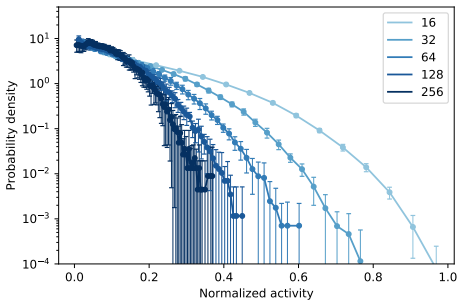
(H) Autocorrelation function.



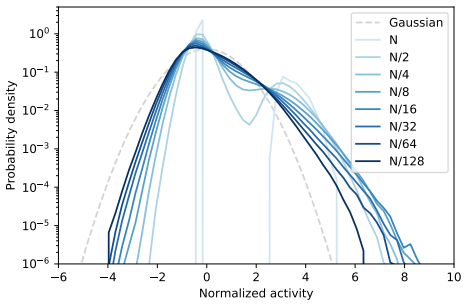
(I) Scaling of the correlation times.



(J) Autocorrelation for rescaled times.

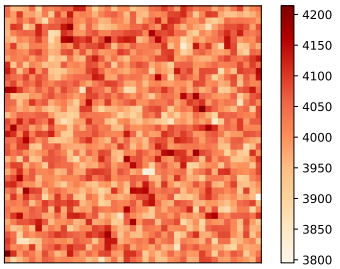


(K) Distribution of normalized non-zero activity.

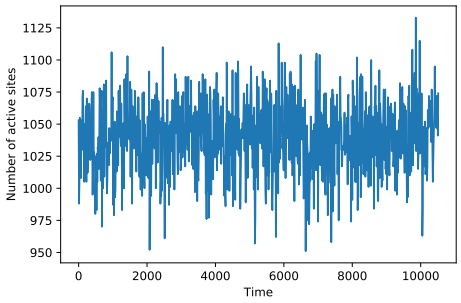


(L) Probability distribution of the projected variables.

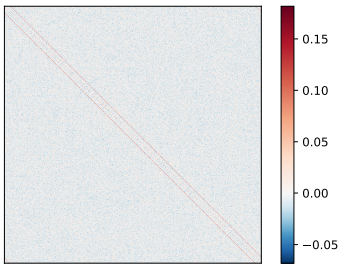
FIGURE 3.7: Results of the coarse-graining for a critical contact process in a $2D$ lattice, with $r = 200$ (cont.).



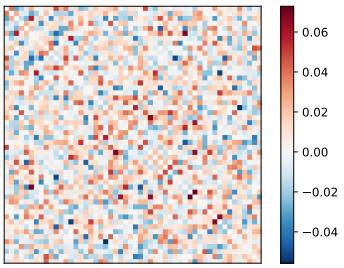
(A) Density of activated sites during the simulation.



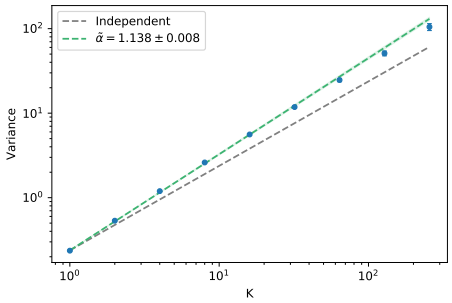
(B) Fluctuations of the numbers of active sites.



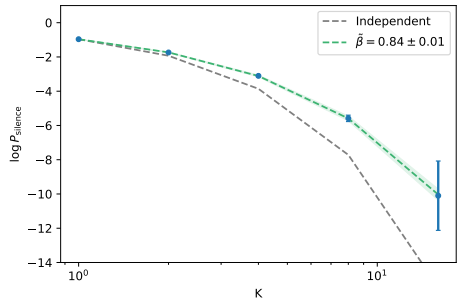
(C) Correlation matrix of the raw variables (diagonal set to zero).



(D) Correlation matrix of clusters of 32 variables (diagonal set to zero).

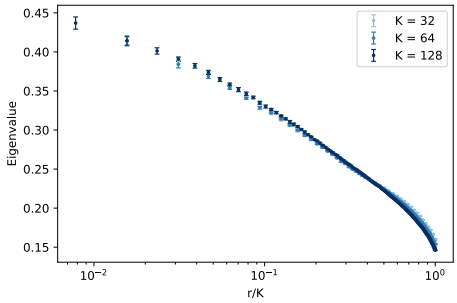


(E) Scaling of the eigenvalues of the variance.

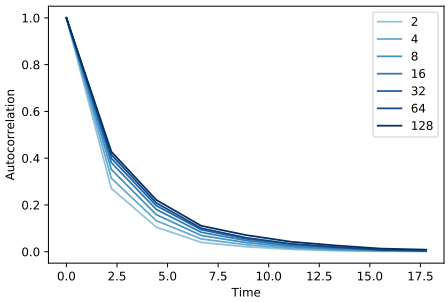


(F) Scaling of the logarithm of the silence probability.

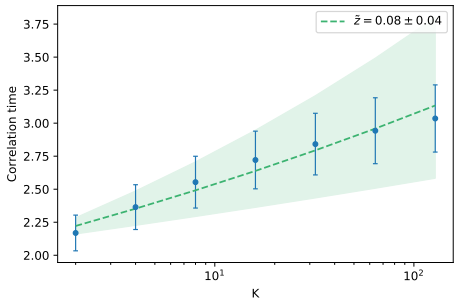
FIGURE 3.8: Results of the coarse-graining for a supercritical contact process in a $2D$ lattice, with $r = N$.



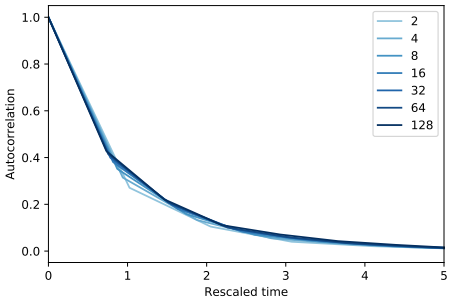
(G) Scaling of the eigenvalues of the covariance matrix.



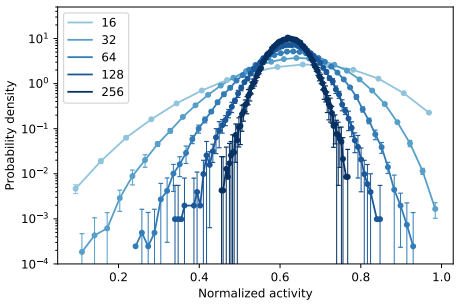
(H) Autocorrelation function.



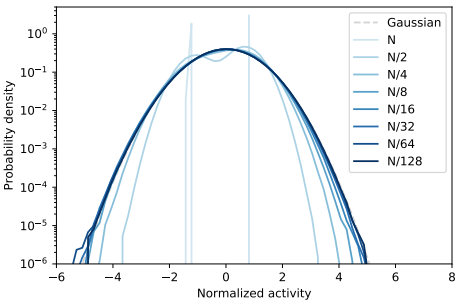
(I) Scaling of the correlation times.



(J) Autocorrelation for rescaled times.



(K) Distribution of normalized non-zero activity.



(L) Probability distribution of the projected variables.

FIGURE 3.8: Results of the coarse-graining for a supercritical contact process in a $2D$ lattice, with $r = N$ (cont.).

being a signature of criticality it should be interpreted as a signature of uncorrelated variables. There is no underlying structure of the correlations.

The result for the normalized non-zero activity, Figure 3.8k, are quite similar to the ones of random variables. The Gaussian-like shape of the distribution is a consequence of the fact that the number of active sites tends to fluctuate around a given value. Moreover, since that activity tends to proliferate in all sites it further implies that the distribution is not centered around $1/2$, but at a higher value. Apart from this, we do not see any particular invariance and the distribution itself is quite different from the critical case.

Still, we believe that the most clear and significant is the one obtained from the coarse-graining in momentum space, in Figure 3.8l. What we see is obviously a consequence of the spectrum of the covariance matrix, that is of the weak correlation of the variables: there is a quite fast convergence of the joint distribution to the CLT fixed point.

Overall, to some extend the coarse-graining procedure does seem to distinguish between the critical and the supercritical contact process. There are, however, many caveats and we might argue that the most promising approach is the one that tackles the correlation structure, both in time and in space: the scaling of the variance and the free energy are much less convincing, especially if we think about applying them in a system where we know almost nothing about the dynamics, like a network of real neurons. In order to further investigate this point, we first try to introduce long range interactions in our problem, and then try to see what happens in a supercritical system that is near criticality, but not quite there.

3.3.3 THE CONTACT PROCESS ON A SMALL-WORLD NETWORK

We now investigate the contact process on a small-world network. In particular, we build a Watts-Strogatz network with an average node degree $\langle k \rangle = K$ and a rewiring probability p. We can think of this topology as the result of N nodes initially ordered in a one-dimensional lattice with periodic boundary conditions, each with K links to its nearest neighbors. Then all the nodes are visited clockwise and each link is rewired with probability p with another node as in Figure 3.9. In this way we naturally introduce long-range interactions in the model, and if p is not too high we are not close to the trivial case of a fully connected network. This is also not too dissimilar from the case of real neurons, where long synaptic connections are present together with local clustering.

The critical behavior of the contact process on small-world networks has been studied numerically in [15], since for low values of p analytical approximations fail. In particular, the critical point depends on the rewiring probability, and for the value $p = 0.001$ it is given by $\lambda \approx 1.7961$. Since we have just seen that some of the scaling behaviors seem to persist in the supercritical case, we are interested in looking at the supercritical contact process in order to check whether the presence of long range interactions strengthen these spurious signatures. Thus, we perform the simulation at $\lambda = 3$.

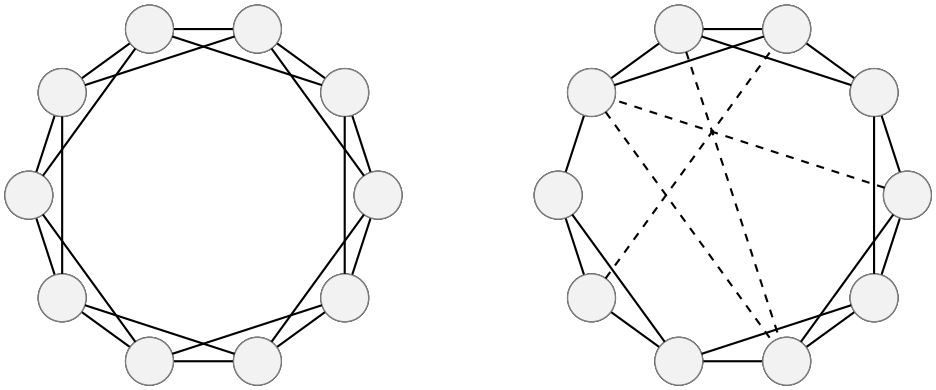


FIGURE 3.9: An example of the procedural generation of a Watts-Strogats network with $\langle k \rangle = 2$. Notice that in the original ring network each node has exactly $k = 2$, which is only true on average for the WS network, but the global number of links is conserved.

Moreover, we do so by implementing a synchronous update algorithm to see if there is any difference with the asynchronous choice. The algorithm works essentially in the same way as Algorithm 2, with a major difference. Let us call $\mathbf{x}(t)$ the state vector of the system at a given time. For every occupied site x_i in $\mathbf{x}(t)$, with probability $p_\lambda = \lambda/(1+\lambda)$ a neighboring empty site x_j is occupied and with probability $a - p_\lambda$ the site x_i is emptied. In both cases, we update their state in a new vector $\mathbf{x}_{\text{new}}$ without changing $\mathbf{x}(t)$. Once every occupied site of $\mathbf{x}(t)$ has been visited, we set $\mathbf{x}(t + \Delta t) = \mathbf{x}_{\text{new}}$. Hence, at each step of the algorithm we update the whole network.

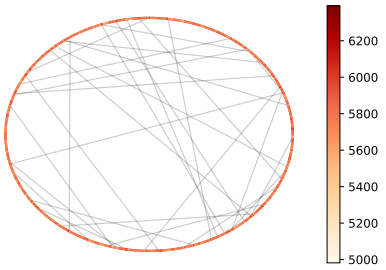
The results of the coarse-graining are shown in Figure 3.10, and it is clear that they hardly changed with respect to the $2D$ lattice. This means that the coarse-graining is indeed stable with respect to the presence of long range interactions, which is exactly the reason we introduced this coarse-graining in the first place. It also means that the caveats that were present for the supercritical regime are neither strengthened nor weakened in the small-world case.

3.3.4 PERSISTENCE OF THE SCALING NEAR CRITICALITY

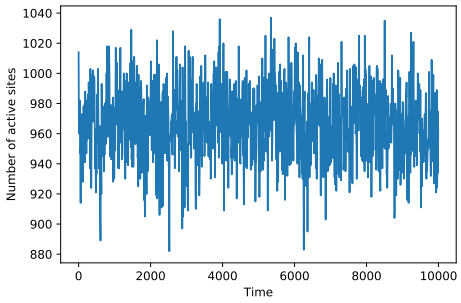
The last test we want to perform, but arguably the most important one, is the stability of these results if we consider a supercritical contact process with a value of λ near the critical one. This is an interesting question, since one might wonder whether the activity in the brain is slightly supercritical rather than critical, and a well defined coarse graining procedure should be sensible enough to these kind of changes.

Let us look at the small-world network previously analyzed, so that we include long range interactions. We perform the simulation for $\lambda = 1.95$, which is roughly a 10% increase with respect to the critical point and thus corresponds to a slightly supercritical activity. The results are shown in Figure 3.11 and as we can see from the number of active sites it is clear that the dynamic is not critical at all: fluctuations are rather bounded and the system is far from the absorbing state.

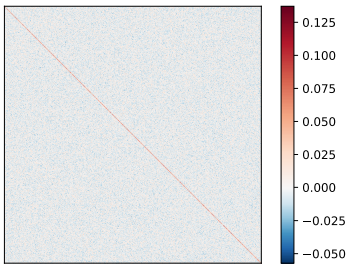
The scaling of the variance and of the free energy are very well present, as if we were looking at a critical system. This confirms the considerations we pointed out for



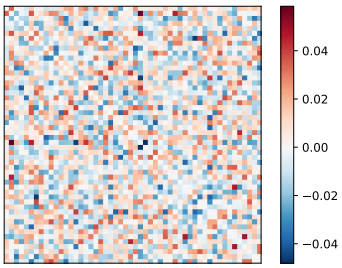
(A) Density of activated sites during the simulation.



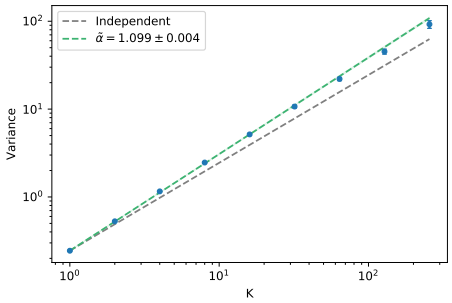
(B) Fluctuations of the numbers of active sites.



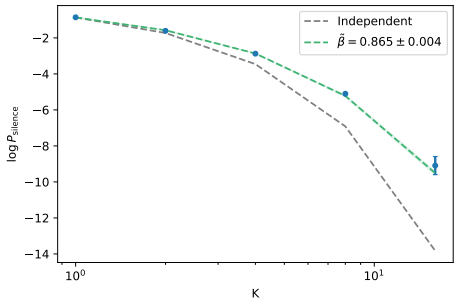
(C) Correlation matrix of the raw variables (diagonal set to zero).



(D) Correlation matrix of clusters of 32 variables (diagonal set to zero).

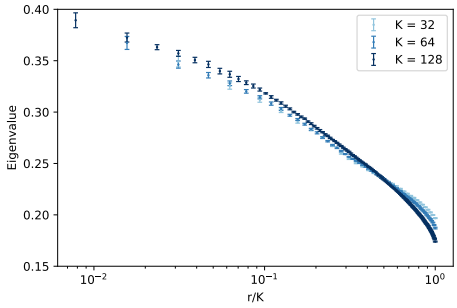


(E) Scaling of the eigenvalues of the variance.

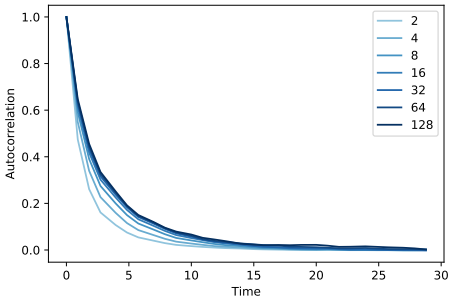


(F) Scaling of the logarithm of the silence probability.

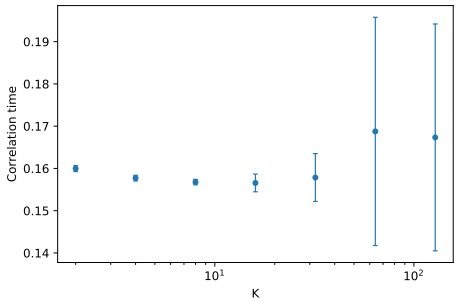
FIGURE 3.10: Results of the coarse-graining for a supercritical contact process in a small-world network.



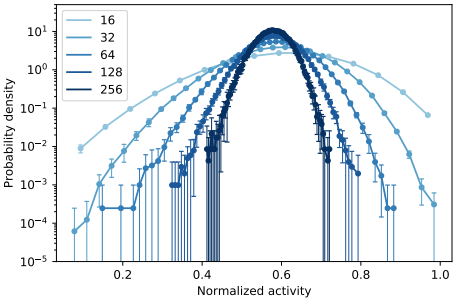
(G) Scaling of the eigenvalues of the covariance matrix.



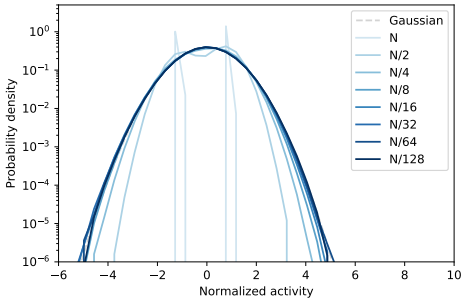
(H) Autocorrelation function.



(I) Scaling of the correlation times.



(J) Distribution of normalized non-zero activity.



(K) Probability distribution of the projected variables.

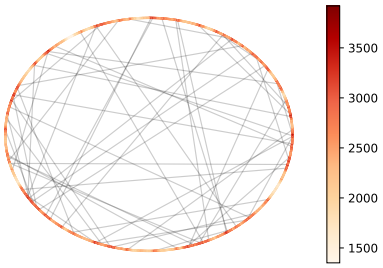
FIGURE 3.10: Results of the coarse-graining for a supercritical contact process in a small-world network (cont.).

the supercritical contact process, where the exponents $\tilde{\alpha}$ and $\tilde{\beta}$ were different from 1 but arguably close. With the results in Figure 3.11e-3.11f it is clear that these two quantity are not really able to gauge the presence of an underlying critical dynamics.

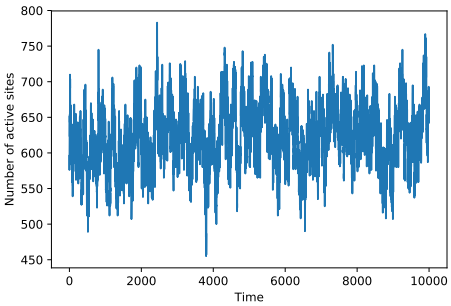
The spectrum of the covariance matrix does not show a power law scaling, but its absence is not as clear as for the contact process deeper in its supercritical phase. Still, the range of values that the eigenvalues take is small, and the plot of Figure 3.11g does not point at criticality. The same considerations can be made for the autocorrelation functions and the scaling of the correlation times: even if they are not as clear as the ones of a deep supercritical phase, they are not compatible with the critical one either. For instance, the power law scaling is not present for the correlation times but they are not as constant as in Figure 3.10i. We might say that these are mixed results: they are good indicators of the absence of criticality, but at the same time they are not fully compatible with the ones we find in the supercritical contact process.

The most interesting result is in Figure 3.11k. Averaging over low variance modes seem to be a quite powerful tool that is capable of distinguish this near-critical case from the critical one. In fact, we do see the convergence to the CLT fixed point, albeit it is slower as we might expect from the spectrum of the covariance matrix.

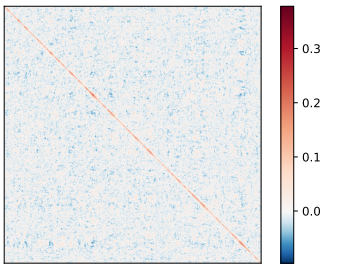
Overall, let us say that the methods that are related to the correlation structure seem to be the most reliable in understanding whether a system is critical. This is not surprising: criticality is deeply related to the correlations, as we saw from the scaling hypothesis and the divergence of the correlation length. This coarse-graining procedure does have the advantage of being very stable with respect to the presence of long range interactions, which is not a trivial statement. However, its results must be interpreted with carefulness, and it is not obvious that they provide a definitive answer to the question of the presence of criticality.



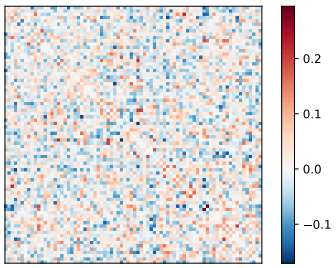
(A) Density of activated sites during the simulation.



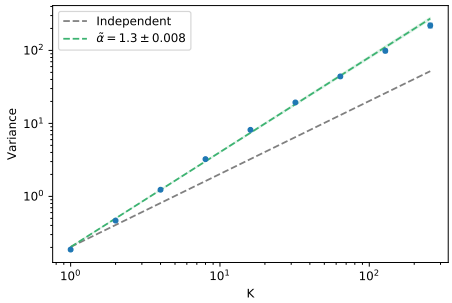
(B) Fluctuations of the numbers of active sites.



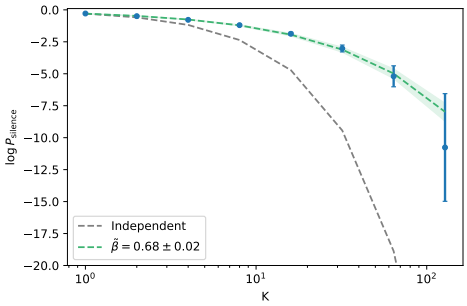
(C) Correlation matrix of the raw variables (diagonal set to zero).



(D) Correlation matrix of clusters of 32 variables (diagonal set to zero).

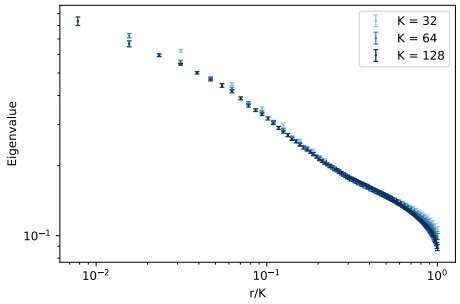


(E) Scaling of the eigenvalues of the variance.

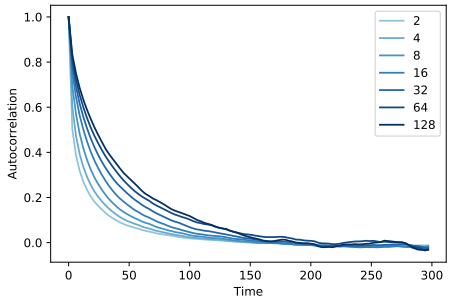


(F) Scaling of the logarithm of the silence probability.

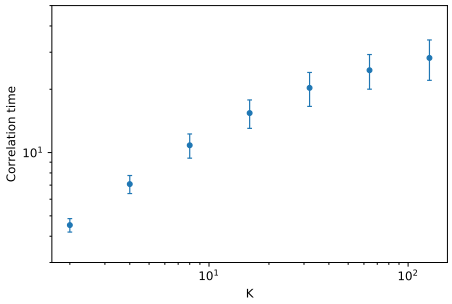
FIGURE 3.11: Results of the coarse-graining for a near-critical contact process in a small-world network.



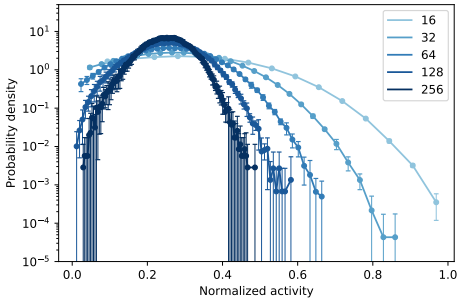
(G) Scaling of the eigenvalues of the covariance matrix.



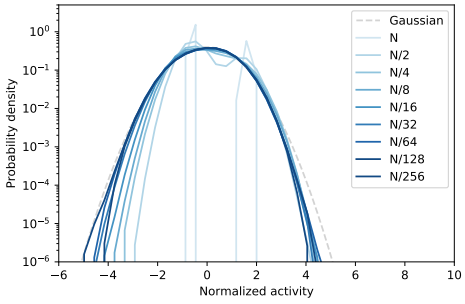
(H) Autocorrelation function.



(I) Scaling of the correlation times.



(J) Distribution of normalized non-zero activity.



(K) Probability distribution of the projected variables.

FIGURE 3.11: Results of the coarse-graining for a near-critical contact process in a small-world network (cont.).

CHAPTER 4

CONCLUSIONS

The most considerable advantage of the phenomenological Renormalization Group proposed in [34] is that it is model independent, in the sense that it can be applied directly to the data. Indeed, given a series of configurations of a system we are always able to estimate the covariance matrix, which is the core of this procedure, even if we do not have an explicit model. This is fundamental in most of the broader applications, because it is often very hard to write down an Hamiltonian that fully accounts for the behavior of complex systems. On the other hand, the absence of a model makes it particularly important to have a comprehensive understanding of the connections between the results of this phenomenological approach and a possible presence of an underlying criticality. This is the direction in which this thesis moves.

We believe that the contact process proves to be an optimal control case to investigate the ability of a phenomenological Renormalization Group to extract signature of criticality. Indeed, we are able to test the RG in the presence of a peculiar kind of phase transition, both in different regions of the control parameter and in different topologies. In particular, the latter allow us to clearly show that the procedure proposed in [34] is stable with respect to the long-range connections that appear in a small-world network. This is promising since it implies that in principle one does not need to worry about the presence of long-range interactions, which might not be known a priori.

Nonetheless, the most important test for the Renormalization Group is how its results change as we vary the control parameter across the phase transition and deep into the supercritical case. A crucial point is its ability to clearly discriminate one phase from the other, since in principle their properties stem from different fixed points of the RG transformation. Aside from the contact process, we must say this phenomenological RG works quite well in equilibrium settings like the Ising model. Admittedly it is a simpler environment where we do not expect to have correlations among subsequent configurations, i.e. correlations in time, and the spontaneous symmetry breaking makes

a subcritical configuration qualitatively different from a supercritical or a critical one. Still, even the Ising model proves to be not trivial, because the spontaneous symmetry breaking makes the system low dimensional in the sense of PCA.

Albeit interesting, statistical criticality arguably plays a smaller role with respect to dynamical criticality in natural systems. Hence our focus on the contact process in this thesis, which is also an archetypal model for neural activity itself. Since the distinction between the supecritical regime and the critical one is less evident than the Ising model, the first question one should ask is whether the results of the phenomenological RG are different in the two phases. In principle, one should expect that for $\lambda > \lambda_c$ the coarse-grained variables would eventually become independent because they are weakly correlated. However, the exponents of the variance $\tilde{\alpha}$ and of the silence probability $\tilde{\beta}$ turn out to be not exactly 1 as we expect. A posteriori, this is not entirely worrying: in fact, these two quantities prove to be the ones most sensible to the variation of the control parameter and they do not give consistent result as we move closer, but still above, the critical point. They show the expected behavior only in the limit $\lambda \gg 1$[1], which is not useful in the problems we are dealing with. We believe that there are other quantities that are much more valuable.

In order to understand which these quantities are, we look at the contact process a $\lambda \gtrapprox \lambda_c$ and, perhaps not surprisingly, we find that the best methods to distinguish this slightly supercritical system from the real critical point are the ones that deal with the correlation structure. Indeed, many interesting properties of critical systems are a consequence of the presence of strong correlations at all scales, hence it is somewhat natural to expect that tests that involve the correlations are going to be the most interesting ones.

We believe that the most fascinating approach is the one in momentum space, because it has both the clearest physical meaning and the clearest interpretation in terms of the Renormalization Group flow. Indeed, the standard approach to momentum space renormalization relies on the removal of short wavelength Fourier modes, that depend on the microscopic details of the system and are not important for the collective behavior that emerges at criticality. This translates quite well into the removal of low variance modes. Furthermore, as we have shown in Chapter 3 and Appendix A, we can think of a RG transformation as a transformation that acts on the probability distribution of the degrees of freedom of the system. Hence, fixed points of the Renormalization Group are fixed forms of such probability distribution and we can probe them using momentum space renormalization. The presence of the central limit theorem fixed point is extremely useful because it tells us that away from criticality the weakly correlated variables will converge there[2], and on the other hand if we see

[1] In the previous chapter the simulations of the supercritical contact process were made with $\lambda = 3$. If we drastically increase the value of λ eventually we find that both $\tilde{\alpha}$ and $\tilde{\beta}$ reach the expected value 1, with the caveat that the silence probability decreases even more and is vanishing even for smaller clusters.

[2] To be precise, this is true if no long-range order is present. In the case of the subcritical Ising model, for instance, the fixed point is trivial but it is not Gaussian: we find a superposition of delta functions because the coarse-graining drives the system towards a configuration with either all spins up or all spins down. Albeit this is not the CLT fixed point, it is still not significant. Of course this is a complication that does not show up in the case of absorbing phase transitions, which is the one we are most interested in.

non-Gaussian tales it is an interesting signatures of a non trivial correlation structure. However, apart from the CLT case we hardly know a priori this fixed form and it may not be trivial to determine whether there is convergence or not, especially for small systems. Since we do not see a clear convergence, the non Gaussian features we find are more of a necessary condition for criticality, but not a sufficient one since they might be just a signature of strong, but not critical, correlations.

Overall we believe that, as it is, this phenomenological approach should be considered as a good starting point rather than a complete method able to provide a definitive answer in complex systems. For instance, if we think about neural activity one might wonder if the brain really operates at criticality or is slightly supercritical, and the fact that the coarse-graining of the contact process at $\lambda \gtrsim \lambda_c$ produces mixed results is exactly the reason for which one should be careful and pursue different approaches as well. Of course, this is a delicate point and further research is needed to refine the analysis in order to quantify how distant a system is from being critical and to take into account finite size effects.

Nonetheless, the scope of a general phenomenological Renormalization Group is extremely broad: in its essence, it is a method that applies directly to the data whenever a complete model is too hard to build, and at least in principle it is able to extract the typical signatures of criticality if there are any. We could apply it to a vast variety of systems, from neural activity to flocks of birds, from the growth of a city to biological networks, and also to many physical models. The Renormalization Group teaches us that the critical behavior is controlled by the critical fixed point, and if we were able to determine such fixed point - possibly in the form of the convergence of the joint probability distribution of coarse-grained variables - we might be able to understand which are the important features of the system in the RG sense and infer an effective model that describes the collective, macroscopic behavior emerging at criticality. This would hugely augment our ability to test the fascinating hypothesis that, after all, all these systems might be described by the same language. And that language is the Statistical Mechanics of phase transitions.

APPENDIX A

THE RG, PROBABILITY AND CRITICALITY

In Chapter 3 we saw that for a sequence of centered and independent random variables and a coarse-graining step defined as

$$X_i^{(n)} = \frac{X_{2i-1}^{(n-1)} + X_{2i}^{(n-1)}}{\zeta} \tag{A.1}$$

for $\zeta = \sqrt{2}$ it exists a Gaussian fixed point in the space of probability distributions. However, we do not typically deal with independent variables: one might try to argue that away from criticality the correlation among them is weak and this result still holds, but we would like to understand what happens at the transition at least qualitatively.

Recall that we wrote the RG transformation for the joint probability distribution of the variables as

$$p_{n+1}(x) = (\mathcal{R}_\zeta p_n)(x) = \zeta \int dx'\, p_n(x') p_n(\zeta x - x').$$

Of course, this is not true for the case of dependent variables and there is not a general form we can write down for the transformation. Now we shall consider a particular case in which it is possible to do so, the one of a hierarchical model.

Consider a spin system in a one dimensional lattice, and the interaction scheme illustrated in Figure A.1. At the lowest level the spins interact in pairs; at the next level the interaction is between pairs; then between quadruplets; and so on. We have $N = 2^n$ variables. This model is extremely interesting because we can write a recursion relation for the Hamiltonian. If we consider the n-th level, we have

$$H_n(x_1, \ldots, x_{2^n}) = H_{n-1}(x_1, \ldots, x_{2^{n-1}}) + H_{n-1}(x_{2^{n-1}+1}, \ldots, x_{2^n}) - c^n \left(\sum_{i=1}^{2^n} \frac{x_i}{2^n} \right)^2$$

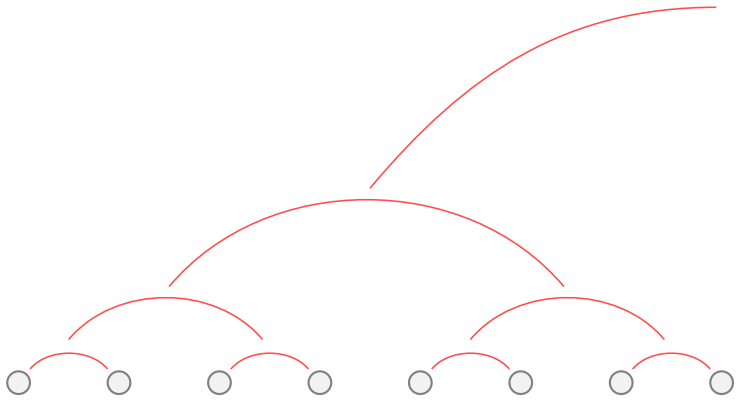


FIGURE A.1: A representation of the interactions in a hierarchical model for the variables $(x_1, \ldots, x_8)$. Notice how the variables are already grouped in pairs, as they would be in the consequent steps of a coarse-graining process.

where the last term is our choice of the interaction and $1 < c < 2$ [7]. Notice that the first terms are

$$H_1(x_1, x_2) = H_0(x_1) + H_0(x_2) - x\left(\frac{x_1 + x_2}{2}\right)^2$$

$$H_2(x_1, x_2, x_3, x_4) = H_1(x_1, x_2) + H_1(x_3, x_4) - x\left(\frac{(x_1 + x_2)/2 + (x_3 + x_4)/2}{2}\right)^2$$

and so on. This kind of hierarchical structure is extremely useful if we deal with a coarse-graining transformation such as (A.1), since the spins interact in pairs in the first place. Hence, each level n of the interaction corresponds to the n-th step of the coarse graining, short of the normalization.

Suppose we know the distribution of the single variables $p_0(x) = \exp[-\beta H_0(x)]$. Then at the first level we can write

$$\begin{aligned} p_1(x) &= \int dx_1 dx_2\, p_1(x_1, x_2)\delta\left(x - \frac{x_1 + x_2}{\zeta}\right) \\ &= \int dx_1 dx_2\, e^{-\beta H_0(x_1)} e^{-\beta H_0(x_2)} e^{\beta c\left(\frac{x_1+x_2}{2}\right)^2} \delta\left(x - \frac{x_1 + x_2}{\zeta}\right) \\ &= \zeta \exp\left[\beta c\left(\frac{\zeta x}{2}\right)^2\right] \int dx_1\, p_0(x_1) p_0(\zeta x - x_1). \end{aligned}$$

Similarly, at the second level

$$p_2(x) = \int dx_1 \ldots dx_4\, e^{-\beta(H_1(x_1,x_2)+H_1(x_3,x_4))} e^{\beta c^2\left(\frac{x_1+\cdots+x_4}{4}\right)^2} \delta\left(x - \frac{x_1 + \cdots + x_4}{\zeta^2}\right)$$

and in order to simplify this expression we introduce the identity

$$1 = \int dy_1 dy_2\, \delta\left(y_1 - \frac{x_1 + x_2}{\zeta}\right) \delta\left(y_1 - \frac{x_3 + x_4}{\zeta}\right) \tag{A.2}$$

where $(x_1 + x_2)/\zeta$ and $(x_1 + x_2)/\zeta$ are the coarse grained variables from the previous level. Doing so yields

$$p_2(x) = \zeta \exp\left[\beta c^2 \left(\frac{\zeta^2 x}{2^2}\right)^2\right] \int dy_1\, p_1(y_1) p_1(\zeta x - y_1).$$

Recursively inserting in the n-th level expression the identity (A.2) at the $n-1$-th level leads to the general expression

$$\begin{aligned} p_n(x) &= L_n \exp\left[\beta c^n \left(\frac{\zeta^n x}{2^n}\right)^2\right] \int dy\, p_{n-1}(x) p_{n-1}(\zeta x - y) \\ &:= g_n(x^2)(\mathcal{R}_\zeta p_n)(x) \end{aligned} \tag{A.3}$$

where L_n is a normalization factor that assures that the transformation stays in the space of probability distributions. The expression (A.3) is the exact transformation of the Renormalization Group for this hierarchical model.

Now suppose that we try to use the same normalization we used to find the CLT fixed point, namely $\zeta = \sqrt{2}$. Then

$$g_n(x^2) = L_n \exp\left[\beta \left(\frac{c}{2}\right)^n x^2\right].$$

For the sake of argument, let us further assume that the single variable distribution is Gaussian,

$$p_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

In this case we are able to perform the iteration exactly [27], since a Gaussian integration yields

$$p_n(x) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-x^2/2\sigma_n^2}$$

with

$$\sigma_n^2 = \left[1 - 2\beta \sum_{k=1}^{n} \left(\frac{c}{2}\right)^k\right]^{-1} \tag{A.4}$$

which is well defined only if $\beta < \beta_{\text{cr}} = 1/c - 1/2$, that is only if we are in the high temperature phase. Then the $n \to \infty$ limit gives a Gaussian fixed point with variance

$$\sigma^2 = \frac{2 - c}{2 - c(1 + 2\beta)}.$$

In particular, at $\beta = \beta_{\text{cr}}$ the variance becomes infinite and our choice of the normalization ζ does not make sense anymore. This somewhat expected since the fluctuations at criticality are not Gaussian, meaning that they increase faster than $\mathcal{O}(2^{n/2})$.

Let us consider the particular case $\zeta = 2c^{-1/2}$, so that

$$g_n(x^2) = g(x^2) = Le^{\beta x^2}.$$

A Gaussian distribution is still a fixed point of (A.3), namely

$$p_G(x) = \sqrt{\frac{a_0}{\pi}} e^{-a_0 x^2}, \quad a_0 = \frac{2\beta}{2-c} \tag{A.5}$$

but one should be careful that this is not the CLT fixed point because the fluctuations are not Gaussian, and indeed the normalization is not the one required by the CLT fixed point, that exists only for $\beta > \beta_{\text{cr}}$.

In fact, (A.5) is not always stable, and it is not even the only fixed point we can find [9]. Let us give an heuristic argument and look for a solution $f(x) = p_G(x)\phi(x)$, which apart from the normalization implies

$$\begin{aligned}
\phi(x) &= Le^{(\beta+a_0)x^2} \int dy\, e^{a_0\left[y^2+(\zeta x-y)^2\right]} \phi(y)\phi(\zeta x - y) \\
&= Le^{(\beta+a_0)x^2} \int dy\, e^{-2a_0 z^2 - \frac{1}{2}a_0\zeta^2 x^2} \phi\left(\frac{1}{2}\zeta x + z\right) \phi\left(\frac{1}{2}\zeta x - z\right) \\
&= L \int dy\, e^{-2a_0 z^2} \phi\left(\frac{1}{2}\zeta x + z\right) \phi\left(\frac{1}{2}\zeta x - z\right)
\end{aligned}$$

where we change the variable to $z = y - \zeta x/2$. Suppose that the perturbation around the Gaussian solution is small, so we choose

$$\phi(x) = 1 + \eta h(x), \quad \eta \ll 1$$

that gives the linear operator

$$(\mathcal{L}h)(x) = \int dz e^{-2a_0 z^2} \left[h\left(\frac{1}{2}\zeta x + z\right) + h\left(\frac{1}{2}\zeta x - z\right)\right].$$

It is possible to show [7] that for $\zeta = 2c^{-1/2}$ the eigenfunctions of this operator are properly rescaled Hermite polynomials of even degree, with eigenvalues

$$\lambda_{2k} = \frac{2}{c^k}, \quad k = 0, 1, 2, \dots.$$

Hence, sufficiently close to the Gaussian solution we may write

$$f(x) = p_G(x)\left[a_0(\beta) + a_2(\beta)H_2(x) + a_4(\beta)H_4(x) + \dots\right].$$

Notably, for $\sqrt{2} < c < 2$ the only unstable direction is given by H_2, since all the other eigenvalues are such that $\lambda < 1$[1]. If $a_2(\beta_{\text{cr}}) = 0$ the Gaussian fixed point (A.5) is stable,

[1] We ignore H_0, since it is a constant and can always be absorbed into the normalization [9].

that is the joint probability distribution of the coarse-grained variables at criticality is still Gaussian even if the normalization is not $\mathcal{O}(N)$ anymore.

However, if $1 < c < \sqrt{2}$ the direction of H_4 becomes unstable as well and the linear analysis tells us that (A.5) cannot be stable. The stable fixed point becomes non-trivial, and to find it one typically needs to rely on perturbation theory. This in no different from the usual exchange of stability between the mean field fixed point and the fixed point for models embedded in $d < d_u$, where d_u is the upper critical dimension.

Let us summarize what this heuristic argument suggests. For $\beta > \beta_{\text{cr}}$, the limit distribution of the coarse-grained variables corresponds the central limit theorem fixed point, that is fluctuations are $\mathcal{O}(\sqrt{N})$. This is the result we see using the PCA-like coarse-graining in the case of the supercritical contact process, for instance, because the initial variables are weakly correlated.

If $\beta = \beta_{\text{cr}}$, the fixed point is not the CLT one anymore. We have two possible solutions: a Gaussian distribution with fluctuations that increase faster than $\sqrt{N}$, which is the case of mean-field theories, and a non-trivial distribution that describes systems embedded in $d < d_{\text{u}}$. Since we we are looking at the contact process below its upper critical dimension, with our approach we expect to see the convergence to this non-Gaussian limit distribution.

APPENDIX B

PRINCIPAL COMPONENT ANALYSIS

Let us denote by X_{ij} the $P \times N$ matrix of P points each with N features, so each row represent a single configuration $\mathbf{x}_i$ that lives in a N-dimensional space. If we refer to the considerations we made in Chapter 3, each row represent a configuration of a square lattice with N sites, for instance, and we have P samples. Each sample represent a point whose coordinates are the N sites. We want to tackle the problem of dimensionality reduction in a slightly different way than what we did, and in particular we want to understand the relation between PCA and the RG procedure in momentum space we introduced.

Singular value decomposition tells us that every $P \times N$ matrix can be written as

$$X = U\Sigma V^T, \quad UU^T = \mathbb{I} = VV^T, \quad \Sigma = \text{diag}(\sigma_1, \dots, \sigma_r) \tag{B.1}$$

where $U \in \mathbb{M}(P \times P)$, $V \in \mathbb{M}(N \times N)$, $\Sigma \in \mathbb{M}(P \times N)$ and $r = \min(P, N)$. For instance, if $r = N$ we have

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_N \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

The values $(\sigma_1, \dots, \sigma_r)$ are called singular values of the matrix X, whereas the columns of U and V are respectively called the left-singular vectors and the right-singular

vectors of X. In fact, we shall notice that

$$\begin{cases} X^T X = V\,\Sigma^T\Sigma\,V^T, & \Sigma^T\Sigma \in \mathbb{M}(N\times N) \\ XX^T = U\,\Sigma\Sigma^T\,U^T, & \Sigma\Sigma^T \in \mathbb{M}(P\times P) \end{cases}$$

which means that the right-singular vectors of X are eigenvectors of X^TX, and the left-singular vectors are eigenvectors for XX^T. This further implies that the non-zero singular values of X are the square roots of the non-zero eigenvalues of both XX^T and X^TX.

With this in mind, a possible approach to dimensionality reduction is to find the best low-rank decomposition of X, where we should note that the rank is equal to the number of non-zero singular values. We want

$$\min||X - X_k||_F^2 = \min\left[\sum_{ij}(X - X_k)_{ij}^2\right]$$

where $||\cdot||_F$ is the Frobenius norm and X_k is a rank k matrix. Let us look for a unitary rank decomposition

$$X_1 = \lambda\mathbf{a}\mathbf{b}^T, \qquad \mathbf{a}^T\mathbf{a} = 1 = \mathbf{b}^T\mathbf{b} \tag{B.2}$$

with $\mathbf{a}\in\mathbb{R}^P$ and $\mathbf{b}\in\mathbb{R}^N$, so that X_k is a $P\times N$ matrix. Hence

$$\partial_{b_\mu}\left[\sum_{ij}(X_{ij} - \lambda a_i b_j)^2\right] = 0 = a_i\sum_i(X_{i\mu} - \lambda a_i b_\mu)$$

and, upon deriving with respect to the components of a as well, we end up with

$$\begin{cases} \mathbf{a}^T X = \lambda\mathbf{b}^T \\ X\mathbf{b} = \lambda\mathbf{a} \end{cases} \implies \begin{cases} XX^T\mathbf{a} = \lambda^2\mathbf{a} \\ X^TX\mathbf{b} = \lambda^2\mathbf{b} \end{cases}.$$

This result means that Equation (B.2) is the best rank 1 decomposition of X if $\mathbf{a}$ is an eigenvector of XX^T and $\mathbf{b}$ is an eigenvector of X^TX, both with the same eigenvalue λ^2. Hence we can write

$$X_1 = \sigma_i\mathbf{u}_i\mathbf{v}_i^T$$

for some $i = 1,\dots,\text{rank}(X)$. The only thing we need to do is to choose the appropriate index i among them.

Since $X = \sum_j \sigma_j\mathbf{u}_j\mathbf{v}_j^T$ and the Frobenius norm is defined as $||X||_F^2 = \sum_i\sigma_i^2$, it is clear that the quantity $||X - X_1||_F^2$ is minimized if we choose the index i to be the one of the highest singular value. In fact,

$$||X - X_1||_F^2 = ||X - \sigma_i\mathbf{u}_i\mathbf{v}_i^T||_F^2 = \sum_{j=1,j\neq i}^{\text{rank}(X)}\sigma_j^2$$

so the best choice we can make is to remove form this sum the highest singular value. This is known as the Young-Eckhart theorem and it is true for any value k: if $i = 1, \dots, k$ are the indexes of the highest singular values and we define

$$X_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

then $||X - X_k||_F < ||X - B||_F$ for every matrix $B \in \mathbb{M}(P \times N)$ of rank k. Hence, the low-rank decomposition is intimately related to the k highest singular values.

Insofar, we said nothing about dimensionality reduction, albeit it is clear that finding a rank $k < N$ decomposition amounts to restrict the data into a k-dimensional subspace of the original N-dimensional space[1]. Thus in light of the Young-Ekchart theorem instead of finding a rank k decomposition we could directly project the variables into the subspace spanned by the k highest right-singular vector of X. This is what PCA does when it is used to reduce the dimensionality of the data.

Suppose that $P > N$, that is we have more points than dimensions. Then the rank of X is at most N, which a trivial statement in the sense that it equivalent to tell that the points live in at most N dimensions. Reducing the dimensionality amounts to find a matrix $X_{P,k} \in \mathbb{M}(P \times k)$ which represents P points in k dimensions. To do so, we introduce the $N \times k$ matrix

$$W = (\mathbf{v}_1, \dots, \mathbf{v}_k)$$

where $\mathbf{v}_i$, being right-singular vectors, are the eigenvectors of $X^T X$. For the sake of argument, suppose that the data are centered: then $X^T X$ is exactly the covariance matrix, and W is the matrix of the k eigenvectors of the highest eigenvalues of the covariance matrix. Thus the low dimensional representation is

$$X_{P,k} = XW \in \mathbb{M}(P \times k).$$

W is called the matrix of the principal directions, and $X_{P,k}$ gives the first k principal components. For instance, an usual choice is $k = 2$ so that we can plot the first two principal components against each other. Notice that by means of the Young-Eckhart theorem this would the best two-dimensional representation of the data, but nothing tells us a priori that the data really live in two dimensions. We need to look at the spectrum of $X^T X$ to understand whether is makes sense or not.

We briefly note that if $N > P$ the rank of the matrix is at most P. Hence, looking for a $k \times N$ representation of the data makes sense. This is achieved by means of the matrix

$$Z = (\mathbf{u}_1, \dots, \mathbf{u}_k)$$

which is $P \times k$. Then

$$X_{k,N} = Z^T X \in \mathbb{M}(k \times N)$$

[1] This is obvious since the rank of a matrix is defined as the maximum number of linearly independent columns or rows, hence the dimensions such vectors are spanning.

is of rank at most k. Notice that $\mathbf{u}_i$ are the eigenvectors of XX^T, which is the Gram matrix and not the covariance matrix. In Chapter 3 we had a physical motivation to use the covariance matrix in the first place, namely that it is directly related with the correlation functions, so we should be careful to always consider a number of samples that is greater than the number of lattice sites.

Finally, let us remark that what we do in our Renormalization Group procedure in momentum space, where we vary a cutoff Λ on the number of eigenvectors we project onto, is essentially equivalent to determine the best low-rank decomposition rather than dimensionality reduction. In fact, in the notation used for the singular value decomposition the projector we used in Chapter 3 is given by

$$(P_k)_{\mu\nu} = \sum_{i=1}^{k} v_{\mu i} v_{\nu i}.$$

But then we recover the result of the Young-Eckhart theorem by means of

$$X_k = XP_k = \sum_{j=1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^T \sum_{i=1}^{k} \mathbf{v}_i \mathbf{v}_i^T = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T .$$

Hence, if we define the $P \times N$ matrix Φ as the matrix of the coarse-grained variables we have

$$\Phi_{\mu\nu} = \sum_{\sigma=1}^{N} X_{\mu\sigma} (P_k)_{\sigma\nu}$$

which is exactly X_k.

APPENDIX C

SAMPLING THE CONTACT PROCESS

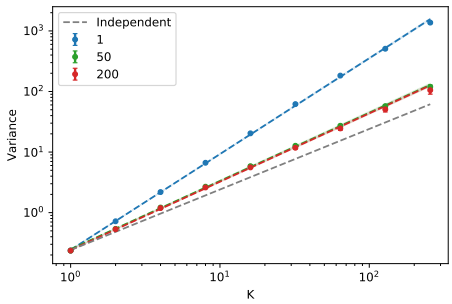
As we stressed out in Chapter 3, sampling the dynamical evolution of the contact process is not trivial due to the interplay between the physical correlation and the spurious one induced by our choice of the algorithm. Moreover, performing long time simulations at criticality is computationally heavy, so we would like to establish a tradeoff among the sampling rate r and the results of the coarse-graining procedure.

Let us look at the supercritical contact process in detail, since in this regime the variables are weakly correlated and therefore we expect that at low r the major contribution to the correlation is the spurious one. In Figure C.1 we see that for $r = 1$ almost all the scaling features are very similar to the ones of a critical system. This result is twofold. On one hand, it tells us that there is indeed a spurious correlation in the simulation, hence $r = 1$ is a poor choice. On the other, it suggests carefulness because we see scaling features even if we have no reason to believe that such spurious correlation is critical in any way.

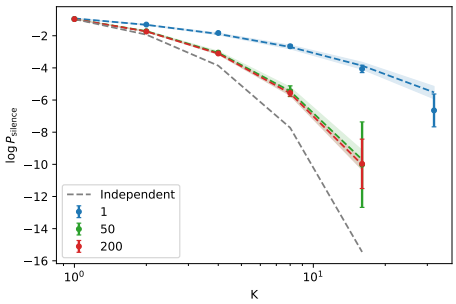
For $r = 50$ and $r = 100$ it is hard to tell the results apart, and the exponents are fully compatible. In particular, we see that both the exponents if the variance and the silence probability are closer to, but not quite at, the independence values $\tilde{\alpha} = \tilde{\beta} = 1$. Hence the considerations that we made in Chapter 3 hold, namely that both the variance and the free energy are not entirely able to clearly distinguish between criticality and supercriticality[1]. The convergence to the CLT fixed point in Figure C.1f is much faster for both these sampling rates and is quite different from what we find at criticality.

Overall we believe that with good approximation only the physical correlation remains for $r > 50$. However, as we said for the supercritical contact process it is not computationally expensive for go as high as $r = N$ because it is highly improbable for

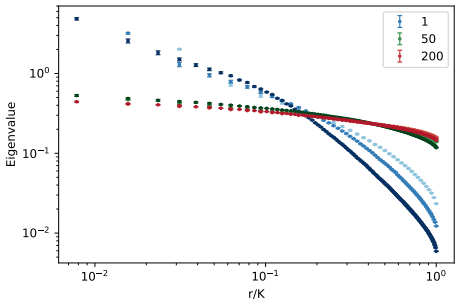
[1] As noted in the last chapter, the exponents become compatible with 1 only if $\lambda \gg \lambda_c$.



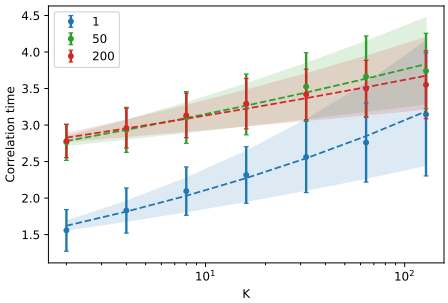
(A) Scaling of the eigenvalues of the variance.



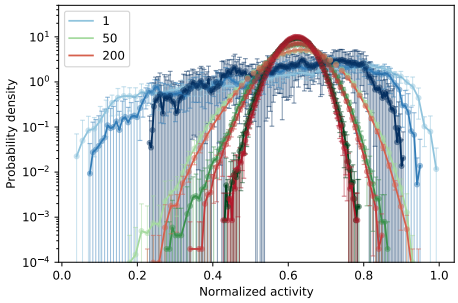
(B) Scaling of the logarithm of the silence probability.



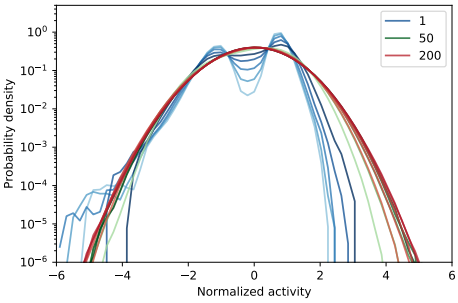
(C) Scaling of the eigenvalues of the covariance matrix.



(D) Scaling of the correlation times.

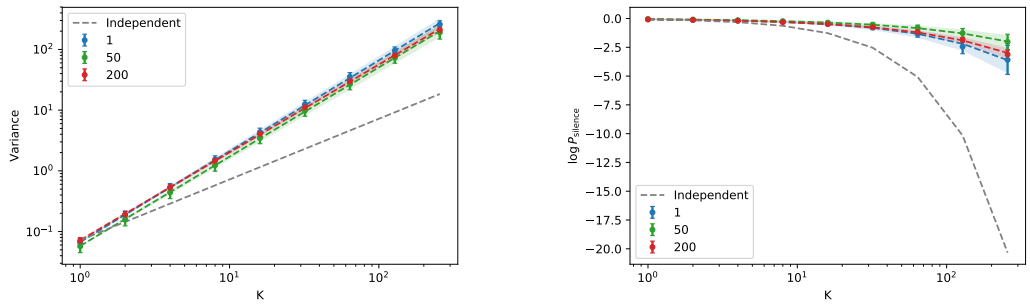


(E) Distribution of normalized non-zero activity.



(F) Probability distribution of the projected variables.

FIGURE C.1: Comparison of the results of the main coarse-graining procedures for the supercritical contact process for different sampling rates $r = 1, 50, 200$. In each figures, the data in blue represents $r = 1$, the data in green $r = 50$ and the data in red $r = 200$.



(A) Scaling of the eigenvalues of the variance.

(B) Scaling of the logarithm of the silence probability.

FIGURE C.2: Comparison of some coarse-graining exponents in the critical contact process for different sampling rates $r = 1, 50, 200$. The data in blue represents $r = 1$, the data in green $r = 50$ and the data in red $r = 200$.

the system to be anywhere near the absorbing state. The results for $r = N$ in Chapter 3 are compatible with the ones of $r = 50$ and $r = 200$ here, as expected.

Since the algorithm does not change, one might argue that taking $r = 200$ is enough to loose the spurious correlation even in the critical regime. However, at criticality the physical correlation is much stronger and what we find is that even for $r = 1$ the spurious correlation has a marginal effect. In fact, the exponents remain compatible from $r = 1$ to $r = 200$ as we can see in Figure C.2. In Chapter 3 we use a simulation with $r = 200$ since it is still a good tradeoff between the stability of the results and the computational time.

BIBLIOGRAPHY

[1] P. Bak. *How Nature Works: the science of self-organized criticality*. Copernicus, New York, 1996.

[2] P. Bak, C. Tang, and K. Wiesenfeld. "Self-organized criticality: An explanation of the 1/f noise". In: *Phys. Rev. Lett.* 59 (4 1987).

[3] J. M. Beggs and D. Plenz. "Neuronal Avalanches in Neocortical Circuits". In: *Journal of Neuroscience* 23.35 (2003).

[4] W. Bialek and T. Mora. "Are Biological Systems Poised at Criticality?" In: *Journal of Statistical Physics* 144.2 (2011).

[5] W. Bialek et al. "Social interactions dominate speed control in poising natural flocks near criticality". In: *Proceedings of the National Academy of Sciences* 111.20 (2014).

[6] J. J. Binney et al. *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*. Oxford University Press, 1992.

[7] P. M. Bleher and J. G. Sinai. "Investigation of the critical point in models of the type of Dyson's hierarchical models". In: *Communications in Mathematical Physics.* 33.1 (1973).

[8] S. Bradde and W. Bialek. "PCA Meets RG". In: *Journal of Statistical Physics* 167.3 (2017).

[9] M. Cassandro and G. Jona-Lasinio. "Critical point behaviour and probability theory". In: *Advances in Physics* 27.6 (1978).

[10] A. Cavagna et al. "Scale-free correlations in starling flocks". In: *Proceedings of the National Academy of Sciences* 107.26 (2010).

[11] W. Cota and S. C. Ferreira. "Optimized Gillespie algorithms for the simulation of Markovian epidemic processes on large and heterogeneous networks". In: *Computer Physics Communications* 219 (2017).

[12] R. Dickman. "Reweighting in nonequilibrium simulations". In: *Physical review. E* 60 (Oct. 1999).

[13] R. Dickman and M. M. de Oliveira. "Quasi-stationary simulation of the contact process". In: *Physica A: Statistical Mechanics and its Applications* 357.1 (2005).

[14] A. E. Ferdinand and M. E. Fisher. "Bounded and Inhomogeneous Ising Models. I. Specific-Heat Anomaly of a Finite Lattice". In: *Phys. Rev.* 185 (2 1969).

[15] R. Ferreira and S. Ferreira. "Critical behavior of the contact process on small-world networks". In: *European Physical Journal B* 86 (July 2013).

[16] C. Furusawa and K. Kaneko. "Adaptation to Optimal Cell Growth through Self-Organized Criticality". In: *Phys. Rev. Lett.* 108 (20 2012).

[17] C. Furusawa and K. Kaneko. "Zipf's Law in Gene Expression". In: *Phys. Rev. Lett.* 90 (8 2003).

[18] C. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. Springer, 2004.

[19] D. T. Gillespie. "Stochastic Simulation of Chemical Kinetics". In: *Annual review of physical chemistry* 58 (2007).

[20] N. Goldenfeld. *Lectures on phase transitions and the renormalization group*. Addison-Wesley, 1992.

[21] T. E. Harris. "Contact Interactions on a Lattice". In: *Ann. Probab.* 2.6 (1974).

[22] M. Henkel, H. Hinrichsen, and S. Lübeck. *Non-Equilibrium Phase Transitions. Volume 1: Absorbing Phase Transitions*. Springer, 2009.

[23] J. Hidalgo et al. "Information-based fitness and the emergence of criticality in living systems". In: *Proceedings of the National Academy of Sciences* 111.28 (2014).

[24] P. Hohenberg and A. Krekhov. "An introduction to the Ginzburg–Landau theory of phase transitions and nonequilibrium patterns". In: *Physics Reports* 572 (2015).

[25] K. Huang. *Statistical Mechanics*. Wiley, 1987.

[26] A. Jiménez-Dalmaroni. "Directed percolation with incubation times". In: *Phys. Rev. E* 74 (1 2006).

[27] G. Jona-Lasinio. "Renormalization group and probability theory". In: *Physics Reports* 352.4 (2001).

[28] K. B. Lauritsen et al. "Directed percolation with an absorbing boundary". In: *Physica A: Statistical Mechanics and its Applications* 247.1 (1997).

[29] A. Lesne and M. Laguës. *Scale Invariance: From Phase Transitions to Turbulence*. Springer, 2012.

[30] R. Livi and P. Politi. *Nonequilibrium Statistical Physics: A Modern Perspective*. Cambridge University Press, 2017.

[31] J. Marro and R. Dickman. *Nonequilibrium Phase Transitions in Lattice Models*. Cambridge University Press, 1999.

[32] M. Marsili and Y.-C. Zhang. "Interacting Individuals Leading to Zipf's Law". In: *Physical Review Letters* 80 (1998).

[33] M. Martinello et al. "Neutral Theory and Scale-Free Neural Dynamics". In: *Phys. Rev. X* 7 (4 2017).

[34] L. Meshulam et al. “Coarse-graining and hints of scaling in a population of 1000+ neurons”. In: *arXiv preprints* (2018). eprint: 1812.11904 (physics.bio-ph).

[35] T. Mora et al. “Maximum entropy models for antibody diversity.” In: *Proceedings of the National Academy of Sciences* 107.12 (2010).

[36] M. A. Muñoz. “Colloquium: Criticality and dynamical scaling in living systems”. In: *Rev. Mod. Phys.* 90 (3 2018).

[37] M. A. Muñoz et al. “Avalanche and spreading exponents in systems with absorbing states”. In: *Phys. Rev. E* 59 (5 1999).

[38] L. Onsager. “Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition”. In: *Phys. Rev.* 65 (3-4 1944).

[39] D. Schwab, I. Nemenman, and P. Mehta. “Zipf’s Law and Criticality in Multivariate Data without Fine-Tuning”. In: *Physical review letters* 113 (2013).

[40] J. P. Sethna. *Entropy, Order Parameters, and Complexity*. University Press, 2006.

[41] G. Tkacik et al. “Thermodynamics for a network of neurons: Signatures of criticality”. In: *Proceedings of the National Academy of Sciences* 112 (2014).

[42] U. C. Täuber. *Critical Dynamics: A Field Theory Approach to Equilibrium and Non-Equilibrium Scaling Behavior*. Cambridge University Press, 2014.

[43] S. Zapperi, K. B. Lauritsen, and H. E. Stanley. “Self-Organized Branching Processes: Mean-Field Theory for Avalanches”. In: *Phys. Rev. Lett.* 75 (22 1995).

[44] G. K. Zipf. *Human behavior and the principle of least effort*. Addison-Wesley, Cambridge, 1949.