



Dynamic cavity method and problems on graphs

Andrey Y. Lokhov

► To cite this version:

Andrey Y. Lokhov. Dynamic cavity method and problems on graphs. Disordered Systems and Neural Networks [cond-mat.dis-nn]. Université Paris Sud - Paris XI, 2014. English. NNT : 2014PA112331 . tel-01127103

HAL Id: tel-01127103

<https://theses.hal.science/tel-01127103>

Submitted on 7 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE 564: PHYSIQUE EN ÎLE-DE-FRANCE
LABORATOIRE DE PHYSIQUE THÉORIQUE ET MODÈLES STATISTIQUES

DISCIPLINE : PHYSIQUE

THÈSE DE DOCTORAT

Soutenue le 14 Novembre 2014 par

Andrey Y. Lokhov

**Dynamic cavity method and
problems on graphs**

Devant le jury composé de :

Satya Majumdar
Marc Mézard
David Saad
Guilhem Semerjian
Lenka Zdeborová
Riccardo Zecchina

Président du jury
Directeur de thèse
Rapporteur
Examineur
Membre invité
Rapporteur

Acknowledgments

I am grateful to my supervisor, Marc Mézard, for giving me the opportunity to spend three wonderful years of PhD program in Laboratoire de Physique Théorique et Modèles Statistiques, an exceptional place for scientific research. During these years, Marc was for me an example of a researcher of extremely high quality, with outstanding intuition and technical skills. It is my pleasure to thank all my co-authors and collaborators, who taught me how to translate the work, patience and enthusiasm into a valuable research. With a special feeling I am grateful to Sergei Nechaev for his kind support and for sharing with me his passion for physics. I am thankful to Sergei for helping me to explore new fields of research, and for showing me how to turn imagination and erudition into amazing discoveries. I am grateful to Lenka Zdeborová for providing me with valuable advice and insight throughout these years. I benefited considerably from the schools and conferences that Lenka organized, they have been a major source of knowledge and interactions with top researches in the field. It was a pleasure to work with Misha Tamm, Olga Valba and Hiroki Ohta. I learnt a lot from them, spending many hours in front of the blackboard and discussing interesting problems.

Step back in the past, I am indebted to all teachers and professors from lycée №130, Novosibirsk State University, Ecole Polytechnique and Ecole Normale Supérieure that influenced my choice of becoming physicist. In particular, I am thankful to Gregory Korchemsky for his time and patience. The experience that I gained during my Master Diploma turned into a deciding argument for continuing to work in theoretical physics.

I would like to thank the members of my defense committee for carefully examining my thesis, and for their comments.

This fascinating experience would not have been so colorful without a great atmosphere of interactions and discussions with my colleagues at LPTMS. I would like to thank all my friends in the lab, in particular Andrii, Arthur, Clélia, François, Dima, Matthieu, Misha, Pierre-Élie, Yasar, as well as many others, who were always ready to share their experience and advice at difficult moments of the PhD student's life. The encouragements of my close friends outside the lab have been a substantial source of motivation for my work. I would like to thank all of them, including university peers and the judo band; I am thinking in particular about Andrei, Jean-Nöel, Masha, Oleg, Sasha, Silèvre, Thomas, Yarik, Yulia, and others. Above all, I am grateful to Margarita for her constant love, encouragements and support.

Finally, I dedicate this thesis to my family: without their love and help I would have never made it to the end. I am very grateful to my brothers Alexey and Roman, for their continuous guidance, care and support, and to my father, for inculcating in me a taste for science and for rising my scientific curiosity. I am particularly indebted to my mother for her immense investment in my education, for her kindness and love throughout all these years.

Andrey Lokhov

Paris, November 14, 2014

Contents

Acknowledgments	i
Abstract	v
Résumé	vii
1 Foreword: statistical physics of complex systems	1
I Dynamic processes on networks	3
2 Introduction	5
2.1 The static spin-glass systems in a nutshell	5
2.1.1 Infinite-range systems	6
2.1.2 Finite connectivity systems	7
2.1.3 Algorithmic applications	7
2.2 Dynamics of the Ising systems	10
2.2.1 Langevin-type dynamics	10
2.2.2 Glauber dynamics	11
2.3 Spreading dynamics in complex systems	13
2.3.1 Some aspects of complex networks	14
2.3.2 Spreading processes	15
2.4 Dynamic belief propagation	20
3 Dynamic processes with unidirectional dynamics	25
3.1 Unidirectional dynamics: preliminary considerations	25
3.2 Models with 2 states	27
3.2.1 SI model	28
3.2.2 Random field Ising model	32
3.3 Models with 3 states	33
3.3.1 SIR model	33
3.3.2 Rumor spreading model	36
3.4 Models with larger number of states	43
4 Application: patient zero problem	45
4.1 Context of the inference problem	45
4.2 DMP algorithm	46
4.3 Results and discussion	47

5	Perspectives	53
	Key results	53
	Future directions	53
II	Planar matching and optimal folding	57
6	Introduction	59
6.1	Disordered planar matching: definitions	59
6.2	Statistical mechanics of RNA secondary structures	61
6.3	Topological transition: numerical	63
6.4	Naive mean-field and correlations	65
7	Topological transition: analytical	69
7.1	Field-theory approach	69
7.2	Combinatorial approach	73
7.2.1	First order of the expansion	74
7.2.2	Beyond the first order	77
8	Application: optimal folding	81
8.1	Connection to the molten-glass transition	81
8.2	Other models for non-integer alphabets	85
8.2.1	Construction of the non-integer alphabets	85
8.2.2	Perfect matching transition for non-integer alphabets	87
9	Perspectives	89
	Key results	89
	Future directions	89
	List of publications	95
	Bibliography	97

Title: Dynamic cavity method and problems on graphs

Author: Andrey Y. Lokhov

Abstract:

A large number of optimization, inverse, combinatorial and out-of-equilibrium problems, arising in the statistical physics of complex systems, allow for a convenient representation in terms of disordered interacting variables defined on a certain network. Although a universal recipe for dealing with these problems does not exist, the recent years have seen a serious progress in understanding and quantifying an important number of hard problems on graphs. A particular role has been played by the concepts borrowed from the physics of spin glasses and field theory, that appeared to be extremely successful in the description of the statistical properties of complex systems and in the development of efficient algorithms for concrete problems.

In the first part of the thesis, we study the out-of-equilibrium spreading problems on networks. Using dynamic cavity method on time trajectories, we show how to derive dynamic message-passing equations for a large class of models with unidirectional dynamics – the key property that makes the problem solvable. These equations are asymptotically exact for locally tree-like graphs and generally provide a good approximation for real-world networks. We illustrate the approach by applying the dynamic message-passing equations for susceptible-infected-recovered model to the inverse problem of inference of epidemic origin.

In the second part of the manuscript, we address the optimization problem of finding optimal planar matching configurations on a line. Making use of field-theory techniques and combinatorial arguments, we characterize a topological phase transition that occurs in the simple Bernoulli model of disordered matching. As an application to the physics of the RNA secondary structures, we discuss the relation of the perfect-imperfect matching transition to the known molten-glass transition at low temperatures, and suggest generalized models that incorporate a one-to-one correspondence between the contact matrix and the nucleotide sequence, thus giving sense to the notion of effective non-integer alphabets.

Keywords: cavity method, out-of-equilibrium dynamics, message-passing, belief propagation, unidirectional dynamics, spreading processes, constrained satisfaction problems, combinatorial optimization, planar matching, phase transitions

Titre: Méthode de cavité dynamique et problèmes sur des graphes

Autheur: Andrey Y. Lokhov

Résumé:

Un grand nombre de problèmes d'optimisation, ainsi que de problèmes inverses, combinatoires ou hors équilibre qui apparaissent en physique statistique des systèmes complexes, peuvent être représentés comme un ensemble de variables en interaction sur un certain réseau. Bien que la recette universelle pour traiter ces problèmes n'existe pas, la compréhension qualitative et quantitative des problèmes complexes sur des graphes a fait de grands progrès au cours de ces dernières années. Un rôle particulier a été joué par des concepts empruntés à la physique des verres de spin et la théorie des champs, qui ont eu beaucoup de succès en ce qui concerne la description des propriétés statistiques des systèmes complexes et le développement d'algorithmes efficaces pour des problèmes concrets.

En première partie de cette thèse, nous étudions des problèmes de diffusion sur des réseaux, avec la dynamique hors équilibre. En utilisant la méthode de cavité sur des trajectoires dans le temps, nous montrons comment dériver des équations dynamiques dites “message-passing” pour une large classe de modèles avec une dynamique unidirectionnelle – la propriété clef qui permet de résoudre le problème. Ces équations sont asymptotiquement exactes pour des graphes localement en arbre et représentent en général une bonne approximation pour des réseaux réels. Nous illustrons cette approche avec une application des équations dynamiques pour résoudre le problème inverse d'inférence de la source d'épidémie dans le modèle “susceptible-infected-recovered”.

Dans la seconde partie du manuscrit, nous considérons un problème d'optimisation d'appariement planaire optimal sur une ligne. En exploitant des techniques de la théorie de champs et des arguments combinatoires, nous caractérisons une transition de phase topologique qui se produit dans un modèle désordonné simple, le modèle de Bernoulli. Visant une application à la physique des structures secondaires de l'ARN, nous discutons la relation entre la transition d'appariement parfait-imparfait et la transition de basse température connue entre les états fondu et vitreux de biopolymère ; nous proposons également des modèles généralisés qui suggèrent une correspondance exacte entre la matrice des contacts et la séquence des nucléotides, permettant ainsi de donner un sens à la notion des alphabets effectifs non-entiers.

Mots-clés: méthode de cavité, dynamique hors équilibre, passage de messages, belief propagation, dynamique unidirectionnelle, processus de diffusion, problèmes de satisfaction des contraintes, optimisation combinatoire, appariement planaire, transitions de phase

Chapter 1

Foreword: statistical physics of complex systems

Modern statistical physics of complex systems can be generally thought of as a collection of concepts and techniques that make it possible to study in detail systems involving many interacting variables: particles, spins, colors, boolean variables that appear in combinatorial optimization and information theory, *etc.* Conceptually, the success of statistical physics lies in the surprising ability to explain an emergent complex behavior of the systems composed of many elementary constituents, from the underlying microscopic interactions between them. Although the manifestations of the systems' global behavior can seem to be very different, they can be often explained by collective effects of elementary components, common to many systems of different nature, and described by very general concepts, such as *phase transitions* – a qualitative change of the system's state.

It is often convenient to represent the topology of the interactions in a complex system by a *network* $G = (V, E)$, with a collection of nodes V , corresponding to variables, and edges E (directed or undirected), representing interactions between them (correspondingly directed or undirected). This interaction network can be extracted from the *real* system one wants to study, or can be artificially generated if one is interested in the general properties of the system: in this case, either the network is pre-designed to have some special structure, or a *random network* is used as an approximation to the real-world interaction graph. In this thesis, we will use both random networks (for physical modeling of the system and systematic study of its properties) and real-world networks (for practical applications and test of performance of the developed algorithms). Depending on the probability distribution which is used for generating the network, it may have some specific properties. In the part I we will mostly use *sparse* and *locally tree-like* random graphs, that are in general defined in the N -dimensional space, where N is the number of nodes of the network. On the contrary, in the part II we will mostly focus on *dense* and *planar* graphs, i.e. graphs that can be drawn without self-intersections on a plane, and hence are defined in a two-dimensional space.

The evolution of a complex system is generically accompanied by some microscopic dynamics. Still, depending on the questions one asks on the behavior or the properties of a complex system, the problem can be classified as *static* or *dynamic*. We will say that the problem is static in the case where one would like to find a particular solution to a related optimization problem, or studies the system at *equilibrium* (i.e. the underlying

microscopic dynamics respects the so-called *detailed balance condition* and hence the emergent global properties do not change). In this thesis, a particular attention will be devoted to the investigation of the *out-of-equilibrium* dynamics of complex systems. Depending on its nature, the system may eventually relax to the equilibrium state, fall down to some final absorbing state, or always stay out of equilibrium. Last but not least, both static and dynamic processes can be defined on static or dynamically-changing networks.

Throughout this thesis, we will see that the description of each problem involve concepts that are related to both statics and dynamics. For instance, in the part I, we will study the physics of the out-of-equilibrium dynamic processes on networks, such as, for instance, avalanches in magnetic systems, epidemic spreading, or propagation of opinions and information in social networks. The description of these processes in terms of time trajectories of individual variables will allow us to reformulate this problem in a purely static setting; however, the easiest way to obtain the solutions of the corresponding static equations consists in using an iterative message-passing algorithm, with its proper convergence dynamics. On the other hand, for many processes of interest, we will be able to write dynamic equations that instead have to be iterated in a physical time. Another example of mixing between the two concepts will be given by one of the problems studied in the part II, where we will solve a static constrained satisfaction problem of planar matching, relevant for the physics of secondary structures of RNA molecules. Again, a numerical algorithm that solves this problem, called dynamical programming algorithm, can be given an interpretation of a dynamic growth and rewiring of a certain graph. The planar matching problem will also have a one-to-one correspondence with a constrained random walk on a line and on a regular tree, which will be used for an analytic study of this optimization problem.

One of the main difficulties for treating the problems considered in this thesis essentially arises from their *disordered* nature. The disorder may be present both in the topology of the problem (in the heterogeneity of the interaction network) and in the parameters of the interactions. The two can be related if one is interested in the problems defined on dynamically-changing graphs: the dynamics can often be encoded into the change of interactions (e.g. a zero coupling would mean that the corresponding edge in the network is absent). If the dynamics of the network is much faster than the dynamics of the corresponding process, we would say that the disorder is *annealed*; in the opposite case, when the network can be considered as static, we will say that the disorder is *quenched*: typically, these problems are hard to solve. Finally, the disorder can be *relevant*, being at the source of a global properties of the system, or *irrelevant*, when it is not crucial. An example of this type will be provided in the part II, with a heterogeneous polymer behaving as a homopolymer at sufficiently high temperature.

All these concepts will be studied and explained in detail in the next chapters. The part I focuses mainly on the dynamic aspects of out-of-equilibrium processes on networks. In the part II, we consider a static optimization problem of planar matching, related to the problem of optimal folding in the physics of RNA. However, we will witness a frequent occurrence of the highlighted concepts in both parts.

Part I

Dynamic processes on networks

Chapter 2

Introduction

Over the past few decades, a serious progress has been made in the theory of disordered systems *at equilibrium*. The mean-field methods of statistical physics allow for the detailed study of the equilibrium problems involving a large number of interacting variables, both in the cases of dense and diluted interaction graphs. The developed mean-field techniques made it possible to introduce efficient algorithms for statistical inference and optimization, and to study the corresponding phase diagrams.

The recent years have seen a growing interest for building new analytical tools for the description of the *out-of-equilibrium* systems defined on heterogeneous networks, but there is still no well-established tractable method for solving the corresponding disordered dynamics in the general case. In this chapter, we will briefly review the existing statistical approaches to the out-of-equilibrium dynamics of classical disordered spin models and complex systems, and introduce a recently suggested approach to the problem, the dynamic cavity method on time trajectories.

2.1 The static spin-glass systems in a nutshell

A large part of theoretical developments in the statistical physics of disordered systems came from the efforts to describe the physics of spin glasses. A concrete physical example of a spin glass is given by an alloy of a noble non-magnetic material, crystallized on a regular lattice, with a small fraction of magnetic impurities, placed randomly inside the lattice. A particularity of the spin glass is that the sign of the interactions between atoms inside the material is an oscillating function of distance, leading to an unusual low temperature phase, in which the magnetizations of the impurities are fixed to some non-zero values, but the long range order is absent. The first model for these alloys has been suggested by Edwards and Anderson in 1975 [EA75]. This EA model is described by the Hamiltonian for N Ising spins $\sigma_i = \pm 1$:

$$H = - \sum_{\langle ij \rangle} J_{ij} \sigma_i \sigma_j, \quad (2.1)$$

where the interactions occur between the neighboring sites of a three-dimensional lattice, and the couplings J_{ij} take positive (ferromagnetic) or negative (anti-ferromagnetic) values. The exact solution of the EA model is a widely open question nowadays: even in the absence of disorder, the partition function of the purely ferromagnetic Ising model has been computed only at two dimensions. The EA model can be generalized to a family of

disordered spin models of the spin glass type, given by the Hamiltonian

$$H = - \sum_{\langle i_1 \dots i_p \rangle} J_{i_1 \dots i_p} \sigma_{i_1} \dots \sigma_{i_p}. \quad (2.2)$$

The definition of the model should be completed by the topological specification of the interaction graph G . The interacting variables σ_i , $i = 1, \dots, N$, take either discrete values $\sigma_i = \pm 1$ (Ising spins), or real values under some additional constraints that have to be included in the Hamiltonian (2.2). In the *soft-spin* Ising representation, the values σ_i are favored to take the values around ± 1 ; another popular choice is the *spherical model*, when the constraint $\sum_{i=1}^N \sigma_i^2 = N$ is imposed. The integer parameter $p \geq 2$ controls the number of spins involved in each interaction. The random couplings $J_{i_1 \dots i_p}$ are fixed (*quenched*) random numbers, drawn from some probability distribution; in the following, this distribution will be specified for each model.

2.1.1 Infinite-range systems

The models (2.2) can be solved in a special case, when the interaction graph is of a fully-connected type: the interactions run over all the interaction variables. Thus defined, the model is of a mean-field type: each variable interacts with a very large number of neighboring ones. In the case of a complete graph, one usually takes Gaussian or bimodal distribution of the couplings $J_{i_1 \dots i_p}$, with zero mean and variance, so that $\langle J_{i_1 \dots i_p}^2 \rangle_J = p! J^2 / (2N^{p-1})$ to ensure a correct thermodynamic limit. The statistical properties of the infinite-range Ising systems have been intensively investigated during several decades, for a review see [BY86, MVP87, FH93, DG06]. The case $p = 2$ of (2.2) for Ising spins corresponds to the so-called Sherrington-Kirkpatrick (SK) model [SK75, KS78] with infinite-range random interactions between Ising spins $\sigma_i = \pm 1$:

$$H = - \sum_{i < j} J_{ij} \sigma_i \sigma_j, \quad (2.3)$$

where the sum runs over all the ordered pairs of N spins. In the original papers, this problem has been attacked by the *replica method*, very often used in the mean-field disordered models: the introduction of n *replicas*, or copies, of the system, allows one to decouple the interaction between the spins; a peculiar limit $n \rightarrow 0$ is then used for computed quantities. Sherrington and Kirkpatrick have used the so-called *replica-symmetric* (RS) ansatz for the structure of the replicas space, but the obtained solution was not correct: the entropy at zero temperature appeared to be negative, and the solution was proved to be unstable [dAT78]. The correct solution of the SK model has been given by Parisi in a series of papers [Par80a, Par80c, Par80b] using the concept of the *replica symmetry breaking* (RSB), and has been rigorously proven by Talagrand [Tal06] using the interpolation method of Guerra [Gue03].

The spherical p -spin model has been introduced in [CS92], and can be studied and solved using the Thouless-Anderson-Palmer (TAP) method [TAP77, KPV93, CS95]. A particularly interesting case is given by the limit $p \rightarrow \infty$, when the energies of a pair of configurations are no longer correlated, and the system is frozen in a small number of configurations of small energy. This model, called the random energy model (REM), has been introduced by Derrida [Der81] as a simplification of the spin glass problem, in which the energies of configurations can instead be regarded as correlated random variables [BM97]. The limit $p \rightarrow \infty$ has been also studied by Gross and Mézard [GM84], while the final p effects have been investigated in [Gar85].

2.1.2 Finite connectivity systems

Obviously, the infinite-range systems differ a lot from their real three-dimensional counterparts: in the realistic models, the connectivity of the interaction graph remains finite (equal to 6 for a three-dimensional hyper-cubic lattice). The advantage of the *diluted models* of spin glasses is twofold: they provide for the finite connectivity even in the thermodynamic limit $N \rightarrow \infty$, but stay of the mean-field nature. Indeed, the neighbors for each node are chosen randomly between N sites of the system, and hence there is no notion of euclidean distance in the interaction graph. Therefore, one could hope that the properties of the diluted systems would be closer to the real spin glass systems.

In the finite connectivity systems, the interactions of (2.2) are drawn from the probability distribution [SCM04]

$$P(J_{i_1 \dots i_p}) = \left(1 - \frac{cp!}{N^{p-1}}\right) \delta(J_{i_1 \dots i_p}) + \frac{cp!}{N^{p-1}} \Pi(J_{i_1 \dots i_p}), \quad (2.4)$$

where δ is the Dirac distribution, Π is a regular distribution, and c is the average ratio of p -spin interactions per variable. The case $p = 2$ corresponds to the Viana-Bray model [VB85], with the couplings distribution of the form

$$P(J_{ij}) = \left(1 - \frac{c}{N}\right) \delta(J_{ij}) + \frac{c}{N} \pi(J_{ij}), \quad (2.5)$$

where, again, π is some regular distribution. However, the analytical treatment of the diluted models is more involved with respect to the fully-connected models: the order parameter in the systems with finite connectivity becomes a function and not a number. The order parameter at the RS level has been studied in the original work [VB85] and in [KS87] in the zero-temperature limit. The RSB phenomenon in weak connectivity systems has been first studied in [DM87]. A breakthrough in the understanding of the disordered diluted systems has been done by Mézard and Parisi [MP01]. These authors have used the *cavity method*, which is equivalent to the use of replicas, but is more intuitive and gives simpler equations for the diluted systems. They showed how to treat the problem at the first step of the RSB, solving the equations numerically with the population dynamics technique.

2.1.3 Algorithmic applications

A special interest for the study of the finite connectivity systems is explained by a mapping between a large number of optimization problems and the diluted spin glass models. An instance of an optimization problem corresponds to a particular realization of disorder in a spin glass system, and the optimal configuration is given by the ground state of the physical model. This analogy appeared to be very fruitful for bringing to the optimization problems the tools developed in statistical physics [MM09]. On the numerical side, the idea to use Monte Carlo for finding the solutions of the optimization problems has led to the development of the simulated annealing technique [KGV83]. Analytical methods from the statistical physics of disordered systems allowed to obtain a number of previously inaccessible results. In the early works, the replica method has been successfully applied to the travelling salesman problem [MP86a, MP86b], the weighted bipartite matching [MP85] and bi-partitioning [FA86]. Later on, a number of other optimization problems have been treated using the tools from the statistical physics of disordered systems.

The examples include the XORSAT problem, represented by a diluted p-spin model [CDMM03, MRTZ03]; the K-SAT problem, introduced to the physics community by Monasson and Zecchina [MZ96, MZ97] and followed by research on the effects of RSB [RTWZ01], leading to the efficient *survey propagation* algorithm [MPZ02, MZ02]; the 2+p-SAT problem [MZK⁺99]; the graph coloring [MPWZ02], and many more.

Importantly, in practical applications it is desired to find a solution to a given instance of the optimization problem on a single network. A very efficient local algorithm of this type, called *belief propagation* (BP), and closely related to the replica symmetric cavity method, allows for a computation of marginal probabilities in the graphical models in a distributed, *message-passing* way. Let us briefly present the ideas behind the BP algorithm that we will use in the following for introducing a general approach to dynamic problems; see [MM09] for more details on the BP approach.

Suppose that we have a graphical model defined on a tree graph by a joint probability distribution

$$P(\underline{\sigma}) = \frac{1}{Z} \prod_{a=1}^M \psi_a(\underline{\sigma}_{\partial a}), \quad (2.6)$$

where Z is the normalization constant. Note that this expression is given in a factorized form, each factor ψ_a representing a local interaction weight. Very often it is convenient to represent the graphical model in a form of a corresponding *factor graph* that reflects this structure of the model (2.6). The factor graph can be thought of as a bipartite graph $G = (V, F, E)$: V is a list of variables, $\underline{\sigma} = \{\sigma_i\}_{i \in V}$, and F represents a list of interactions, or function nodes, so that an edge $(i, a) \in E$ is present if the interaction $a \in F$ involves a variable σ_i in the node $i \in V$. The neighboring nodes in the factor graphs are denoted by ∂i and ∂a for the variables and function nodes, correspondingly. The concept of factor graph is illustrated in the Fig. 2.1 for the samples of EA, SK and p -spin models, mentioned previously. A review on the factor graph representations is given in [KFL01].

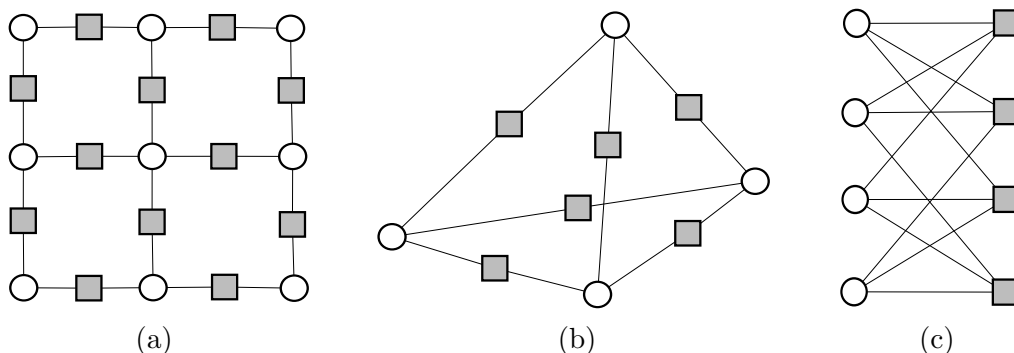


Figure 2.1: Factor graph representation of spin glass models with N spins: (a) two-dimensional EA model with $N = 9$, (b) SK model with $N = 4$, and (c) 3-spin model with $N = 4$. The hatched squares represent interactions between spins.

The *marginal* probability distribution (also called *belief*) that the variable on the site i takes value σ_i , is defined as

$$m^i(\sigma_i) = \sum_{\underline{\sigma}_{\setminus i}} P(\underline{\sigma}). \quad (2.7)$$

The basic idea of the method consists in the following observation: since the model is defined on a tree, when one removes a site i from the graph G and cuts the corresponding

connection to the neighboring interactions, the resulting *cavity graph* $G^{(i)}$ is given by a collection of independent and *statistically uncorrelated* branches of a tree. Therefore, the marginal $m^i(\sigma_i)$ can be expressed simply as a product over the conditional probabilities that represent the contributions of these branches:

$$m^i(\sigma_i) = \frac{1}{Z^i} \prod_{a \in \partial i} \hat{m}^{a \rightarrow i}(\sigma_i). \quad (2.8)$$

In this expression, Z^i is the normalization factor that ensures that $\sum_{\sigma_i} m^i(\sigma_i) = 1$, and ∂i stands for the neighbors of i in the factor graph. The probability $\hat{m}^{a \rightarrow i}(\sigma_i)$, called *message*, is defined as the marginal probability that node i takes value σ_i in the modified graph, in which all the interactions around i except a have been cut out. Now, in order to compute $m^i(\sigma_i)$, one needs to know the values of $\hat{m}^{a \rightarrow i}(\sigma_i)$. These quantities obey the coupled self-consistency equations

$$m^{i \rightarrow a}(\sigma_i) = \frac{1}{Z^{i \rightarrow a}} \prod_{b \in \partial i \setminus a} \hat{m}^{b \rightarrow i}(\sigma_i), \quad (2.9)$$

$$\hat{m}^{a \rightarrow i}(\sigma_i) = \frac{1}{Z^{a \rightarrow i}} \sum_{\underline{\sigma}_{\partial a \setminus i}} \psi_a(\underline{\sigma}_{\partial a}) \prod_{k \in \partial a \setminus i} m^{k \rightarrow a}(\sigma_k), \quad (2.10)$$

where we have also introduced another sort of messages, $m^{i \rightarrow a}(\sigma_i)$, which is defined as the marginal probability that node i takes value σ_i in the modified graph, in which the interaction a has been deleted. In these equations, $Z^{i \rightarrow a}$ and $Z^{a \rightarrow i}$ are the normalization constants. The coupled equations (2.9) and (2.10) are usually solved by iteration: first, one initialises all the messages to some value, and iterate the equations (2.9) and (2.10) until convergence. Then, the final values for the messages $\hat{m}^{a \rightarrow i}(\sigma_i)$ are used in (2.8) for computing the *exact values* of the marginal probability distributions $m^i(\sigma_i)$. This procedure explains the fine terminology of the BP algorithm: the iteration of equations (2.9) and (2.10) can be thought of as a message-passing protocol, each message holding a conditional information on the probability of the corresponding variable; the marginal is then given by a belief shaped by the information contained in all the messages arriving to the node.

Note that although the equations (2.8)-(2.10) have been derived for a tree graph, they can be viewed as an algorithm that can be run on an arbitrary interaction graph. They will provide accurate estimations of the marginals as long as the replica symmetric assumption holds for the interaction graph, i.e. that the correlations induced by loops decay fast enough, so that the approximation (2.8) as a product over neighboring interactions is correct. In particular, the BP equations (2.8)-(2.10) give asymptotically exact (in the thermodynamic limit $N \rightarrow \infty$) expressions for the beliefs on the only *locally tree-like* networks; in what follows, we will see that sparse random graphs, as well as many real-world networks of interest, fall into this category.

Sometimes, it is easier to eliminate one sort of messages in (2.9) and (2.10) and to use a single iteration equation for messages instead of two:

$$\hat{m}^{a \rightarrow i}(\sigma_i) = \frac{1}{Z^{\rightarrow i}} \sum_{\underline{\sigma}_{\partial a \setminus i}} \psi_a(\underline{\sigma}_{\partial a}) \prod_{k \in \partial a \setminus i} \prod_{b \in \partial k \setminus a} \hat{m}^{b \rightarrow k}(\sigma_k), \quad (2.11)$$

with $Z^{\rightarrow i} = Z^{a \rightarrow i} \prod_{k \in \partial a \setminus i} Z^{k \rightarrow a}$. This expression further simplifies in the important case of pairwise models, when the variables interact pairs by pairs, and the joint probability

distribution factorizes over the links in the graph:

$$P(\underline{\sigma}) = \frac{1}{Z} \prod_{(ij)} \psi_{ij}(\sigma_i, \sigma_j). \quad (2.12)$$

In this case, the update equation (2.11) can be rewritten exclusively in terms of messages $m^{i \rightarrow j}(\sigma_i)$, a short-cut for $m^{i \rightarrow (ij)}(\sigma_i)$:

$$m^{i \rightarrow j}(\sigma_i) = \frac{1}{Z^{i \rightarrow j}} \prod_{k \in \partial i \setminus j} \sum_{\sigma_k} \psi_{ik}(\sigma_i, \sigma_k) m^{k \rightarrow i}(\sigma_k). \quad (2.13)$$

Belief propagation, also known as the *sum-product* algorithm, has been rediscovered several times independently and in different contexts. It has been introduced by Pearl [Pea88] for the inference problems in acyclic Bayesian networks. In the field of coding theory, the BP equations appeared as an application to the low-density parity-check codes [Gal62, BG96, KS98]. In physics, the history can be traced back to the mean-field approximations to the Ising model, developed in terms of pseudo-marginals by Bethe [Bet35]. The replica symmetric cavity method applied to sparse spin-glass models has been discussed for the first time by Mézard and Parisi in [MP87a]. An overview of the belief propagation algorithm in the framework of factor graphs is given in [KFL01] and in [YFW00, YFW03, YFW05]. The last papers suggested a variational formulation of belief propagation in terms of stationary points of the Bethe free energy, and introduced the so-called generalized BP algorithm, related to the cluster variational method of Kikuchi [Kik51]. The loop corrections to BP have been considered in [MR05] and in [CC06a, CC06b], where it has been shown that the BP fixed points appear as a first-order approximation in a perturbative loop calculus procedure.

2.2 Dynamics of the Ising systems

There is no intrinsic dynamics in the static model defined by the Hamiltonian (2.2), and it has to be defined on a separate basis. In this section, we briefly discuss two popular choices for the dynamics of the spin-glass systems. We mostly focus on the basic definitions and techniques that are traditionally used in the studies of the disordered systems and deliberately put aside some interesting aspects about the aging of the spin glasses, the phenomenology of the dynamical phase transitions, mode-coupling approximations, as well as connections to the physics of structural glasses and hard spheres systems; for a review of these topics see, e.g., [BCKM96, BCKM98, Cug03].

2.2.1 Langevin-type dynamics

For a system defined by the Hamiltonian (2.2), one can define a relaxation dynamics of the Langevin type

$$\frac{d\sigma_i(t)}{dt} = -\frac{\partial H}{\partial \sigma_i} + \eta_i(t), \quad (2.14)$$

where $\eta_i(t)$ is a Gaussian thermal noise with zero mean and white-noise statistics.

The Langevin dynamics of the fully-connected models has been studied for the soft-spin version of the SK model by Sompolinsky and Zippelius in [SZ81, SZ82, Som81], see

also [MVP87, CK94]. The solution is obtained perturbatively with respect to the four-spin coupling constant (that large limit of this constant corresponds to the limit of the Ising spins) and allows to study the long-time asymptotic behavior of the spin correlators. A review on the relaxation dynamics of the SK model is given in [CK08].

The Langevin relaxation dynamics due to the Hamiltonian (2.2) for a spherical p -spin model has been studied by Crisanti, Horner and Sommers [CHS93] in the high-temperature phase. In the low-temperature phase, this system has been solved by Cugliandolo and Kurchan [CK93]. The dynamical TAP approach based on the path-integral techniques for p -spin model has been analysed in [Bir99].

Understanding the dynamics of the diluted p -spin model is a particularly important problem for the same reason as in the static case: still being of a mean-field nature, this system is closer to the finite dimensional case. In the thermodynamic limit $N \rightarrow \infty$, the description of the disordered average dynamics is not limited to the linear response and two-time auto-correlation functions, but should involve many-time correlation functions. The Monte Carlo simulations of the diluted p -spin model [RTZ00, MRT03a] have pointed out the heterogeneous character of the dynamics and demonstrated the validity of the out-of-equilibrium fluctuation-dissipation relations. The analytical investigation of the relaxation dynamics of the diluted spherical $p = 2$ spin model has been done in [SC03] for a case of a diagonal interaction matrix. The paper [SCM04] makes use of the generating functional techniques [MSR73, De 78] in order to construct a series of approximated equations for the averaged dynamics; note, however, that only typical properties can be studied within this approach, and a detailed description of a single spin dynamics in a given heterogeneous environment is not accessible.

2.2.2 Glauber dynamics

Another setting consists in imposing the Glauber dynamics for the Ising spins in the SK model. The stochastic Glauber dynamics [Gla63] at each type step is defined by a *local transition probability* $w_i(\sigma_i^{t+1} | \{\sigma_j^t\}_{j \in \partial i})$ that node i takes value σ_i^{t+1} at time $t + 1$ given the values $\{\sigma_j^t\}$ of its neighbors at time t :

$$w_i(\sigma_i^{t+1} | \{\sigma_j^t\}_{j \in \partial i}) = \frac{\beta \sigma_i^{t+1} \Delta_i^t}{2 \cosh(\beta \Delta_i^t)}, \quad (2.15)$$

where Δ_i^t is the local field at time t , created by the set of neighbors ∂i of i in the interaction graph:

$$\Delta_i^t = \sum_{j \in \partial i} J_{ij} \sigma_j^t. \quad (2.16)$$

In the fully-connected SK model, ∂i corresponds to all spins except i . The inverse temperature β controls the randomness of the dynamics: it is completely random for $\beta \rightarrow 0$ and deterministic for $\beta \rightarrow \infty$. One has to specify the choice of the time step and of the spin for each update. A simultaneous update of all nodes $\sigma^t \equiv \{\sigma_i^t\}_{i \in V}$ at once corresponds to the Markovian *parallel dynamics*:

$$P(\sigma^{t+1}) = \sum_{\sigma^t} W(\sigma^{t+1} | \sigma^t) P(\sigma^t), \quad (2.17)$$

with the transition probability

$$W(\sigma^{t+1} | \sigma^t) = \prod_{i \in V} w_i(\sigma_i^{t+1} | \{\sigma_j^t\}_{j \in \partial i}). \quad (2.18)$$

In the case of symmetric interactions $J_{ij} = J_{ji}$, the dynamics (2.17) obeys the detailed balance condition $W(\sigma' | \sigma)P(\sigma) = W(\sigma | \sigma')P(\sigma')$, and leads to the equilibrium distribution that can be formally written in the Boltzmann form

$$P_{eq}(\sigma) \propto e^{-\beta H_\beta(\sigma)}, \quad (2.19)$$

where $H_\beta(\sigma)$ is the Peretto's temperature-dependent *pseudo-Hamiltonian* [Per84]:

$$H_\beta(\sigma) = -\frac{1}{\beta} \sum_{i \in V} \log 2 \cosh(\beta \Delta_i). \quad (2.20)$$

A convergence to the equilibrium distribution with the Hamiltonian (2.3) happens under an asynchronous update of spins: node i is chosen randomly and independently from the set V , and the spin σ_i is updated with the probability (2.15) at each time step (which is now proportional to $1/N$ so that all the spins have been updated on average on a time scale $\mathcal{O}(N^0)$). This gives rise to the *sequential dynamics*, defined by the Markov chain (2.17), but now with the transition probability [Coo01]

$$W(\sigma^{t+1} | \sigma^t) = \frac{1}{N} \sum_{i \in V} \left[\prod_{j \neq i} \delta_{\sigma_j^{t+1}, \sigma_j^t} \right] w_i(\sigma_i^{t+1} | \{\sigma_j^t\}_{j \in \partial i}). \quad (2.21)$$

In the continuous time limit, the process defined by (2.17) and (2.21) gives rise to the master equation [BLLS71]

$$\frac{d}{dt} P(\sigma^t) = \sum_{i \in V} [w_i(F_i \sigma^t) P(F_i \sigma^t) - w_i(\sigma^t) P(\sigma^t)], \quad (2.22)$$

where $F_i \sigma^t \equiv (\sigma_1^t, \dots, -\sigma_i^t, \dots, \sigma_N^t)$ is the spin-flip operator, and

$$w_i(\sigma^t) = \frac{1}{2} [1 - \sigma_i^t \tanh \beta \Delta_i^t]. \quad (2.23)$$

In the case of symmetric interactions, this dynamics converges to the unique (for ergodic processes) equilibrium distribution

$$P_{eq}(\sigma) \propto e^{-\beta H_{eq}(\sigma)}, \quad (2.24)$$

with H_{eq} given by (2.3).

The dynamics of the SK model under Glauber dynamics has been considered in the original papers [SK75, KS78]. A systematic path-integral approach has been developed in [Som87], using the formalism introduced by De Dominicis [De 78]. Above the spin-glass transition, the Glauber dynamics of the SK model via the high-temperature expansion has been studied in [NY96]. Another approach in terms of deterministic flow equations for macroscopic state variables has been proposed by Laughton, Coolen and Sherrington [CS93, CS94, LCS96]. A development of this approach in the spirit of the continuous fraction expansion is given in [Sza98a, Sza98b].

More recently, several mean-field approximations have been suggested for the study of the Glauber dynamics of Ising systems, motivated by inference in the kinetic inverse Ising model [MS11, RH11, SRMH12, MS13, HK14]. In contrast to the symmetric networks ($J_{ij} = J_{ji}$), in the *fully asymmetric* case of SK model [CS88] the stationary state is not

described by Gibbs measure, but the dynamics exhibits small correlations among spins at different times. This property allows to use the Gaussian approximation and develop an exact formalism for the corresponding parallel dynamics [MS11], reminiscent to what has been done in earlier works in the context of asymmetric neural networks [DGZ87, GM88]: the method gives the exact values of local magnetizations and an exact relation between equal time and time-delayed correlations. The incorporation of information from previous times allows to get a better approximation for the application to the symmetric networks [MS13].

The numerical studies of the Glauber dynamics in diluted systems [BZ99, MRT03b] highlighted the effects of heterogeneity on the dynamics. The sequential dynamics of the Bethe lattice spin glass with binary couplings has been treated in [KH08] using cavity-like arguments. Hatchett *et al.* [HWC⁺04] have applied the generating functional analysis techniques [MSR73, De 78] to the parallel dynamics of Ising model on a Poissonian network. This approach has been generalized in [MC09a] for graphs with a given degree distribution. Although being exact, the number of order parameters (single-site spin and field paths) grows exponentially with time, making it intractable for a reasonable number of time steps. An alternative approach [SW04, HCCS05, MC08, MC09b], based on (or equivalent to) the dynamical replica analysis [CS94, LCS96] for the SK model, does not suffer from the growth of order parameters with time, but is not exact: one has to increase the set of observables for a better precision, which is not easy to control. Finally, the cavity method has been recently used to solve the Glauber dynamics on random graphs in terms of single-spin time trajectories [NB09, KM11, AM12]. This approach has a benefit of providing exact answers on locally tree-like graphs, but suffers from the problems similar to the generating functional method when applied to symmetric networks: the number of variables grows exponentially, making it difficult to solve the problem beyond only a few time steps. We will return to this point in more details in the section 2.4, where an equivalent formulation of this approach is presented in the form of dynamic belief propagation.

2.3 Spreading dynamics in complex systems

Before proceeding to the formulation of dynamic belief propagation as an approach to deal with dynamic problems, let us take a short survey into the dynamics of some prototypical complex system models of spreading processes, which are used to describe very diverse diffusion phenomena: propagation of infections [Mur89, AM91, Het00, BLM⁺06] and innovations [Rog10, CWW⁺08, LGLDG03, GDGP⁺02, SS98] in a social media, spread of blackouts and failures in power grid and computer networks [DCLN07, KCAL05], communication protocols, such as gossip algorithms and peer-to-peer file sharing [DGG⁺87, VRB03], avalanche dynamics in biological and neural networks [OCK13, KYK12], magnetic materials [DSS97, OS10] and financial markets [Kim04, Kos98]. Although all these cascading processes refer to different systems, their mathematical formulation relies on similar models that characterize the emergence of a collective behavior starting from the microscopic interactions between a large number of individuals in spatially extended systems. Conceptually, this setting is close to the statistical physics approach to the study of the non-equilibrium phase transitions in disordered systems [BBV08, CFL09, Ves11]. This fact explains the success of the techniques akin to the statistical physics of mean-field disordered models applied to the study of these complex systems. For instance, the bi-

nary variables used in voter [Gal02], majority-rule [KR03] and Sznajd [SWS00] models of opinion formation are similar to Ising spins, with continuous opinion variables [DNAW00] close to a soft-spin representation, while the few-state discrete variables used in epidemic [KM27] and other models can be viewed as Potts spins. However, as we will see, the dynamics may be defined not by a Hamiltonian, but rather through a set of dynamic transition rules. Typically, one is interested in the study of these models on realistic networks, either real, extracted from a particular system, or random, that models the system of interest in a satisfactory way. In this section, we will first mention some features of these networks that will be used through the rest of the thesis, and define several typical and common models used to approximate the spreading processes.

2.3.1 Some aspects of complex networks

It turns out that a careful study of a dynamic process can not be carried out without accounting for the properties of the underlying network, as the topology of the interaction graph enters in the very definition of the model. To cite an example, we will see that the topological parameters enter in the expressions of the critical thresholds for dynamic processes. For this reason, the network science itself became fundamental in the investigation of the out-of-equilibrium dynamics occurring on top of a complex topology. Recent years have seen a large number of works on the structure and emergence of complex networks, for a review see [AB02, New03, BLM⁺06, DGM08, BBV08, CH10].

A network $G = (V, E)$, represented by a collection of nodes (V) and edges (E) between them, is characterized by a number of parameters: node degrees and degree correlations, number of motifs (small subgraphs) and communities, diameter, betweenness, clustering coefficient, resilience, *etc.* Depending on the particular application, it is useful to construct and to study a relevant class of artificial graphs that mimic the properties of interest of a given real network. Hence, a primary role is played by the theory of *random graphs*, pioneered by Erdős and Rényi [ER60]. A random graph with a constrained degree profile can be constructed using the *configuration model* [Bol80, MR95]: given a *degree sequence* of N nodes in a network, $\{k_1, \dots, k_N\}$, the connections are made by choosing uniformly at random the matchings between the nodes' *stubs* (half-edges); in order to get a simple graph, one should explicitly forbid the formation of self-loops and multi-edges. The degree sequence can be drawn from some probability distribution $P(k)$. A random graph is called *regular*, see Fig. 2.2(a), if each degree is equal to some fixed value c :

$$P(k) = \delta(k - c). \quad (2.25)$$

A Poissonian random graph, also called *Erdős-Rényi* graph (drawn on the Fig. 2.2(b)), is obtained if the degrees are chosen according to the Poissonian law with a mean value c

$$P(k) = e^{-c} \frac{c^k}{k!}. \quad (2.26)$$

Many real-world networks are characterized by heavy-tailed and skewed degree distribution, with a few nodes (*hubs*) with very high degree distributions. These networks are often approximated by a *power-law* degree distribution of the form

$$P(k) \sim k^{-\gamma}, \quad (2.27)$$

giving rise to a high level of large-scale fluctuations [BA99]. A cartoon picture of this type of graph is given in the Fig. 2.2(c).

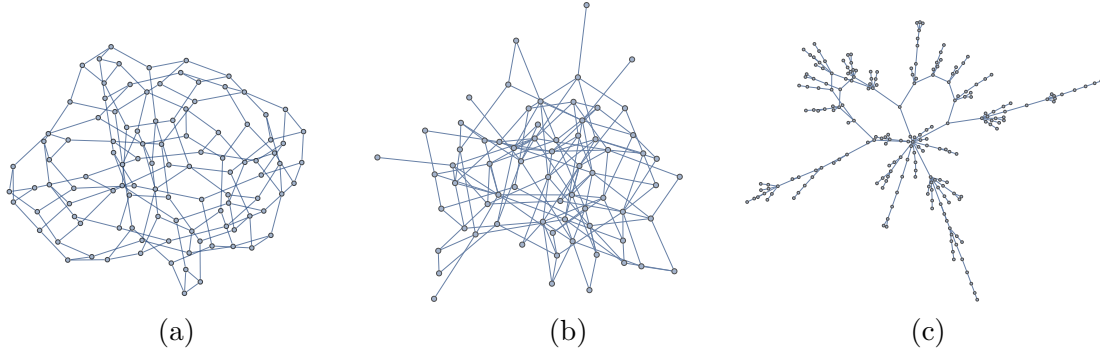


Figure 2.2: Cartoon examples of connected components of random graphs: (a) random regular, (b) Erdős-Rényi, and (c) power-law.

A random graph will be referred to as *dense*, if an average degree of nodes is of order of N , $\langle k \rangle = \mathcal{O}(N)$, as opposed to the *sparse* graph, in which $c = \langle k \rangle$ is of order of one. The examples of sparse networks include the finite-connectivity diluted model defined by (2.5), or a power-law graph with a finite variance $\langle k^2 \rangle$. The sparse networks with finite variance share a particular *tree-like* property: they look as trees at a distance scaling as $\log N$ [AS04]. More precisely, if l_i is the length of the shortest loop going through a randomly chosen node i , then with high probability $l_i \sim \log N / \log c$.

2.3.2 Spreading processes

Let us give a common example of a spreading process on a network, typically used to describe a propagation of pathogens or information in a network. The mathematical modeling of epidemic spreading is a subject of growing interest because of its importance for practical applications, such as analysis, evaluation and prevention of consequences of epidemiological processes [AM91]. The simplest epidemiological model is the *susceptible-infected-recovered* (SIR) model, introduced in 1927 [KM27]. In this model, a node in a network $G = (V, E)$ at each time can be in either of three states: susceptible (S), infected (I), or recovered (R). The propagation of infection on this model occurs due to the pairwise interactions between individuals: an individual in the state S can be infected by one of its I neighbors. The infected node can then switch to a recovered state with a certain rate, leading to an ultimate depletion of infected agents. These transition rules can be summarized by the following diagram [BLM⁺06]:

$$S(i) + I(j) \xrightarrow{\lambda_{ji}} I(i) + I(j) \quad (2.28)$$

$$I(i) \xrightarrow{\mu_i} R(i). \quad (2.29)$$

Another possibility of the recovery mechanism consists in a return of an infected individual to the susceptible state after some time: it corresponds to the so-called *susceptible-infected-susceptible* (SIS) model, introduced on a regular lattice by Harris in 1974 [Har74] under the name of the *contact process*, and which is used to model the behavior of endemic diseases, in general leading to an indefinite persistence of the disease. This recovery mechanism is encoded in the equation $I(i) \xrightarrow{\mu_i} S(i)$ that should be used instead of the equation (2.29) in the definition above.

As a first approximation, the dynamics (2.28)-(2.29) is studied within the *homogeneous mixing hypothesis*. It consists in neglecting the actual topology of the network, assuming

that the individuals interact through the whole population. This assumption allows one to write simple *naïve mean-field* (NMF) equations on the densities of susceptible (s), infected (i) and recovered (r) nodes in the population. For example, in the continuous time limit, and for homogeneous infection (λ) and recovery (μ) rates, one gets

$$\frac{ds(t)}{dt} = -\lambda\langle k \rangle i(t)s(t), \quad (2.30)$$

$$\frac{di(t)}{dt} = -\mu i(t) + \lambda\langle k \rangle i(t)s(t), \quad (2.31)$$

$$\frac{dr(t)}{dt} = \mu i(t), \quad (2.32)$$

with $\langle k \rangle$ being the average number of contacts of each individual. Although the mass-action mixing hypothesis is obviously a very crude approximation, the analysis of the equations (2.30)-(2.32) predicts the existence of the epidemic threshold λ_c , separating in the thermodynamic limit $N \rightarrow \infty$ the regions of parameters that lead either to the zero epidemic incidence $r(\infty) = 0$, or to a percolation of the infection to a finite fraction of population $r(\infty) > 0$. This threshold corresponds to the condition $\mathcal{R}_0 = 1$ on the *reproductive number* $\mathcal{R}_0 = \lambda\langle k \rangle / \mu$ that controls the average number of secondary infection cases, generated by a primary one in an entirely susceptible population, yielding $\lambda_c = \mu / \langle k \rangle$.

A more refined framework is given by the *heterogeneous mean-field* (HMF) approach [PSV01b, PSV01a, New02, MPSV02], in which the variables are aggregated according to their degree k , assuming that the nodes are statistically equivalent inside each degree block. This gives rise to the mean-field equations for the degree class densities $s_k(t)$, $i_k(t)$ and $r_k(t)$, similar to the equations (2.30)-(2.32):

$$\frac{ds_k(t)}{dt} = -\lambda k s_k(t) \Theta(t), \quad (2.33)$$

$$\frac{di_k(t)}{dt} = -\mu i_k(t) + \lambda k s_k(t) \Theta(t), \quad (2.34)$$

$$\frac{dr_k(t)}{dt} = \mu i_k(t), \quad (2.35)$$

where $\Theta(t)$ has a meaning of the probability that a given link points to an infected node [PSV01b, PSV01a]:

$$\Theta(t) = \frac{\sum_k k P(k) i_k(t)}{\langle k \rangle}, \quad (2.36)$$

with an assumption that the network is uncorrelated. The global quantities are obtained by averaging over the degree distribution $P(k)$: for example, the fraction of recovered nodes $r(t)$ at time t will be given by $r(t) = \sum_k P(k) r_k(t)$. The analysis of the large-time limit of the equations (2.33)-(2.35) yields the reproductive number $\mathcal{R}_0 = \lambda \mu^{-1} \langle k^2 \rangle / \langle k \rangle$, locating the critical threshold value at $\lambda_c = \mu \langle k \rangle / \langle k^2 \rangle$. This threshold has a finite value for networks with $\langle k^2 \rangle < \infty$, but vanishes in the graphs with strong connectivity fluctuations, for instance in the *scale-free* networks with $2 < \gamma \leq 3$ with $\langle k^2 \rangle \rightarrow \infty$ in the thermodynamic limit. The large-time limit of the SIR epidemic can also be studied using the similar heterogeneous mean-field and generating functions techniques, mapping the problem to the bond percolation model [New02, KR07].

The absence of the epidemic threshold has been first reported by analysing a variant of the equations (2.33)-(2.35) for the SIS model in [PSV01b, PSV01a]. Since then, a number

of rigorous results for the thresholds as a function of the network topology have been derived, see, e.g., [CPS10, Dur10, KNZ14, HP14]. In particular, many results relating the critical value to the largest eigenvalue of the adjacency matrix (or, more recently, of the so-called non-backtracking matrix [KNZ14, HP14]) have been established for the quenched networks, for which the adjacency matrix is fixed during the spreading process, as opposed to the assumptions of NMF and HMF, in which the edges are continuously reshuffled. Hence, HMF can in principle be used only to model the influenza-like illnesses with an infection period long with respect to the duration of transmission contacts, and is not suitable to the systems with an opposite time scale separation, where the infection spreads very rapidly.

Several recent investigations addressed this issue by carrying out a more advanced approach to the problem, providing the differential equations for the fraction of nodes in a heterogeneous environment that are susceptible, infected, or recovered at a given time [Vol08, KN10, Mil11], see also [MSV12]. Although averaged over the graph ensembles and initial conditions, these equations should not be confused with the NMF or HMF approximations that were derived under perfect mixing assumptions: the equations [Vol08, Mil11] appear to be exact in the thermodynamic limit, $N \rightarrow \infty$, for locally tree-like random graphs. Karrer and Newman [KN10] wrote a message-passing version of these equations by treating a more general SIR model (in which the transmission and recovery distributions are non-exponential), however, in a not very transparent convolutional form and averaged over initial conditions.

The derivation of these exact equations relies on identifying correct dynamic variables that are required to obtain the closed-form expressions. Let us present such a derivation for a message-passing version of these equations, applicable on a *single instance* of the contact network for a *given* initial condition. These *dynamic message-passing* (DMP) have been reported in [P-2], and are derived using a cavity-like argument, similar to what has been used for the derivation of the equation (2.8). In particular, we will show that the probabilities of being susceptible, infected, or recovered at a given time t as provided by the DMP equations are exact for all initial conditions and every realization of the transmission and recovery probabilities λ_{ij} and μ_i if the graph of contacts is a tree, and give asymptotically exact results for locally tree-like random graphs. When averaged over initial conditions or/and the graph ensemble, the DMP equations are equivalent to those of [Vol08, Mil11] and to a special case of [KN10].

For the purpose of derivation, we assume that the graph $G = (V, E)$ is a tree, and associate to each node i a Potts-like discrete-time variable σ_i^t that can take one of three values: susceptible, $\sigma_i^t = S$, infected, $\sigma_i^t = I$, or recovered, $\sigma_i^t = R$. Then, we define the discrete parallel dynamics of the SIR model: at each time step, an infected node i will recover with probability μ_i , and a susceptible node i will become infected with probability $1 - \prod_{k \in \partial i} (1 - \lambda_{ki} \delta_{\sigma_k(t), I})$, where ∂i is the set of neighbors of node i , and λ_{ki} measures the efficiency of spread from node k to node i . The recovered nodes never change their state. The set of quenched probabilities $\{\lambda_{ij}\}_{(ij) \in E}$, $\{\mu_i\}_{i \in V}$ determine a single instance of the problem.

Let us define $P_S^i(t)$, $P_I^i(t)$ and $P_R^i(t)$ as the marginal probabilities that $\sigma_i^t = S$, $\sigma_i^t = I$ and $\sigma_i^t = R$. These marginals sum to one and thus

$$P_I^i(t+1) = 1 - P_S^i(t+1) - P_R^i(t+1). \quad (2.37)$$

Since the recovery process from state I to state R is independent of neighbors, we have

$$P_R^i(t+1) = P_R^i(t) + \mu_i P_I^i(t). \quad (2.38)$$

The epidemic process on a graph can be interpreted as the propagation of infection signals from infected to susceptible nodes. The infection signal $d^{i \rightarrow j}(t)$ is defined as a random variable which is equal to one with probability $\delta_{\sigma_i^{t-1}, I} \lambda_{ij}$, and equal to zero otherwise. Consider an auxiliary dynamics D_j where node j receives infection signals, but ignores them and thus is fixed to the S state at all times. Since the infection cannot propagate through node j in this dynamic setting, different graph branches rooted at node j become independent if the underlying graph is a tree. This is the natural generalization of the cavity method to dynamic processes. Notice that the auxiliary dynamics D_j is identical to the original dynamics D for all times such that $\sigma_j^t = S$. We also define an auxiliary dynamics D_{ij} in which the state of a pair of neighboring nodes i and j is always S .

In order to obtain a closed system of message-passing equations, we write the remaining update rules in terms of three kinds of cavity messages, defined as follows. We first define the message $\theta^{k \rightarrow i}(t)$ as the probability that the infection signal has not been passed from node k to node i up to time t in the dynamics D_i :

$$\theta^{k \rightarrow i}(t) = \text{Prob}^{D_i} \left(\sum_{t'=0}^t d^{k \rightarrow i}(t') = 0 \right). \quad (2.39)$$

The quantity $\phi^{k \rightarrow i}(t)$ is the probability that the infection signal has not been passed from node k to node i up to time t in the dynamics D_i and that node k is in the state I at time t :

$$\phi^{k \rightarrow i}(t) = \text{Prob}^{D_i} \left(\sum_{t'=0}^t d^{k \rightarrow i}(t') = 0, \sigma_k^t = I \right). \quad (2.40)$$

Finally, $P_S^{k \rightarrow i}(t)$ is the probability that node k is in the state S at time t in the dynamics D_i :

$$P_S^{k \rightarrow i}(t) = \text{Prob}^{D_i} (\sigma_k^t = S). \quad (2.41)$$

In what follows, we prove that

$$P_S^{i \rightarrow j}(t+1) = P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(t+1), \quad (2.42)$$

where $\partial i \setminus j$ means the set of neighbors of i excluding j . Indeed, by definition

$$P_S^{i \rightarrow j}(t+1) = \text{Prob}^{D_j} (\sigma_i^{t+1} = S) = P_S^i(0) \text{Prob}^{D_j} \left(\sum_{k \in \partial i \setminus j} \sum_{t'=0}^{t+1} d^{k \rightarrow i}(t') \right). \quad (2.43)$$

Since the auxiliary dynamics D_{ij} coincides with dynamics D_j as long as node i is in the S state, we can write

$$P_S^{i \rightarrow j}(t+1) = P_S^i(0) \text{Prob}^{D_{ij}} \left(\sum_{k \in \partial i \setminus j} \sum_{t'=0}^{t+1} d^{k \rightarrow i}(t') \right). \quad (2.44)$$

Since different branches of the graph containing nodes $k \in \partial i \setminus j$ are connected only through node i , they are independent of each other, hence

$$P_S^{i \rightarrow j}(t+1) = P_S^i(0) \prod_{k \in \partial i \setminus j} \text{Prob}^{D_{ij}} \left(\sum_{t'=0}^{t+1} d^{k \rightarrow i}(t') \right). \quad (2.45)$$

Moreover, for nodes $k \in \partial i \setminus j$, the dynamics D_{ij} is equivalent to the dynamics D_i , so we can replace D_{ij} by D_i in the last expression and hence, using the definition (2.39), we obtain equation (2.42). We complete the updating rules by writing the equations for $\theta^{k \rightarrow i}(t)$ and $\phi^{k \rightarrow i}(t)$. The only way in which $\theta^{k \rightarrow i}(t)$ can decrease is by actually transmitting the infection signal from node k to node i , and this happens with probability λ_{ki} multiplied by the probability that node k was infected, so we have

$$\theta^{k \rightarrow i}(t+1) - \theta^{k \rightarrow i}(t) = -\lambda_{ki} \phi^{k \rightarrow i}(t). \quad (2.46)$$

The change for $\phi^{k \rightarrow i}(t)$ at each time step comes from three different possibilities: either node k actually sends the infection signal to node i (with probability λ_{ki}), either it recovers (with probability μ_k), or it switches to I at this time step, being previously in the S state (this happens with probability $P_S^{i \rightarrow j}(t-1) - P_S^{i \rightarrow j}(t)$):

$$\begin{aligned} \phi^{k \rightarrow i}(t) - \phi^{k \rightarrow i}(t-1) = & -\lambda_{ki} \phi^{k \rightarrow i}(t-1) - \mu_k \phi^{k \rightarrow i}(t-1) + \lambda_{ki} \mu_k \phi^{k \rightarrow i}(t-1) \\ & + P_S^{k \rightarrow i}(t-1) - P_S^{k \rightarrow i}(t). \end{aligned} \quad (2.47)$$

The third compensation term on the right-hand side of the previous equation has been introduced in order to avoid double-counting in the situation when node k transmits the infection and recovers at the same time step.

This completes the update rules for cavity messages. These equations can be iterated in time starting from initial conditions for cavity messages:

$$\theta^{i \rightarrow j}(0) = 1, \quad (2.48)$$

$$\phi^{i \rightarrow j}(0) = \delta_{\sigma_i^0, I}. \quad (2.49)$$

The marginal probability in the original dynamics D is obtained by including all the neighbor nodes $k \in \partial i$ in eq. (2.42):

$$P_S^i(t+1) = P_S^i(0) \prod_{k \in \partial i} \theta^{k \rightarrow i}(t+1). \quad (2.50)$$

Let us summarize the closed set of recursion rules, given by the combination of (2.37), (2.38), (2.42), (2.46), (2.47), (2.50):

$$P_S^{i \rightarrow j}(t+1) = P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(t+1), \quad (2.51)$$

$$\theta^{k \rightarrow i}(t+1) - \theta^{k \rightarrow i}(t) = -\lambda_{ki} \phi^{k \rightarrow i}(t), \quad (2.52)$$

$$\phi^{k \rightarrow i}(t) = (1 - \lambda_{ki})(1 - \mu_k) \phi^{k \rightarrow i}(t-1) - [P_S^{k \rightarrow i}(t) - P_S^{k \rightarrow i}(t-1)]. \quad (2.53)$$

The marginal probabilities that node i is in a given state at time t are then given as

$$P_S^i(t+1) = P_S^i(0) \prod_{k \in \partial i} \theta^{k \rightarrow i}(t+1), \quad (2.54)$$

$$P_R^i(t+1) = P_R^i(t) + \mu_i P_I^i(t), \quad (2.55)$$

$$P_I^i(t+1) = 1 - P_S^i(t+1) - P_R^i(t+1). \quad (2.56)$$

Together with the initial conditions (2.48-2.49), these equations give the exact values of marginal probabilities $P_S^i(t)$, $P_I^i(t)$ and $P_R^i(t)$ on a tree graph. The algorithmic complexity of DMP equations for a given vertex i is $\mathcal{O}(tNc)$, where c is the average degree of the graph.

Note that despite the similarity with the BP equations (2.9) and (2.10) that need to be iterated until convergence, the DMP equations are iterated in the physical time: starting from the initial conditions at time zero, by applications of the equations (2.51)-(2.56), one gets the values of marginal probabilities at time one, *etc.* In general, it is very difficult to guess the right dynamic messages (analogous to (2.39)-(2.41) in the SIR model) that should be used in the dynamic equations for more complicated models, involving a larger number of states and several non-trivial transitions. In the chapter 3, we will develop a general procedure for deriving the DMP equations for these complicated models, based on the dynamic belief propagation approach, presented in the next section.

2.4 Dynamic belief propagation

In this section, we will describe the cavity approach to the dynamics on networks, already mentioned in the section 2.2, in a simple form of belief propagation equations, described in the section 2.1. We will use this *dynamic belief propagation* (DBP) formulation as a starting point for obtaining new results, presented in the chapter 3. Equivalent formulations have appeared in [NB09, KM11, AM12, ABDZ13b]. The presentation of the dynamic belief propagation follows the publication [P-1].

Consider again a graph $G = (V, E)$, defined by a vertex set V and a set of edges E . Each vertex $i \in V$ is characterized by a variable, taking the value σ_i^t at time t . We assume here that the set of possible values of σ_i^t is a finite set of size K . We consider a generic dynamic process defined in a discrete-time parallel dynamics and described by a local transition probability $w_i(\sigma_i^{t+1} | \{\sigma_j^t\}_{j \in \partial i})$ that node i takes value σ_i^{t+1} at time $t+1$ given the values $\{\sigma_j^t\}$ of its neighbors at time t (the set of neighbors of i on the graph is denoted by ∂i).

If we denote by $\vec{\sigma}_i = (\sigma_i^0, \dots, \sigma_i^T)$ the trajectory of variable i at times $t = 0, \dots, T$, where T is defined as the stopping time, the joint probability distribution of the trajectories $P(\{\vec{\sigma}_i\}_{i \in V})$ can be written as follows:

$$P(\{\vec{\sigma}_i\}_{i \in V}) = \prod_{i \in V} \prod_{t=0}^{T-1} w_i(\sigma_i^{t+1} | \{\sigma_j^t\}_{j \in \partial i}) P_0, \quad (2.57)$$

where $P_0 \equiv P(\{\sigma_i^0\}_{i \in V})$ is the distribution of variables at initial time.

It is a well-known fact that BP equations are exact for static graphical models when the factor graph is a tree. However, when we consider the factor graph of the model defined in (2.57) (where the variables are time trajectories $\vec{\sigma}_i$), it turns out that the factor graph contains many loops, even in the case where G is a tree, see Fig. 2.3.

A way to fix this problem consists in exploiting the duality between variables and interactions by putting the variables on the edges. To this purpose, we introduce a different representation of the problem, that uses auxiliary variables (time-trajectories) $\vec{\sigma}_{i \rightarrow j}$ on each directed edge $(i, j) \in E$. For a given i , all the variables $\vec{\sigma}_{i \rightarrow j}$ are supposed to be copies of the original $\vec{\sigma}_i$. They should thus be all equal, and we implement this by adding for each i an additional constraint $\vec{\sigma}_{i \rightarrow j} = \vec{\sigma}_{i \rightarrow k}$ for all $j, k \in \partial i$. The joint

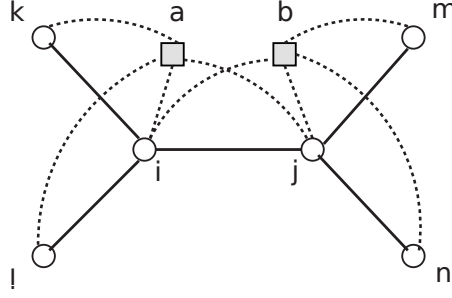


Figure 2.3: An example of a factor graph of the graphical model at two nearest times described by the joint probability distribution $P(\{\vec{\sigma}_i\}_{i \in V})$. The check node a represents interaction between the variable σ_i^{t+1} and the variables $\{\sigma_j^t\}_{j \in \partial i}$ at a previous time step. This factor graph is characterized by systematic short loops.

probability distribution (2.57) of time trajectories can hence be written in terms of these new variables:

$$P(\{\vec{\sigma}_{i \rightarrow j}, \vec{\sigma}_{j \rightarrow i}\}_{(i,j) \in E}) = \prod_{i \in V} \prod_{t=0}^{T-1} \left[w_i(\sigma_i^{t+1} \mid \{\sigma_{k \rightarrow i}^t\}_{k \in \partial i}) \prod_{k \in \partial i \setminus l} \delta_{\sigma_{i \rightarrow l}^t, \sigma_{i \rightarrow k}^t} \right] P_0, \quad (2.58)$$

where l is any of the variables influenced directly by i , and $k \in \partial i \setminus l$ means the set of nodes neighboring node i , excluding l . This new form of the probability distribution is described by a factor graph which is very closely related to G : the new variables $\vec{\sigma}_{i \rightarrow j}, \vec{\sigma}_{j \rightarrow i}$ live on each edge $(ij) \in E$, and there is a function node (interaction) associated with every vertex $i \in V$. If the original graph G is a tree, the factor graph is also a tree, see Fig. 2.4. This crucial property allows to use the BP method in terms of this new description for studying the dynamics, with the guarantee that the resulting equations are exact if G is a tree.

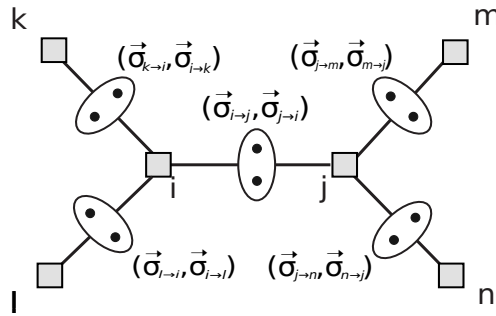


Figure 2.4: An example of a factor graph of the graphical model at all times described by the joint probability distribution $P(\{\vec{\sigma}_{i \rightarrow j}, \vec{\sigma}_{j \rightarrow i}\}_{(i,j) \in E})$. The check node i represents interaction between trajectories $\vec{\sigma}_i$ and $\{\vec{\sigma}_j\}_{j \in \partial i}$. This factor graph is characterized by the underlying structure of the original graph.

Let us now write the BP equations, using the form of BP for single type of messages (2.11). Using the fact that $\vec{\sigma}_{i \rightarrow j} = \vec{\sigma}_{i \rightarrow k}$ for all $j, k \in \partial i$, it is convenient to rename the variables $\{\vec{\sigma}_{i \rightarrow j}, \vec{\sigma}_{j \rightarrow i}\}_{(i,j) \in E}$ to $\{\vec{\sigma}_i, \vec{\sigma}_j\}_{(i,j) \in E}$. We can write the BP equation for the joint probability (2.58) in terms of messages $m^{i \rightarrow j}(\vec{\sigma}_i, \vec{\sigma}_j)$:

$$m^{i \rightarrow j}(\vec{\sigma}_i, \vec{\sigma}_j) = \frac{1}{Z^{i \rightarrow j}} \sum_{\{\vec{\sigma}_k\}_{k \in \partial i \setminus j}} \left[\prod_{t=0}^{T-1} w_i(\sigma_i^{t+1} \mid \{\sigma_k^t\}_{k \in \partial i \setminus j}, \sigma_j^t) P_0 \right] \prod_{k \in \partial i \setminus j} m^{k \rightarrow i}(\vec{\sigma}_k, \vec{\sigma}_i). \quad (2.59)$$

The normalization constant $Z^{i \rightarrow j}$ can be calculated explicitly for the markovian dynamics from the normalization condition

$$\sum_{\vec{\sigma}_i, \vec{\sigma}_j} m^{i \rightarrow j}(\vec{\sigma}_i, \vec{\sigma}_j) = 1 \quad (2.60)$$

For example, in the case of a general Markov dynamics we use the fact that $m^{k \rightarrow i}(\vec{\sigma}_k, \vec{\sigma}_i)$ does not depend on σ_i^T and perform the summation first over σ_j^T , then over σ_i^T , and so on for the times $T-1, \dots, 0$. Finally, we get the normalization factor

$$Z^{i \rightarrow j} = \frac{1}{2^{(T+1)(d_i-2)}} \quad (2.61)$$

for this case, where d_i is the number of neighbors of the node i in the initial graph.

The message $m^{i \rightarrow j}(\vec{\sigma}_i, \vec{\sigma}_j)$ has the meaning of the probability for the trajectories $\vec{\sigma}_i, \vec{\sigma}_j$ in the transformed cavity graph, where the factor node j has been removed. We can also rewrite the equation (2.59) in terms of conditional probabilities $m^{i \rightarrow j}(\vec{\sigma}_i \mid \vec{\sigma}_j)$ in the cavity graph. Thus, for the dynamics obeying the Markov property we have

$$\sum_{\vec{\sigma}_i} m^{i \rightarrow j}(\vec{\sigma}_i, \vec{\sigma}_j) = \frac{1}{2^{T+1}}, \quad (2.62)$$

and hence get

$$m^{i \rightarrow j}(\vec{\sigma}_i \mid \vec{\sigma}_j) = \sum_{\{\vec{\sigma}_k\}_{k \in \partial i \setminus j}} \left[\prod_{t=0}^{T-1} w_i(\sigma_i^{t+1} \mid \{\sigma_k^t\}_{k \in \partial i \setminus j}, \sigma_j^t) P_0 \right] \prod_{k \in \partial i \setminus j} m^{k \rightarrow i}(\vec{\sigma}_k \mid \vec{\sigma}_i). \quad (2.63)$$

The normalization factor in this equation is exactly equal to 1, due to the Markov property of the dynamics. Note that, again, by construction, $m^{k \rightarrow i}(\vec{\sigma}_k \mid \vec{\sigma}_i)$ does not depend on σ_i^T , so this variable can be erased. Then, as far as $\prod_{t=0}^{T-1} w_i(\sigma_i^{t+1} \mid \{\sigma_k^t\}_{k \in \partial i \setminus j}, \sigma_j^t)$ does not depend on σ_k^T , we can perform the sum over σ_k^T in the right-hand side of the equation (2.63), which yields the special form of the equation, appeared in the work [KM11] for the study of the voter model:

$$m_{T+1}^{i \rightarrow j}(\vec{\sigma}_i \mid \vec{\sigma}_j) = \sum_{\{\sigma_k^0, \dots, \sigma_k^{T-1}\}_{k \in \partial i \setminus j}} \left[\prod_{t=0}^{T-1} w_i(\sigma_i^{t+1} \mid \{\sigma_k^t\}_{k \in \partial i \setminus j}, \sigma_j^t) P_0 \right] \prod_{k \in \partial i \setminus j} m_T^{k \rightarrow i}(\vec{\sigma}_k \mid \vec{\sigma}_i), \quad (2.64)$$

where we denote $m_{T+1}^{i \rightarrow j}(\vec{\sigma}_i \mid \vec{\sigma}_j) = m^{i \rightarrow j}(\vec{\sigma}_i \mid \vec{\sigma}_j)$ and $m_T^{i \rightarrow j}(\vec{\sigma}_i \mid \vec{\sigma}_j) = \sum_{\sigma_i^T} m_{T+1}^{i \rightarrow j}(\vec{\sigma}_i \mid \vec{\sigma}_j)$.

The message $m^{i \rightarrow j}(\vec{\sigma}_i \mid \vec{\sigma}_j)$ in (2.63) has the meaning of the probability for the trajectory $\vec{\sigma}_i$ given the trajectory $\vec{\sigma}_j$ in the transformed cavity graph, where the factor node j has been removed. We denote the dynamics in the corresponding cavity graph as D_j .

The equation (2.63) can be iterated until convergence, and the corresponding marginal probability of a time trajectory $\vec{\sigma}_i$ will be given by

$$m^i(\vec{\sigma}_i) = \sum_{\{\vec{\sigma}_k\}_{k \in \partial i}} \left[\prod_{t=0}^{T-1} w_i(\sigma_i^{t+1} \mid \{\sigma_k^t\}_{k \in \partial i}) P_0 \right] \prod_{k \in \partial i} m^{k \rightarrow i}(\vec{\sigma}_k \mid \vec{\sigma}_i). \quad (2.65)$$

Note that in the general case, it takes an exponential number of operations in the duration of the process to solve the equations (2.63) and (2.65), since each message has K^T components, where K is the number of values that each variable σ_i^t may take, and the sum in (2.63) is performed over $K^{T(d_i-1)}$ variables for each node i , with d_i being the number of neighbors of i . However, we will see that a crucial simplification occurs for the models with unidirectional dynamics, introduced in the chapter 3.

If the graph G is purely directed, meaning that the direct influence between neighboring nodes i and j runs only in one direction, another immediate simplification occurs in the DBP approach. In this case, $m^{i \rightarrow j}(\vec{\sigma}_i \mid \vec{\sigma}_j)$ does not depend on the variable $\vec{\sigma}_j$, and the equation (2.63) is reduced to

$$m^{i \rightarrow j}(\vec{\sigma}_i) = \sum_{\{\vec{\sigma}_k\}_{k \in \partial_{in} i}} \left[\prod_{t=0}^{T-1} w_i(\sigma_i^{t+1} \mid \{\sigma_k^t\}_{k \in \partial_{in} i}) P_0 \right] \prod_{k \in \partial_{in} i} m^{k \rightarrow i}(\vec{\sigma}_k). \quad (2.66)$$

Therefore, writing the marginal in a factorized form

$$m^i(\vec{\sigma}_i) = \prod_t m_t^i(\sigma_i^t), \quad (2.67)$$

we immediately get from (2.65) for $t > 0$

$$m_{t+1}^i(\sigma_i^{t+1}) = \sum_{\{\sigma_k^t\}_{k \in \partial_{in} i}} w_i(\sigma_i^{t+1} \mid \{\sigma_k^t\}_{k \in \partial_{in} i}) \prod_{k \in \partial_{in} i} m_t^k(\sigma_k^t). \quad (2.68)$$

Note, however, that in the case of the dynamics that depends on the state of the same node at the previous time, the factorization (2.67) does not lead to a decoupled expression, and we have

$$\prod_t m_{t+1}^i(\sigma_i^{t+1}) = \prod_t \left[\sum_{\{\sigma_k^t\}_{k \in \partial_{in} i}} w_i(\sigma_i^{t+1} \mid \{\sigma_k^t\}_{k \in \partial_{in} i}, \sigma_i^t) P_0 \prod_{k \in \partial_{in} i} m_t^k(\sigma_k^t) \right]. \quad (2.69)$$

The DBP equations (2.63) and (2.65) are equivalent to those of [NB09, KM11], but are presented in a simple belief-propagation-like form for a generic local transition probability $w_i(\sigma_i^{t+1} \mid \{\sigma_j^t\}_{j \in \partial i})$. Although the time factorization ansatz (2.67) is exact only for the fully asymmetric networks, it has been also used (under the name “one-time approximation”) in [NB09] and in [AM12] for symmetric networks, breaking the exactness of the approach even for tree graphs. We will not use this approximation since we are interested in the exact results, and instead look on the dynamic side of what can be done from the DBP equations. We will see that the exact dynamic message-passing equations can be derived in the DBP framework for a large class of models with unidirectional dynamics, which is the subject of the next chapter.

Chapter 3

Dynamic processes with unidirectional dynamics

In this chapter, using the dynamic cavity method on time trajectories, we construct a general procedure for deriving the dynamic message-passing equations for a large class of models with unidirectional dynamics – the key ingredient that makes the problem solvable. These equations are applicable to single instances of the corresponding problems with arbitrary initial conditions, and are asymptotically exact for problems defined on locally tree-like graphs. They also provide a good analytic approximation of the real dynamics on a number of tested real-world networks. The presented results are reported in detail in [P-1]; in this chapter, we will summarize the principle results, and discuss some unpublished details – in particular, the large-time limit of the DMP equations for the SIR and the rumor spreading models.

3.1 Unidirectional dynamics: preliminary considerations

We have seen in the section 2.4 that the solution of the dynamic belief propagation equations (2.63) and (2.65) require in general an exponential number of operations in the duration of the process. However, in a few special cases, some progress has been recently made by a number of authors that were able to write, using cavity-like arguments, tractable asymptotically-exact mean-field dynamic equations for several models defined on locally tree-like graphs, such as the zero-temperature random field Ising model (RFIM) [SDS⁺93, OS10], the susceptible-infected-recovered model [Vol08, KN10, Mil11, MSV12] (cf. section 2.3), and the threshold models [ABDZ13a, SM14]: the precise definitions of these models will be provided below. All these models share a common property: they describe a *unidirectional dynamics* involving one complex transition to the active state; the derivation of the corresponding equations is typically based on identifying correct dynamic variables that are required to obtain the closed-form expressions. These examples lead to the hypothesis that the *microscopic irreversibility* of the dynamics is a key property that makes it possible to derive such equations [Luc12, ABDZ13b]. However, in general it is very difficult to guess the right dynamic variables that should be used in the dynamic equations for more complicated models, involving a larger number of states and several non-trivial transitions. Perhaps, the simplest model of this kind is the so-called rumor spreading model [DK64, DK65, MT73], which is a three-state dynamic model with

two neighbors-dependent transitions.

In this chapter, we develop a general procedure for deriving, starting from the DBP equations, the *dynamic message-passing* (DMP) equations for models with unidirectional dynamics and arbitrary number of states. They allow one to estimate the marginal probabilities of each variable at each time on a given network of contacts, using a number of operations that is polynomial both in the size of the network and in the duration of the dynamic process. These equations are applicable to single-instance problems with arbitrary initial conditions, they are asymptotically exact on locally tree-like networks, and typically provide a good approximation for real-world networks. We show that the unidirectional nature of the dynamics is indeed a crucial element that makes the problem solvable. More precisely, the time trajectories in these models can be fully parametrized with only a few flipping times, leading to a significant simplification of the corresponding dynamic BP equations. As a result, these equations can be rewritten in terms of closed-form DMP equations with a computational complexity which turns out to be reduced from an exponential in the duration of the process to a polynomial. This simplification occurs thanks to the use of dynamic variables that appear naturally to be the weighted sums of messages of the DBP equations on trajectories.

Let us give here some preliminary definitions of the unidirectional processes. We assume that in the expression for the transition probability $w_i(\sigma_i^{t+1} \mid \{\sigma_j^t\}_{j \in \partial i})$, the value σ_i^t for some node i and time t takes one of the K ordered discrete values that we denote $\Omega_1, \Omega_2, \dots, \Omega_K$. We call the dynamic process *unidirectional* if each node can change its state only in a directed and irreversible way:

$$\Omega_1 \rightarrow \Omega_2 \rightarrow \dots \rightarrow \Omega_K, \quad (3.1)$$

and the transition to one of the previous states is forbidden by the dynamic rules.

Among unidirectional processes with $K = 2$ states (σ_i^t can take one of the two values $-1 \equiv \downarrow$ or $1 \equiv \uparrow$), one can mention the zero-temperature random field Ising model with homogeneous initial condition, considered in [OS10]:

$$\prod_{i \in V} \delta_{\sigma_i^0, -1} = (\downarrow \downarrow \dots \downarrow). \quad (3.2)$$

Each spin can flip only if the local field created by its neighbors is positive (precise definitions will be given in the next section). Once being flipped, the spins in this avalanche dynamics remain in the position \uparrow for all times, since the local field is a monotone non-decreasing function of time. Therefore, this system has a unidirectional dynamics with two ordered states, \downarrow and \uparrow . The model can be generalized to any initial condition if the backward transition from \uparrow to \downarrow is explicitly forbidden by the dynamic rules. The situation is different from the standard Glauber dynamics of the Ising model with non-zero temperature (2.15), or from the majority dynamics of voters that switch to one of the alternative opinions according to the majority of their neighbors [KM11], where each variable is free to flip an arbitrary number of times. Note that the *linear threshold model* (LTM) with random thresholds studied in [ABDZ13a, ABDZ13b] is equivalent to this formulation of the RFIM.

Another example of model with two states and unidirectional dynamics is given by the *susceptible-infected* (SI) model, in which the node can be in either of two states: susceptible (S), or infected (I). The propagation of infection on this model occurs due to the pairwise interactions between individuals: for instance, the S individual can be

infected by one of its I neighbors at each time step, and then remains infected forever. Obviously, the SI model is directly related to the SIR model (2.28)-(2.29). Note also that if there exists a recovery mechanism that allows an infected individual to become susceptible again after some time, the resulting SIS model, considered in the section 2.3, does not belong to the class of unidirectional models anymore.

Unidirectional dynamic processes with $K = 3$ states include the previously defined susceptible-infected-recovered (SIR) model (2.28)-(2.29). Another well known model with unidirectional dynamics and three states is given by the so-called *rumor spreading*, or the *ignorant-spreader-stifler* (ISS) model, introduced by Daley and Kendall [DK64, DK65] and further revisited by Maki and Thompson [MT73], which describes the propagation of information by spreaders to ignorants that are unaware of rumor, and takes into account the possibility that the spreader can become uninterested in the rumor under the influence of its neighbors. Depending on the setting, the rumor spreading can be regarded as an adverse problem to the epidemic spreading: besides an accurate modeling of the infection outbreak, the aim is to develop the mitigating strategies in order to minimize the impact of the disease spread. On the contrary, the rumor spreading models are usually designed in a way to maximize the information dissemination. The precise formulation of the ISS model will be given in the section 3.3.

In what follows, we discuss the dynamic message-passing equations for the models listed above, illustrating the general method to derive such equations for other models with arbitrary K . Typically, these equation would allow to answer the following question: *what is the probability that a certain node i is in a certain state Ω_a at time t ?* As we have seen, for some of these models, the equivalent equations have already appeared in the literature, however, in a form averaged over the ensemble of random graphs and/or over the initial conditions, and not suitable to the algorithmic purposes for single-instance problems. For others, the DMP equations have never been stated previously. For each model considered in the following, we will discuss the relation of our DMP equations to those existing in the literature, if any.

3.2 Models with 2 states

As a first example, we illustrate the method in detail on the most general case of unidirectional dynamic model with two states and pairwise interactions between nodes: the generalized SI model. The derivation of the DMP equations for other models will follow a similar scheme: starting from the general DBP equation (2.63) on trajectories, parametrized by a few flipping times (thanks to the irreversibility of the dynamics), and using a properly defined dynamic kernel for a particular model, we form the dynamic messages of interest, typically represented by the marginal probabilities for different states. These marginals, defined in the cavity graph, are given by the corresponding sums of DBP messages. After sometimes cumbersome, but straightforward algebra, we are able to obtain the closed-form equations, using the dynamic messages that emerge automatically as the weighted sums of DBP messages. These expressions of dynamic messages allow us to give them a concrete physical sense. Although for some simple models as the SI process these messages can be guessed from the beginning (see, e.g., the derivation in the section 2.3), it becomes very hard to get them for more complicated models: the examples of such models will be given further, in the sections 3.3 and 3.4, where we will point out some subtleties that appear for models with a larger number of states.

3.2.1 SI model

The definition of the generalized SI model in discrete time can be represented as follows:

$$S(i) + S(j) \xrightarrow{\epsilon_{ji}} I(i) + S(j), \quad (3.3)$$

$$S(i) + I(j) \xrightarrow{\lambda_{ji}} I(i) + I(j), \quad (3.4)$$

$$S(i) \xrightarrow{\nu_i} I(i). \quad (3.5)$$

This diagram represents the dynamic rules at each time step. Here, i and j mean two neighboring nodes in the network, and ϵ_{ji} , λ_{ji} and ν_i correspond to the transition probabilities at each time step. Each of these independent interactions may lead to a transition to the final state.

Since there are only two possible states and the dynamics is unidirectional, the time trajectory $\vec{\sigma}_i$ of node i has a typical form $|S_0 SSSSSSI_{\tau_i} I I I I I_T\rangle$, and hence can be parametrized by a single time, τ_i , when the spin flips from the state S to the state I (τ_i is the first time for which $\sigma_i^{\tau_i} = I$). If node i is initially in the state I , we set $\tau_i = 0$, and we put by definition $\tau_i = T$ if the flipping happens after the observation time T , or never happens (T is the stopping time, i.e. the condition $\tau_i = T$ summarizes all the events that happen after the time T).

Therefore, the dynamic cavity equation (2.63) for the generalized SI model takes the following form for $\tau_i < T$:

$$m^{i \rightarrow j}(\tau_i | \tau_j) = \sum_{\{\tau_k\}_{k \in \partial i \setminus j}} W_{SI} \prod_{k \in \partial i \setminus j} m^{k \rightarrow i}(\tau_k | \tau_i), \quad (3.6)$$

where W_{SI} is the kernel that resumes the dynamics of the model up to the final time T :

$$\begin{aligned} W_{SI} &= P_I^i(0) \mathbb{1}[\tau_i = 0] \\ &+ P_S^i(0) \mathbb{1}[\tau_i > 0] \prod_{t'=0}^{\tau_i-2} (1 - \nu_i) \prod_{k \in \partial i} (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t' + 1]) (1 - \lambda_{ki} \mathbb{1}[t' \geq \tau_k]) \\ &\times \left(1 - (1 - \nu_i) \prod_{k \in \partial i} (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq \tau_i]) (1 - \lambda_{ki} \mathbb{1}[\tau_i \geq \tau_k + 1]) \right). \end{aligned} \quad (3.7)$$

Here and in what follows we use a convention

$$\prod_{t=a}^{a-\epsilon} (\dots) \equiv 1 \quad (3.8)$$

for any fixed a and $\epsilon > 0$.

The messages $m^{i \rightarrow j}(\tau_i | \tau_j)$ allow one to define the marginal probabilities describing the dynamics of the SI model:

$$P_S^i(t) = \sum_{\tau_i > t} m^i(\tau_i), \quad (3.9)$$

$$P_I^i(t) = 1 - P_S^i(t). \quad (3.10)$$

It is also useful to define the marginal probability that node i is in the state S at a given time in the cavity graph D_j , in which the node j is fixed to the state S for all times:

$$P_S^{i \rightarrow j}(t) = \sum_{\tau_i > t} m^{i \rightarrow j}(\tau_i | T). \quad (3.11)$$

The DMP equations for this model can be derived, starting from the equation (3.6), and using the definitions (3.9), (3.10), as well as the elementary properties of the messages, such as normalization and causality constraints:

Property 1. For every fixed τ_j

$$\sum_{\tau_i=0}^T m^{i \rightarrow j}(\tau_i | \tau_j) = 1. \quad (3.12)$$

Property 2. If $\tau_j > \tau_i$, then for every $t' > \tau_i$

$$m^{i \rightarrow j}(\tau_i | \tau_j) = m^{i \rightarrow j}(\tau_i | t'). \quad (3.13)$$

Using the definition (3.9), we get for $t > 0$

$$\begin{aligned} & P_S^{i \rightarrow j}(t+1) \\ &= P_S^i(0) \sum_{\tau_i > t+1} \sum_{\{\tau_k\}_{k \in \partial i \setminus j}} \mathbb{1}[\tau_j = T] \prod_{t'=0}^{\tau_i-2} (1 - \nu_i) \prod_{k \in \partial i} (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t' + 1]) (1 - \lambda_{ki} \mathbb{1}[t' \geq \tau_k]) \\ & \times \left(1 - (1 - \nu_i) \prod_{k \in \partial i} (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq \tau_i]) (1 - \lambda_{ki} \mathbb{1}[\tau_i \geq \tau_k + 1]) \right) \prod_{k \in \partial i \setminus j} m^{k \rightarrow i}(\tau_k | \tau_i). \end{aligned} \quad (3.14)$$

An important observation that can be made on this expression is that one can replace $\prod_{k \in \partial i \setminus j} m^{k \rightarrow i}(\tau_k | \tau_i)$ in the right-hand side of the last expression by $\prod_{k \in \partial i \setminus j} m^{k \rightarrow i}(\tau_k | T)$ for arbitrary value of the stopping time $T > t + 1$. The easiest way to see it consists in observing that the value of the probability $P_S^i(t+1)$ should not depend on the value of the stopping time provided that $T > t + 1$, and can be assigned to an arbitrary value. Directly from the definition (3.11), $P_S^{i \rightarrow j}(t+1)$ involves only the messages $m^{i \rightarrow j}(\tau_i | T)$ with $\tau_i > t + 1$, and in particular we can choose $T = t + 2$. Since the stopping time comprises by definition all the events that happen after the time T , we get

$$\begin{aligned} & P_S^{i \rightarrow j}(t+1) = P_S^i(0) (1 - \nu_i)^{t+1} (1 - \epsilon_{ji})^{t+1} \\ & \times \sum_{\{\tau_k\}_{k \in \partial i \setminus j}} \prod_{k \in \partial i \setminus j} \prod_{t'=0}^t (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t' + 1]) (1 - \lambda_{ki} \mathbb{1}[t' \geq \tau_k]) m^{k \rightarrow i}(\tau_k | T) \\ & \times \left[\sum_{\tau_i > t+1} \mathbb{1}[\tau_j = T] \prod_{t'=t+1}^{\tau_i-2} (1 - \nu_i) \prod_{k \in \partial i} (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t' + 1]) (1 - \lambda_{ki} \mathbb{1}[t' \geq \tau_k]) \right. \\ & \times \left. \left(1 - (1 - \nu_i) \prod_{k \in \partial i} (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq \tau_i]) (1 - \lambda_{ki} \mathbb{1}[\tau_i \geq \tau_k + 1]) \right) \right]. \end{aligned} \quad (3.15)$$

The sum in the square brackets gives exactly one, and we obtain a simple factorized expression

$$P_S^{i \rightarrow j}(t+1) = P_S^i(0) (1 - \lambda_{ji})^{t+1} (1 - \nu_i)^{t+1} \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(t+1), \quad (3.16)$$

where $\theta^{k \rightarrow i}(t+1)$ are given by

$$\theta^{k \rightarrow i}(t+1) = \sum_{\tau_k} \prod_{t'=0}^t (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t' + 1]) (1 - \lambda_{ki} \mathbb{1}[t' \geq \tau_k]) m^{k \rightarrow i}(\tau_k | T). \quad (3.17)$$

In order to close the equations on $P_S^{k \rightarrow i}(t)$, we recover the computational scheme for $\theta^{k \rightarrow i}(t+1)$:

$$\begin{aligned} \theta^{k \rightarrow i}(t+1) - \theta^{k \rightarrow i}(t) &= \sum_{\tau_k} \left(\prod_{t'=0}^{t-1} (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t' + 1]) (1 - \lambda_{ki} \mathbb{1}[t' \geq \tau_k]) \right) \\ &\quad \times (-\epsilon_{ki} \mathbb{1}[\tau_k \geq t+1] - \lambda_{ki} \mathbb{1}[t \geq \tau_k]) m^{k \rightarrow i}(\tau_k | T) \\ &= -\epsilon_{ki} \sum_{\tau_k} \left(\prod_{t'=0}^{t-1} (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t' + 1]) \right) m^{k \rightarrow i}(\tau_k | T) \mathbb{1}[\tau_k \geq t+1] \\ &\quad - \lambda_{ki} \sum_{\tau_k} \left(\prod_{t'=0}^{t-1} (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t' + 1]) (1 - \lambda_{ki} \mathbb{1}[t' \geq \tau_k]) \right) m^{k \rightarrow i}(\tau_k | T) \mathbb{1}[t \geq \tau_k] \\ &\equiv -\epsilon_{ki} \phi_1^{k \rightarrow i}(t) - \lambda_{ki} \phi_2^{k \rightarrow i}(t), \end{aligned} \quad (3.18)$$

where, using the identity $\mathbb{1}[\tau_k \geq t+1] = \mathbb{1}[\tau_k \geq t] - \delta(\tau_k, t)$, we get for $\phi_1^{k \rightarrow i}(t)$

$$\begin{aligned} \phi_1^{k \rightarrow i}(t) &= \sum_{\tau_k} \left(\prod_{t'=0}^{t-2} (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t' + 1]) \right) (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t]) m^{k \rightarrow i}(\tau_k | T) \mathbb{1}[\tau_k \geq t+1] \\ &= (1 - \epsilon_{ki}) \phi_1^{k \rightarrow i}(t-1) - \left(\prod_{t'=0}^{t-1} (1 - \epsilon_{ki}) \right) m^{k \rightarrow i}(t | T) \\ &= (1 - \epsilon_{ki}) \phi_1^{k \rightarrow i}(t-1) - (1 - \epsilon_{ki})^t (P_S^{k \rightarrow i}(t-1) - P_S^{k \rightarrow i}(t)), \end{aligned} \quad (3.19)$$

and for $\phi_2^{k \rightarrow i}(t)$

$$\begin{aligned} \phi_2^{k \rightarrow i}(t) &= \sum_{\tau_k} \left(\prod_{t'=0}^{t-2} (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t' + 1]) (1 - \lambda_{ki} \mathbb{1}[t' \geq \tau_k]) \right) (1 - \epsilon_{ki} \mathbb{1}[\tau_k \geq t]) \\ &\quad \times (1 - \lambda_{ki} \mathbb{1}[t \geq \tau_k + 1]) m^{k \rightarrow i}(\tau_k | T) \mathbb{1}[t \geq \tau_k] \\ &= (1 - \lambda_{ki}) \phi_2^{k \rightarrow i}(t-1) + (1 - \epsilon_{ki})^t (P_S^{k \rightarrow i}(t-1) - P_S^{k \rightarrow i}(t)). \end{aligned} \quad (3.20)$$

As follows from their explicit mathematical definitions, the introduced dynamic variables can be given the physical interpretations, similar to (2.39)-(2.41):

- $\theta^{k \rightarrow i}(t)$ is the probability that neither of both ϵ and λ infection signals has been passed from node k to node i up to time t in the cavity dynamics D_i ;
- $\phi_1^{k \rightarrow i}(t)$ is the probability that the ϵ infection signal has not been passed from node k to node i up to time t in the cavity dynamics D_i and that k is in the state S at time t ;
- $\phi_2^{k \rightarrow i}(t)$ is the probability that neither of both ϵ and λ infection signals has been passed from node k to node i up to time t in the cavity dynamics D_i and that k is in the state I at time t ;

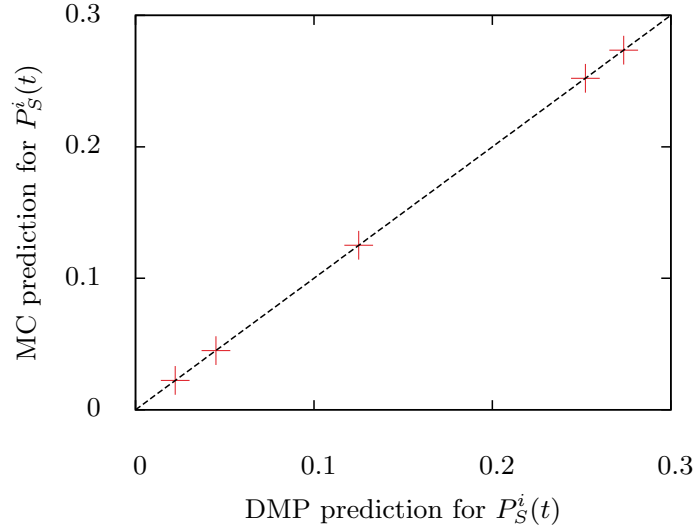


Figure 3.1: Comparison of prediction of the DMP equations for the generalized SI model with the Monte Carlo simulations. Marginal probabilities $P_S^i(t)$ are presented for five nodes from a tree graph with $N = 20$ nodes and $t = 5$, the parameters of the model are $\lambda = 0.5$, $\epsilon = 0.1$ and $\nu = 0.1$, there is one infected node at initial time. The MC average is performed over 10^7 instances. The error bars are negligible and are not shown.

- $P_S^{k \rightarrow i}(t)$ is the probability that k is in the state S at time t in the cavity dynamics D_i .

Let us recapitulate the DMP equations for the generalized SI model:

$$P_S^{k \rightarrow i}(t) = P_S^k(0)(1 - \epsilon_{ik})^t(1 - \nu_k)^t \prod_{l \in \partial k \setminus i} \theta^{l \rightarrow k}(t), \quad (3.21)$$

$$\theta^{k \rightarrow i}(t) = \theta^{k \rightarrow i}(t-1) - \epsilon_{ki} \phi_1^{k \rightarrow i}(t-1) - \lambda_{ki} \phi_2^{k \rightarrow i}(t-1), \quad (3.22)$$

$$\phi_1^{k \rightarrow i}(t) = (1 - \epsilon_{ki}) \phi_1^{k \rightarrow i}(t-1) - (1 - \epsilon_{ki})^t (P_S^{k \rightarrow i}(t-1) - P_S^{k \rightarrow i}(t)), \quad (3.23)$$

$$\phi_2^{k \rightarrow i}(t) = (1 - \lambda_{ki}) \phi_2^{k \rightarrow i}(t-1) + (1 - \epsilon_{ki})^t (P_S^{k \rightarrow i}(t-1) - P_S^{k \rightarrow i}(t)). \quad (3.24)$$

The initial conditions are

$$\theta^{i \rightarrow j}(0) = 1, \quad (3.25)$$

$$\phi_1^{i \rightarrow j}(0) = \delta_{\sigma_i^0, S} = P_S^i(0), \quad (3.26)$$

$$\phi_2^{i \rightarrow j}(0) = \delta_{\sigma_i^0, I} = P_I^i(0) = 1 - P_S^i(0). \quad (3.27)$$

Finally, the marginal probabilities for nodes to be in the states S or I at time t are computed via

$$P_S^i(t) = P_S^i(0)(1 - \nu_i)^t \prod_{k \in \partial i} \theta^{k \rightarrow i}(t), \quad (3.28)$$

$$P_I^i(t) = 1 - P_S^i(t). \quad (3.29)$$

The exactness of the DMP equations for the generalized SI model on tree graphs is demonstrated in the Fig. 3.1. Their computational complexity is $\mathcal{O}(Nct)$, where c is the average degree of the graph. These equations have never appeared in the literature. However, in some sense this model is a straightforward generalization of the SIR model that is considered further, see the next section.

3.2.2 Random field Ising model

The RFIM Hamiltonian reads

$$H = - \sum_{\langle ij \rangle} J_{ij} \sigma_i \sigma_j - \sum_i (h + h_i) \sigma_i, \quad (3.30)$$

where J_{ij} is a non-negative interaction between spins i and j , h is an external uniform magnetic field, and h_i is a random magnetic field on site i , extracted from some probability distribution. At zero temperature spin i tends to be aligned with its local magnetic field

$$\Delta_i^t = h + h_i + \sum_{k \in \partial i} J_{ki} \sigma_k^t. \quad (3.31)$$

Consider the initial condition in the form $P(\{\sigma_i^0\}_{i \in V}) = \prod_{i \in V} \delta_{\sigma_i^0, -1}$ (as it has already been mentioned, one could choose any initial condition, provided the dynamics is such that the transition from \uparrow to \downarrow is forbidden). Define the zero temperature stochastic dynamics respecting the following property: spin $\sigma_i = -1$ with a positive local field Δ_i flips with the probability $1/\tau$, and does not flip otherwise. Again, each spin flips at most only once, so the trajectory $\vec{\sigma}_i$ has a typical form $|\downarrow_0 \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \uparrow_{\tau_i} \uparrow \uparrow \uparrow \uparrow \uparrow \uparrow_T\rangle$ and is in the one-to-one correspondence with the flipping time τ_i (τ_i is the first time for which $\sigma_i = 1$). If the spin does not flip for all the times $0, \dots, T-1$, then by definition we set $\tau_i = T$ (T is the stopping time, i.e. the condition $\tau_i = T$ summarizes all the events that happen after the time T).

Using the representation in terms of flipping times, the equation (2.63) can be expressed as follows:

$$m^{i \rightarrow j}(\tau_i | \tau_j) = \sum_{\{\tau_k\}_{k \in \partial i \setminus j}} W_{RFIM} \prod_{k \in \partial i \setminus j} m^{k \rightarrow i}(\tau_k | \tau_i), \quad (3.32)$$

where

$$W_{RFIM} = \prod_{t'=0}^{\tau_i-2} \left(1 - \frac{1}{\tau} \mathbb{1}[\Delta_i^{t'} > 0] \right) \frac{1}{\tau} \mathbb{1}[\Delta_i^{\tau_i-1} > 0]. \quad (3.33)$$

From the conditional messages $m^{i \rightarrow j}(\tau_i | \tau_j)$, we can define two quantities:

$$p^{i \rightarrow j}(t) = \sum_{\tau_i > t} m^{i \rightarrow j}(\tau_i | T), \quad (3.34)$$

$$q^{i \rightarrow j}(t) = \sum_{\tau_i \leq t} m^{i \rightarrow j}(\tau_i | T). \quad (3.35)$$

These quantities characterize the marginals of the zero-temperature RFIM in the cavity dynamics \mathcal{D}_j , in which $\sigma_j^t = -1$ for every t and never flips, even if $\Delta_j > 0$. In this dynamics, $p^{i \rightarrow j}(t)$ is the probability that spin i stays in the state $\sigma_i^t = -1$ at time t , and $q^{i \rightarrow j}(t)$ is defined as the probability that spin i has already flipped, and hence $\sigma_i^t = 1$.

Of course, we will be ultimately interested in writing a closed equation for the marginals in the original dynamics:

$$p^i(t) = \sum_{\tau_i > t} m^i(\tau_i), \quad (3.36)$$

$$q^i(t) = 1 - p^i(t). \quad (3.37)$$

The derivation of the DMP equations for this model follows very closely the scheme used in the case of the SI model, for details see [P-1]. The resulting DMP equations can be shown to take the following form in discretized time notations:

$$q^{i \rightarrow j}(t+1) = \left(1 - \frac{1}{\tau}\right) q^{i \rightarrow j}(t) + \frac{1}{\tau} \sum_{\{\sigma_k\}_{k \in \partial i \setminus j}} \mathbb{1} \left[h + h_i + \sum_{k \in \partial i \setminus j} J_{ki} \sigma_k - J_{ji} > 0 \right] \\ \times \prod_{k \in \partial i \setminus j: \sigma_k = +1} q^{k \rightarrow i}(t) \prod_{k \in \partial i \setminus j: \sigma_k = -1} [1 - q^{k \rightarrow i}(t)]. \quad (3.38)$$

Therefore, the marginal probability for spin i to be in the state $+1$ at time $t+1$ is given by $q^i(t+1)$, that can be computed according to the following expression:

$$q^i(t+1) = \left(1 - \frac{1}{\tau}\right) q^i(t) + \frac{1}{\tau} \sum_{\{\sigma_k\}_{k \in \partial i}} \mathbb{1} \left[h + h_i + \sum_{k \in \partial i} J_{ki} \sigma_k > 0 \right] \\ \times \prod_{k \in \partial i: \sigma_k = +1} q^{k \rightarrow i}(t) \prod_{k \in \partial i: \sigma_k = -1} [1 - q^{k \rightarrow i}(t)]. \quad (3.39)$$

The probability that spin i is still in the state -1 at time $t+1$ is then given by $p^i(t+1) = 1 - q^i(t+1)$. Note that the DMP equations (3.38) and (3.39) can now be run in the real time, starting with initial conditions $q^i(0) = q^{i \rightarrow j}(0) = 0$ for each node i and j ; these equations have a closed self-consistent form, so we no longer need to compute the messages using (3.32). In the most straightforward implementation, the computational complexity of the DMP equations for the zero-temperature RFIM is $\mathcal{O}(N2^c t)$, where c is the average degree of the graph.

The averaged form of the DMP equations has been first derived in [OS10] using a cavity-like argument for the dynamic variables $q^{i \rightarrow j}(t)$ and $q^i(t)$. The derivation, which is close to ours, have been provided in [ABDZ13a], where an equivalent linear threshold model has been investigated. This model has been also studied in a different setting in the form of the voter model in [KM11].

3.3 Models with 3 states

3.3.1 SIR model

Note that the SIR model, defined by equations (2.28) and (2.29), represents in some sense a particular case of the generalized SI model, with $\alpha_{ij} = 0$ and $\nu_i = 0$ for all i and j . At the same time, a trivial (independent on the state of neighbors) transition to the R state with probability μ_i is added. The main difference with respect to the SI model is that now the time trajectory for a node i can be fully parametrized by two flipping times: $\vec{\sigma}_i = |S_0 SSSSSSI_{\tau_i} IIIIIIR_{\omega_i} RRRRRR_T\rangle \longleftrightarrow (\tau_i, \omega_i)$. Hence, the marginals of interest in the SIR model are defined as

$$P_S^i(t) = \sum_{\tau_i > t} \sum_{\omega_i > \tau_i} m^i(\tau_i, \omega_i), \quad (3.40)$$

$$P_I^i(t) = \sum_{\tau_i \leq t} \sum_{\omega_i > t} m^i(\tau_i, \omega_i), \quad (3.41)$$

$$P_R^i(t) = \sum_{\omega_i \leq t} \sum_{\tau_i < \omega_i} m^i(\tau_i, \omega_i). \quad (3.42)$$

We also define the marginal probability for the susceptible state in the corresponding cavity graph:

$$P_S^{i \rightarrow j}(t) = \sum_{\tau_i > t} \sum_{\omega_i > \tau_i} m^{i \rightarrow j}(\tau_i, \omega_i \mid T, T). \quad (3.43)$$

Let us point out the properties of the messages.

Property 1. $m^{i \rightarrow j}(\tau_i, \omega_i \mid T, T) = 0$ if $\tau_i \geq \omega_i$;

Property 2. If $\tau_j \geq \tau_i$, then $m^{i \rightarrow j}(\tau_i, \omega_i \mid \tau_j, \omega_j) = m^{i \rightarrow j}(\tau_i, \omega_i \mid t', \omega_j)$ for every $\tau_i \leq t' < \omega_j$;

Property 3. $\sum_{\tau_i, \omega_i} m^{i \rightarrow j}(\tau_i, \omega_i \mid T, T) = 1$;

Property 4. $m^{i \rightarrow j}(\tau_i, \omega_i + 1 \mid T, T) = (1 - \mu_i) m^{i \rightarrow j}(\tau_i, \omega_i \mid T, T)$.

The properties, equivalent to 1, 3 and 4, are also valid for the marginals $m^i(\tau_i, \omega_i)$. It is straightforward to establish first two evolution equations on the quantities $P_S^i(t)$, $P_I^i(t)$ and $P_R^i(t)$. According to the definitions,

$$\begin{aligned} P_R^i(t+1) &= \sum_{\omega_i \leq t+1} \sum_{\tau_i < \omega_i} m^i(\tau_i, \omega_i) = \sum_{\omega_i \leq t} \sum_{\tau_i < \omega_i} m^i(\tau_i, \omega_i) + \delta_{\omega_i, t+1} \sum_{\tau_i \leq t} m^i(\tau_i, \omega_i) \\ &= P_R^i(t) + \mu_i P_I^i(t), \end{aligned} \quad (3.44)$$

where we have used an equivalent of the property 4 for the marginals, because

$$\sum_{\omega_i \geq t+1} m^i(\tau_i, \omega_i) = \frac{1}{1 - (1 - \mu_i)} m^i(\tau_i, t+1) = \frac{1}{\mu_i} m^i(\tau_i, t+1). \quad (3.45)$$

Since the expressions defined in (3.40)-(3.42) sum to one, it is obvious that

$$P_I^i(t+1) = 1 - P_S^i(t+1) - P_R^i(t+1). \quad (3.46)$$

The rest of the derivation follows the SI case. Again, we show that we can put $P_S^{i \rightarrow j}(t+1)$ in the form

$$P_S^{i \rightarrow j}(t+1) = P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(t+1), \quad (3.47)$$

where $\theta^{k \rightarrow i}(t+1)$ (with a definition obviously adapted to the case of two flipping times) can be calculated via $P_S^{k \rightarrow i}(t)$ at each time step. This leads to the DMP equations (2.51)-(2.56), presented in the section 2.3, where their relation to those existing in the literature is also discussed. Here we note thereupon that the recent work [SM14] presented a generalization of the SIR model to the threshold models where a transition happens only if the information is received from a certain number A of neighbors. This model can also be readily solved within the DMP approach. Indeed, the expression (2.54) for the marginal probability that node i is in the state S would take a form similar to the second term in the right-hand side of the equation (3.39) in the RFIM, with $q^{k \rightarrow i}(t)$ replaced by $\theta^{k \rightarrow i}(t+1)$: one would need to sum over all the subsets of ∂i that correspond to the transmission of information by at least A neighbors. Hence, this model represents a three-state model with a RFIM-like non-trivial transition to the infected state.

Large-time limit

Let us discuss the large-time limit of the dynamic equations (2.51)-(2.56), also considered in the work [KN10], leading to the static percolation-like equations. We will use the following preliminary formulae. The solution of the differential equation

$$\frac{d}{dt}f(t) = -\alpha f(t) + g(t) \quad (3.48)$$

is given by

$$f(t) = e^{-\alpha t}f(0) + \int_0^t e^{\alpha(\tau-t)}g(\tau)d\tau. \quad (3.49)$$

Moreover, if we assume that the limit $\lim_{t \rightarrow +\infty} g(t) \equiv g(\infty)$ exists, then performing the integration by parts and using $\lim_{t \rightarrow +\infty} g'(t) = 0$, we get

$$\lim_{t \rightarrow +\infty} \int_0^t e^{\alpha(\tau-t)}g(\tau)d\tau = \frac{1}{\alpha}g(\infty). \quad (3.50)$$

In order to take the large-time limit in the SIR model, let us rewrite the DMP equations in the differential form:

$$\frac{d}{dt}\phi^{i \rightarrow j}(t) = -\mu_i \phi^{i \rightarrow j}(t) - \lambda_{ij} \phi^{i \rightarrow j}(t) - \frac{d}{dt}P_S^{i \rightarrow j}(t), \quad (3.51)$$

$$\frac{d}{dt}\theta^{i \rightarrow j}(t) = -\lambda_{ij} \theta^{i \rightarrow j}(t), \quad (3.52)$$

$$P_S^{i \rightarrow j}(t) = P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(t). \quad (3.53)$$

Solving the first equation, we get

$$\phi^{i \rightarrow j}(t) = e^{-(\lambda_{ij} + \mu_i)t} \left[\phi^{i \rightarrow j}(0) - \int_0^t e^{(\lambda_{ij} + \mu_i)\tau} \frac{d}{d\tau} P_S^{i \rightarrow j}(\tau) d\tau \right]. \quad (3.54)$$

Performing the integration by parts, we get

$$\begin{aligned} \phi^{i \rightarrow j}(t) &= e^{-(\lambda_{ij} + \mu_i)t} [\phi^{i \rightarrow j}(0) + P_S^{i \rightarrow j}(0)] - P_S^{i \rightarrow j}(t) \\ &\quad + (\lambda_{ij} + \mu_i) \int_0^t e^{(\lambda_{ij} + \mu_i)(\tau-t)} P_S^{i \rightarrow j}(\tau) d\tau. \end{aligned} \quad (3.55)$$

This expression can be further simplified by noticing that $\phi^{i \rightarrow j}(0) + P_S^{i \rightarrow j}(0) = \theta^{i \rightarrow j}(0) = 1$. Now solving the equation over $\theta^{i \rightarrow j}(t)$, we get

$$\theta^{i \rightarrow j}(t) = \theta^{i \rightarrow j}(0) - \lambda_{ij} \int_0^t \phi^{i \rightarrow j}(\tau) d\tau, \quad (3.56)$$

and hence

$$\begin{aligned} \theta^{i \rightarrow j}(t) &= 1 + \frac{\lambda_{ij}}{\lambda_{ij} + \mu_i} [e^{-(\lambda_{ij} + \mu_i)t} - 1] + \lambda_{ij} \int_0^t P_S^{i \rightarrow j}(\tau) d\tau \\ &\quad - \lambda_{ij}(\lambda_{ij} + \mu_i) \int_0^t e^{-(\lambda_{ij} + \mu_i)\tau} \int_0^\tau e^{(\lambda_{ij} + \mu_i)s} P_S^{i \rightarrow j}(s) ds. \end{aligned} \quad (3.57)$$

Again, performing the integration by parts in the last term, we get

$$\theta^{i \rightarrow j}(t) = 1 + \frac{\lambda_{ij}}{\lambda_{ij} + \mu_i} [e^{-(\lambda_{ij} + \mu_i)t} - 1] + \lambda_{ij} \int_0^t e^{(\lambda_{ij} + \mu_i)(\tau - t)} P_S^{i \rightarrow j}(\tau) d\tau. \quad (3.58)$$

Now, taking the large t limit in the last expression and assuming that $P_S^{i \rightarrow j}(t)$ has a large t limit, we get

$$\theta^{i \rightarrow j}(\infty) = 1 - \frac{\lambda_{ij}}{\lambda_{ij} + \mu_i} [1 - P_S^{i \rightarrow j}(\infty)] = 1 - \frac{\lambda_{ij}}{\lambda_{ij} + \mu_i} \left[1 - P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(\infty) \right]. \quad (3.59)$$

In order to demonstrate explicitly the physical sense of the corresponding terms, we can rewrite this expression as follows:

$$\theta^{i \rightarrow j}(\infty) = P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(\infty) + \frac{\mu_i}{\lambda_{ij} + \mu_i} \left[1 - P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(\infty) \right]. \quad (3.60)$$

The first term corresponds to the case where the infection has never been transmitted to node i in the cavity graph, and the second corresponds to receiving the infection, but recovering before transmitting it to j . The message-passing equations (3.60) have recently appeared in [KNZ14] (see also [KN10]) as an exact solution to the bond percolation problem on locally tree-like networks. This connection to percolation is not a surprise, since the correspondence between epidemic spreading and percolation has been widely used in the literature, see e.g. [New02]. In percolation, the bonds are independently occupied with some probability p , and one is usually interested in the resulting sizes of percolation clusters. In the same way, watching the epidemic outbreak in a network, one may mark links as “occupied” if the infection has been transmitted across it, which happens with the disease transition probability. The final size of the outbreak is then the size of the “percolation cluster” of vertices, reachable from the initially infected nodes. For regular random graphs these percolation-like equations can be iterated, providing the value $\theta(\infty)$ (the same for all links). In the general type of random graph with degree probability distribution $P(k)$, one can write a *density evolution* equation for the distribution of $\theta(\infty)$:

$$P(\theta) = \sum_k P(k) \int d\theta_1 \dots d\theta_k P(\theta_1) \dots P(\theta_k) \delta \left(\theta - \left[1 - \frac{\lambda_{ij}}{\lambda_{ij} + \mu_i} (1 - P_S^i(0) \theta_1 \dots \theta_k) \right] \right). \quad (3.61)$$

3.3.2 Rumor spreading model

The definition of the rumor spreading model can be summarized as follows [BLM⁺06]. For the sake of simplicity, we keep the same notations for the states, as in the SIR model. Each node $i \in V$ at discrete time t can be in one of three states σ_i^t : ignorant, $\sigma_i^t = S$, spreader, $\sigma_i^t = I$, or stifter, $\sigma_i^t = R$. At each time step, an “infected” node i will recover with probability $1 - \prod_{k \in \partial i} (1 - \alpha_{ki} \delta_{\sigma_k^t, I})$, and a “susceptible” node i will become infected with probability $1 - \prod_{k \in \partial i} (1 - \lambda_{ki} \delta_{\sigma_k^t, I})$, where ∂i is the set of neighbors of node i . The recovered nodes never change their state. These rules can be summarized by the following

scheme:

$$S(i) + I(j) \xrightarrow{\lambda_{ji}} I(i) + I(j) \quad (3.62)$$

$$I(i) + I(j) \xrightarrow{\alpha_{ji}} R(i) + I(j). \quad (3.63)$$

The interpretation of this model is as follows: a spreader node can either inform one of its ignorant neighbors on the rumor, in which case they start to communicate the rumor to their neighbors, or become uninterested in the rumor and turn to the R state if the rumor loses its “news value”. This happens in a directed way when the spreader gets in contact with another spreader. Note that some rumor spreading models include an additional modeling of such a spreading decay, described by a contact of a spreader with a stifer with the same probability α . This additional transition can be easily included in our approach, but for simplicity we stick to this “minimal” version of the ISS model that captures the main features of the rumor spreading process and its difference with respect to the epidemiological spreading models, such as the SIR model. As in the previous cases, the irreversibility of the dynamics of the rumor spreading model makes it possible to parametrize the time trajectory of a node by only two flipping times: $\vec{\sigma}_i = (\tau_i, \omega_i)$ for node i .

The rumor spreading model, defined via the transition rules (3.62)-(3.63), is notably more complicated than the SIR model because it has two non-trivial transitions, dependent on the state of neighbors. As we will see, it is not easy to obtain the corresponding DMP equations for this model by guessing the correct dynamic variables since the computation of the very DBP messages is required. On the other hand, they appear automatically in the dynamic cavity approach. Let us state the DMP computational scheme for this model.

The marginal probabilities $P_S^i(t+1)$, $P_I^i(t+1)$ and $P_R^i(t+1)$ that node i is in the state S , I and R respectively at time t are given by the following equations that can be iterated in time starting from the initial conditions at time $t=0$:

$$P_S^i(t+1) = P_S^i(0) \prod_{k \in \partial i} \theta^{k \rightarrow i}(t+1), \quad (3.64)$$

$$P_R^i(t+1) = P_R^i(t) + \sum_{\tau_i \leq t} m^i(\tau_i, t+1), \quad (3.65)$$

$$P_I^i(t+1) = 1 - P_S^i(t+1) - P_R^i(t+1), \quad (3.66)$$

where $m^i(\tau_i, t+1)$ has the physical meaning of the marginal probability that node i has switched to the state I at time τ_i and to the state R at time $t+1$. The remaining computational scheme serves to compute this probabilities explicitly. To this purpose, we introduce a number of auxiliary dynamic messages that can be computed iteratively. Again, these messages may be given a physical interpretation, and are defined in the corresponding cavity dynamics. As an illustration, consider the message $P_S^{i \rightarrow j}(t)$, defined as the probability for node i to be in the state S at time t in the cavity graph, in which all the connections of node j , except to i , has been removed. It is updated as follows:

$$P_S^{i \rightarrow j}(t+1) = P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(t+1), \quad (3.67)$$

$$\theta^{k \rightarrow i}(t+1) - \theta^{k \rightarrow i}(t) = -\lambda_{ki} \phi^{k \rightarrow i}(t), \quad (3.68)$$

$$\begin{aligned} \phi^{k \rightarrow i}(t) &= (1 - \lambda_{ki}) \phi^{k \rightarrow i}(t-1) + P_S^{k \rightarrow i}(t-1) \\ &\quad - P_S^{k \rightarrow i}(t) - \sum_{\tau_k \leq t-1} (1 - \lambda_{ki})^{t-\tau_k} m^{k \rightarrow i}(\tau_k, t \mid T, T). \end{aligned} \quad (3.69)$$

In these equations, the dynamic messages $\theta^{k \rightarrow i}(t+1)$, $\phi^{k \rightarrow i}(t)$ and $m^{k \rightarrow i}(\tau_k, t \mid T, T)$ have the following physical sense (for precise mathematical expressions, refer to [P-1]):

- $\theta^{k \rightarrow i}(t+1)$ is the probability that the infection signal λ has not been passed from node k to node i up to time $t+1$ in the cavity dynamics D_i ;
- $\phi^{k \rightarrow i}(t)$ is the probability that the infection signal λ has not been passed from node k to node i up to time t in the cavity dynamics D_i and that k is in the state I at time t ;
- $m^{k \rightarrow i}(\tau_k, t \mid T, T)$ is the marginal probability that node k has the trajectory (τ_k, t) in the cavity dynamics D_i .

Hence, the last term in the equation (3.69) represents a contribution to the change of $\phi^{k \rightarrow i}(t)$ due to the recovery of node k exactly at time t in the cavity dynamics D_i . The initial conditions are given by $\theta^{k \rightarrow i}(0) = 1$, and $\phi^{k \rightarrow i}(0) = \delta_{\sigma_k^0, I}$.

So far, the equations (3.67)-(3.69) are not in a closed form, we still need to know how to compute $m^{i \rightarrow j}(\tau_i, t \mid T, T)$ for $\tau_i < t$. We have for each t

$$m^{i \rightarrow j}(0, t \mid \tau_j, T) = P_I^i(0) \left[f_{\rho_1, \chi_1}^{i \rightarrow j}(-2, t-2 \mid \tau_j) - f_{\rho_1, \chi_1}^{i \rightarrow j}(-2, t-1 \mid \tau_j) \right], \quad (3.70)$$

$$\begin{aligned} m^{i \rightarrow j}(\tau_i, t \mid \tau_j, T) &= P_S^i(0) \left[f_{\rho_1, \chi_1}^{i \rightarrow j}(\tau_i - 2, t-2 \mid \tau_j) - f_{\rho_1, \chi_1}^{i \rightarrow j}(\tau_i - 2, t-1 \mid \tau_j) \right. \\ &\quad \left. - f_{\rho_2, \chi_2}^{i \rightarrow j}(\tau_i - 1, t-2 \mid \tau_j) + f_{\rho_2, \chi_2}^{i \rightarrow j}(\tau_i - 1, t-1 \mid \tau_j) \right], \end{aligned} \quad (3.71)$$

for $1 \leq \tau_i \leq t-1$ and $0 \leq \tau_j \leq t$. The functional $f_{\rho, \chi}^{i \rightarrow j}(t_1, t_2 \mid \tau_j)$ is defined as follows:

$$f_{\rho, \chi}^{i \rightarrow j}(t_1, t_2 \mid \tau_j) = \rho^{j \rightarrow i}(t_1, t_2 \mid \tau_j) \prod_{k \in \partial i \setminus j} \chi^{k \rightarrow i}(t_1, t_2), \quad (3.72)$$

where the τ_j -dependent coefficients, characterizing the influence of node j on the dynamics of i , read for $t_2 = t-2$ or $t_2 = t-1$:

$$\rho_1^{j \rightarrow i}(\tau_i - 2, t_2 \mid \tau_j) = (1 - \lambda_{ji})^{\tau_i - \tau_j - 1} \left(\prod_{t'=\tau_i}^{t_2} (1 - \alpha_{ji} \mathbb{1}[\tau_j \leq t']) \right), \quad (3.73)$$

$$\rho_2^{j \rightarrow i}(\tau_i - 1, t_2 \mid \tau_j) = (1 - \lambda_{ji})^{\tau_i - \tau_j} \left(\prod_{t'=\tau_i}^{t_2} (1 - \alpha_{ji} \mathbb{1}[\tau_j \leq t']) \right). \quad (3.74)$$

Let us remind at this point, that the convention (3.8) is used here and in the following. Note that in the update equation (3.69) we are only interested in the messages of the

form $m^{i \rightarrow j}(\tau_i, t \mid T, T)$ that correspond to $\tau_j = t$, and for which the j -influence is not present: in this case $\rho_1^{j \rightarrow i}(\tau_i - 2, t_2 \mid \tau_j)$ and $\rho_2^{j \rightarrow i}(\tau_i - 1, t_2 \mid \tau_j)$ are simply equal to one. Still, in the computation scheme for $\chi_1^{k \rightarrow i}(\tau_i - 2, t_1)$ and $\chi_2^{k \rightarrow i}(\tau_i - 1, t_1)$ all the values $0 \leq \tau_j \leq t - 1$ are also required, since the remaining update equations read

$$\chi_1^{k \rightarrow i}(\tau_i - 2, t - 1) = \chi_1^{k \rightarrow i}(\tau_i - 2, t - 2) - \alpha_{ki} \psi_1^{k \rightarrow i}(\tau_i - 2, t - 1), \quad (3.75)$$

$$\chi_2^{k \rightarrow i}(\tau_i - 1, t - 1) = \chi_2^{k \rightarrow i}(\tau_i - 1, t - 2) - \alpha_{ki} \psi_2^{k \rightarrow i}(\tau_i - 1, t - 1), \quad (3.76)$$

and

$$\begin{aligned} \psi_1^{k \rightarrow i}(\tau_i - 2, t - 1) &= P_S^{k \rightarrow i}(t - 2 \mid \tau_i) - P_S^{k \rightarrow i}(t - 1 \mid \tau_i) + (1 - \alpha_{ki} \mathbb{1}_{\tau_i \neq t-1}) \psi_1^{k \rightarrow i}(\tau_i - 2, t - 2) \\ &\quad - \sum_{\tau_k \leq t-2} (1 - \lambda_{ki})^{\tau_i - \tau_k - 1} \left(\prod_{t'=\tau_k}^{t-2} (1 - \alpha_{ki} \mathbb{1}[\tau_k \leq t']) \right) m^{k \rightarrow i}(\tau_k, t - 1 \mid \tau_i, T), \end{aligned} \quad (3.77)$$

$$\begin{aligned} \psi_2^{k \rightarrow i}(\tau_i - 1, t - 1) &= P_S^{k \rightarrow i}(t - 2 \mid \tau_i) - P_S^{k \rightarrow i}(t - 1 \mid \tau_i) + (1 - \alpha_{ki} \mathbb{1}_{\tau_i \neq t-1}) \psi_2^{k \rightarrow i}(\tau_i - 1, t - 2) \\ &\quad - \sum_{\tau_k \leq t-2} (1 - \lambda_{ki})^{\tau_i - \tau_k} \left(\prod_{t'=\tau_k}^{t-2} (1 - \alpha_{ki} \mathbb{1}[\tau_k \leq t']) \right) m^{k \rightarrow i}(\tau_k, t - 1 \mid \tau_i, T). \end{aligned} \quad (3.78)$$

The conditional quantity $P_S^{k \rightarrow i}(t_1 \mid \tau_i)$ is defined as

$$P_S^{k \rightarrow i}(t_1 \mid \tau_i) = P_S^k(0) (1 - \lambda_{ik})^{t_1 - \tau_i} \prod_{l \in \partial k \setminus i} \theta^{l \rightarrow k}(t_1). \quad (3.79)$$

The necessary initial conditions are given by $\chi_1^{k \rightarrow i}(-2, -1) = 1$ and $\psi_1^{k \rightarrow i}(-2, 0) = \phi^{k \rightarrow i}(0)$. The following border conditions are used for $\tau_i = t - 1$:

$$\chi_1^{k \rightarrow i}(t - 3, t - 2) = \theta^{k \rightarrow i}(t - 2), \quad (3.80)$$

$$\chi_2^{k \rightarrow i}(t - 2, t - 2) = \theta^{k \rightarrow i}(t - 1), \quad (3.81)$$

$$\psi_1^{k \rightarrow i}(t - 3, t - 2) = \phi^{k \rightarrow i}(t - 2), \quad (3.82)$$

$$\psi_2^{k \rightarrow i}(t - 2, t - 2) = (1 - \lambda_{ki}) \phi^{k \rightarrow i}(t - 2). \quad (3.83)$$

and $\chi_1^{k \rightarrow i}(t - 3, t - 1)$, $\chi_2^{k \rightarrow i}(t - 2, t - 1)$, $\psi_1^{k \rightarrow i}(t - 3, t - 1)$, $\psi_2^{k \rightarrow i}(t - 2, t - 1)$ follow the equations (3.75)-(3.78).

Therefore, the computation of $m^{i \rightarrow j}(\tau_i, t \mid T, T)$ for $\tau_i < t$ involves messages $m^{i \rightarrow j}(\tau_i, t - 1 \mid \tau_j, T)$ for $\tau_i < t - 1$ and $\tau_j \leq t - 1$, computed at a previous step. Finally, the marginal probabilities $m^i(\tau_i, t + 1)$ are computed via equations (3.70) and (3.71), with replacement of the indices $i \rightarrow j$ simply by i , and the corresponding change of product over $k \in \partial i \setminus j$ in the definition (3.72) by the product over all the neighboring nodes $k \in \partial i$. The computational complexity of DMP equations for rumor spreading model is $\mathcal{O}(Nct^3)$, where c is the average degree of the graph. The details of the derivation are presented in [P-1].

The validity of these equations has been checked numerically via comparison with the Monte Carlo (MC) simulation: as expected, the marginals given by the DMP equations appear to be exact on any tree graph. Although *a priori* the DMP equations are not guaranteed to be exact on networks that do not have a locally tree-like structure, they

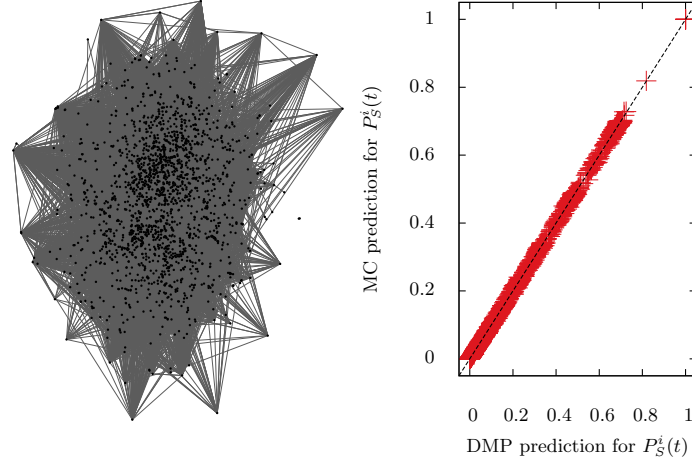


Figure 3.2: Right: Comparison of prediction of the DMP equations for the rumor spreading model with the Monte Carlo (MC) simulations. Marginal probabilities $P_S^i(t)$ are presented for the Facebook-like social network with $N = 1899$ nodes and $t = 10$, the parameters of the model are $\lambda = 0.3$, $\alpha = 0.2$, there is one infected node at initial time. The MC average is performed over 10^4 instances. The error bars are smaller than the symbol size on the plots and are not shown. Left: A representation of the topology of the network, generated with Gephi [BHJ09]. The high-degree nodes (hubs) are placed on the periphery.

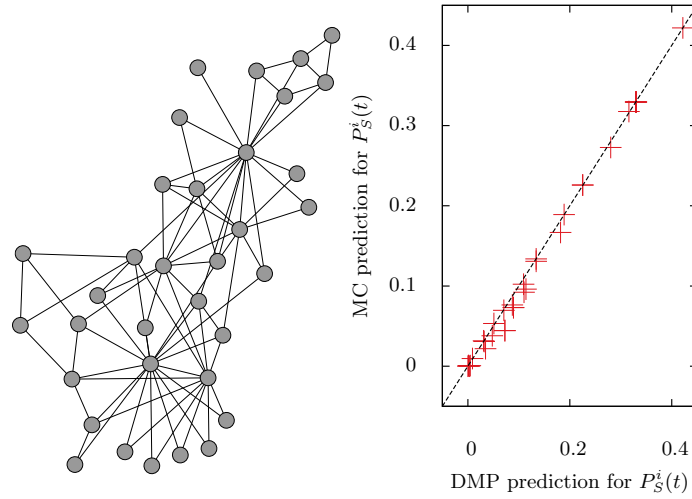


Figure 3.3: Right: Comparison of prediction of the DMP equations for the rumor spreading model with the Monte Carlo (MC) simulations. Marginal probabilities $P_S^i(t)$ are presented for the Zachary's carate club network with $N = 34$ nodes and $t = 10$, the parameters of the model are $\lambda = 0.3$, $\alpha = 0.2$, there is one infected node at initial time. The MC average is performed over 10^4 instances. The error bars are smaller than the symbol size on the plots and are not shown. Left: A representation of the topology of the network, generated with Gephi [BHJ09]. This network has a block structure and contains many loops of small length.

provide remarkably accurate predictions even for small and loopy networks. For example, we have tested the performance of the DMP equations for the rumor spreading model on two real-world networks. The first example is a Facebook-like social network with 1899 nodes and 20296 edges that represents an online community for students at University of

California, Irvine [OP09]; the predictions for the marginals given by DMP are compared with the values obtained from 10^4 MC simulations, see Fig. 3.2. Another test has been performed for the small Zachary's karate club network of friendships between 34 members of a karate club at a US university [Zac77], the results are presented in the Fig. 3.3. In both cases, the predictions of the DMP equations appear to be very accurate with respect to the true values of the marginals.

The DMP equations for the rumor spreading model have never been reported so far. The existing approaches include the naive mean-field equations that are derived under the complete-mixing assumption and completely neglect the topology of the network, or the so-called heterogeneous mean-field equations that assume equivalent behavior for different nodes of the same degree. These equations are averaged over the ensemble of random graphs and are not applicable on a single instance of the network; cf. section 2.3 for a discussion of these mean-field techniques.

Large-time limit

Let us discuss what happens in the large-time limit of the DMP equations for the rumor spreading model. We start with the corresponding differential form:

$$\frac{d}{dt}\phi^{i \rightarrow j}(t) = -\lambda_{ij}\phi^{i \rightarrow j}(t) - \frac{d}{dt}P_S^{i \rightarrow j}(t) - f^{i \rightarrow j}(t), \quad (3.84)$$

$$\frac{d}{dt}\theta^{i \rightarrow j}(t) = -\lambda_{ij}\phi^{i \rightarrow j}(t), \quad (3.85)$$

$$P_S^{i \rightarrow j}(t) = P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(t). \quad (3.86)$$

where

$$f^{i \rightarrow j}(t) = \int_0^{t-1} (1 - \lambda_{ij})^{t-1-\tau} m^{i \rightarrow j}(\tau, t \mid T, T) d\tau. \quad (3.87)$$

Proceeding as in the DMP equations for the SIR model, we would like to close the equation on the variable $\theta^{i \rightarrow j}(t)$, using

$$\theta^{i \rightarrow j}(t) = \theta^{i \rightarrow j}(0) - \lambda_{ij} \int_0^t \phi^{i \rightarrow j}(\tau) d\tau. \quad (3.88)$$

Hence, we first solve the equation for $\phi^{i \rightarrow j}(t)$, obtaining

$$\phi^{i \rightarrow j}(t) = e^{-\lambda_{ij}t} \left[\phi^{i \rightarrow j}(0) - \int_0^t e^{\lambda_{ij}\tau} \left(\frac{d}{d\tau} P_S^{i \rightarrow j}(\tau) + f^{i \rightarrow j}(\tau) \right) d\tau \right]. \quad (3.89)$$

Performing the integration by parts, we get

$$\phi^{i \rightarrow j}(t) = e^{-\lambda_{ij}t} - P_S^{i \rightarrow j}(t) + \lambda_{ij} \int_0^t e^{\lambda_{ij}(\tau-t)} P_S^{i \rightarrow j}(\tau) d\tau - \int_0^t e^{\lambda_{ij}(\tau-t)} f^{i \rightarrow j}(\tau) d\tau. \quad (3.90)$$

Putting this expression in (3.88) gives

$$\theta^{i \rightarrow j}(t) = e^{-\lambda_{ij}t} + \lambda_{ij} \int_0^t e^{\lambda_{ij}(\tau-t)} P_S^{i \rightarrow j}(\tau) d\tau + \int_0^t [1 - e^{\lambda_{ij}(\tau-t)}] f^{i \rightarrow j}(\tau) d\tau. \quad (3.91)$$

Now we take the large t limit in the last expression, assuming that $P_S^{i \rightarrow j}(t)$ and $f^{i \rightarrow j}(t)$ have a large t limit:

$$\theta^{i \rightarrow j}(\infty) = P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(\infty) - \frac{1}{\lambda_{ij}} f^{i \rightarrow j}(\infty) + \lim_{t \rightarrow +\infty} \int_0^t f^{i \rightarrow j}(\tau) d\tau. \quad (3.92)$$

At the same time, if $P_S^{i \rightarrow j}(t)$ has a limit, then directly from (3.85) we get $\phi^{i \rightarrow j}(\infty) = 0$. Hence, on one hand, from (3.90) we obtain

$$\phi^{i \rightarrow j}(\infty) = -\frac{1}{\lambda_{ij}} f^{i \rightarrow j}(\infty) = 0, \quad (3.93)$$

and on the other, since

$$\begin{aligned} f^{i \rightarrow j}(\infty) &= \lim_{t \rightarrow +\infty} \int_0^{t-1} e^{\gamma_{ij}(\tau-t+1)} m^{i \rightarrow j}(\tau, t \mid T, T) d\tau \\ &= \frac{1}{\gamma_{ij}} \lim_{t \rightarrow +\infty} m^{i \rightarrow j}(t-1, t \mid T, T), \end{aligned} \quad (3.94)$$

where $\gamma_{ij} = -\log(1 - \lambda_{ij})$, we get $\lim_{t \rightarrow +\infty} m^{i \rightarrow j}(t-1, t \mid T, T) = 0$. Hence, we finally get

$$\theta^{i \rightarrow j}(\infty) = P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(\infty) + \lim_{t \rightarrow +\infty} \int_0^t f^{i \rightarrow j}(\tau) d\tau. \quad (3.95)$$

In this expression, again, the first term corresponds to the case where the rumor has never been transmitted to the node i in the cavity graph, and the second corresponds to receiving the rumor, but recovering before transmitting it to j . Now we need to study only the large t limit of $\int_0^t f^{i \rightarrow j}(\tau) d\tau$. Note that this limit is trivial in the case $\lambda_{ij} \rightarrow 0$:

$$\begin{aligned} \lim_{t \rightarrow +\infty} \int_0^t f^{i \rightarrow j}(\tau) d\tau &\leq \lim_{t \rightarrow +\infty} \int_0^t \lim_{\lambda_{ij} \rightarrow 0} f^{i \rightarrow j}(\tau) d\tau \\ &= \lim_{t \rightarrow +\infty} \int_0^t dt' \int_0^{t'-1} m^{i \rightarrow j}(\tau, t' \mid T, T) d\tau = 1. \end{aligned} \quad (3.96)$$

Using the integration by parts, we get

$$\lim_{t \rightarrow +\infty} \int_0^t f^{i \rightarrow j}(\tau) d\tau = \frac{1}{\gamma_{ij}} \lim_{t \rightarrow +\infty} \left[\int_0^t m^{i \rightarrow j}(\tau-1, \tau \mid T, T) d\tau - f^{i \rightarrow j}(t) \right], \quad (3.97)$$

or, returning to the discrete form,

$$\lim_{t \rightarrow +\infty} \int_0^t f^{i \rightarrow j}(\tau) d\tau = \frac{1}{\gamma_{ij}} \lim_{t \rightarrow +\infty} \sum_{\tau=1}^t m^{i \rightarrow j}(\tau-1, \tau \mid T, T). \quad (3.98)$$

Explicitely, from the DMP equations, we get

$$\begin{aligned} &\lim_{t \rightarrow +\infty} \sum_{\tau=0}^t m^{i \rightarrow j}(\tau, \tau+1 \mid T, T) \\ &= 1 - \prod_{k \in \partial i \setminus j} [1 - \alpha_{ki} \phi^{k \rightarrow i}(0)] + P_S^{i \rightarrow j}(0) \lim_{t \rightarrow +\infty} \sum_{\tau=0}^t \left(\prod_{k \in \partial i \setminus j} [\theta^{k \rightarrow i}(\tau) - \alpha_{ki} \phi^{k \rightarrow i}(\tau)] \right. \\ &\quad \left. - \prod_{k \in \partial i \setminus j} \left[\theta^{k \rightarrow i}(\tau) - \frac{\alpha_{ki}}{1 - \lambda_{ki}} (\phi^{k \rightarrow i}(\tau+1) - \lambda_{ki} [P_S^{k \rightarrow i}(\tau) - P_S^{k \rightarrow i}(\tau+1)]) \right] \right). \end{aligned} \quad (3.99)$$

We see that although $m^{i \rightarrow j}(\tau, \tau + 1 \mid T, T)$ is a decreasing function in τ , we have to solve the dynamics up to some time t_* to have a lower-bounded estimate for the expression (3.95). This approximation can be controlled by fixing some threshold ε , and defining t_* as the first time for which $m^{i \rightarrow j}(t_*, t_* + 1 \mid T, T) < \varepsilon$. Therefore, a lower bound to $\theta^{i \rightarrow j}(\infty)$ can be found with a complexity proportional to $\mathcal{O}(t_*^3(\varepsilon))$. An exact limit (as it has been done for the SIR model) is, however, not accessible, due to the intrinsic properties of the microscopic dynamics of the rumor spreading model.

3.4 Models with larger number of states

Following the general derivation procedure, the DMP equations for the rumor spreading model, described in the previous section, can be easily generalized to a more complicated pairwise model with arbitrary number of states, similar to the generalized SI model. The models with $K \geq 3$ states that include direct transitions that skip some number of intermediate states can also be taken into account in this approach. The procedure for deriving these equations from the general dynamic cavity equation (2.63) is very similar to the derivation of the DMP equations for the rumor spreading model. In [P-1], the method is illustrated using a “minimal” model with $K = 4$ states, which is an extension of the rumor spreading model with an additional non-trivial transition to the final state. Let us consider four states S , I_1 , I_2 and R , and the following dynamic rules:

$$S(i) + I_1(j) \xrightarrow{\lambda_{ji}} I_1(i) + I_1(j), \quad (3.100)$$

$$I_1(i) + I_1(j) \xrightarrow{\alpha_{ji}} I_2(i) + I_1(j), \quad (3.101)$$

$$I_2(i) + I_2(j) \xrightarrow{\beta_{ji}} R(i) + I_2(j). \quad (3.102)$$

Now the time trajectory of node i can be parametrized by three flipping times: τ_i (first time in I_1), ω_i (first time in I_2) and ε_i (first time in R). The trajectory of spin i is hence described by $\vec{\sigma}_i(t) = (\tau_i, \omega_i, \varepsilon_i)$, and the corresponding marginal of the dynamic cavity equation (2.65) could be written as $m^i(\tau_i, \omega_i, \varepsilon_i)$. Similarly to the SI, the SIR and the rumor spreading model, we might expect that the expressions for the marginal probabilities at time t could be written in the following form:

$$P_S^i(t+1) = P_S^i(0) \prod_{k \in \partial i} \theta^{k \rightarrow i}(t+1), \quad (3.103)$$

$$P_R^i(t+1) = P_R^i(t) + \sum_{\substack{\omega_i \leq t \\ \tau_i + 1 \leq \omega_i}} m^i(\tau_i, \omega_i, t+1), \quad (3.104)$$

$$P_{I_2}^i(t+1) = P_{I_2}^i(t) + \sum_{\substack{\tau_i \leq t \\ \varepsilon_i > t+1}} m^i(\tau_i, t+1, \varepsilon_i), \quad (3.105)$$

$$P_{I_1}^i(t+1) = 1 - P_S^i(t+1) - P_{I_2}^i(t+1) - P_R^i(t+1). \quad (3.106)$$

The apparent difficulty in the equation (3.105) is that the sum runs over all the flipping times $\varepsilon_i > t+1$, and the number of terms can potentially be very big, of the order of the stopping time T . In [P-1] it is shown that this difficulty can be overcome if one defines a new sort of messages:

$$\mu^{k \rightarrow i}(\tau_k, t \mid T, T) = \sum_{\varepsilon_k \geq t+1} m^{k \rightarrow i}(\tau_k, t, \varepsilon_k \mid T, T, T). \quad (3.107)$$

The evolution of $\mu^{k \rightarrow i}(\tau_k, t \mid T, T)$ follows the same equations as for the rumor spreading model (3.70)-(3.83), except that now we will require the computation of $\mu^{k \rightarrow i}(\tau_k, t \mid \tau_i, \omega_i)$. The details of derivation for this case are presented in [P-1].

The variants of this model may describe different models with 4 states, for example the generalization of the SIR model that include immunized or exposed states [AM91]. As it has been expected, the computational complexity for this $K = 4$ model is higher: $\mathcal{O}(Nct^5)$, where c is the average degree of the graph. For a general model with unidirectional dynamics and M non-trivial transitions, the computational complexity grows as t^{2M-1} . Note that M is not always equal to $K - 1$, for instance, compare the SIR ($K = 3$, $M = 1$) and the rumor spreading ($K = 3$, $M = 2$) models. The growth of the number of operations with the number of states of the model in the DMP equations is essentially due to the local effects of retro-action that have to be taken into account. Nevertheless, the DMP approach opens a way to a number of applications aimed at a better control of the cascading processes on networks. The fact that the DMP equations can be applied to a single instance of a graph has been recently used for the algorithmic application to an inverse problem in the context of epidemic spreading: the inference of the origin of an epidemic outbreak [P-2]. This application is presented in the next chapter, 4.

Chapter 4

Application: patient zero problem

In this chapter, we address an application of the DMP equations for the SIR model to the problem of estimation of the origin of an epidemic outbreak (the so-called patient zero): given a contact network and a snapshot of epidemic spread at a certain time, determine the infection source. The DMP approach is in the core of the inference algorithm that leads to a significant improvement of performance compared to previous approaches. The results presented in this chapter are published in [P-2].

4.1 Context of the inference problem

Whereas the dynamics and the prediction of epidemic spreading in networks have attracted a considerable number of works, the problem of estimating the epidemic origin has been mathematically formulated only recently [SZ10], followed by a burst of research on this practically important problem [Cd11, SZ11, PVF12, FC12, PTV12, ZY13, DZT13, BH13, ABD⁺14b]. In order to make the estimation of the origin of spreading a well-defined problem we need to have some knowledge about the spreading mechanism. We shall adopt here the same framework as in existing works, namely we assume that the epidemic spread follows the susceptible-infected-recovered model, discussed in the previous chapters. Note, however, that depending on a particular area of application, the use of a different avalanche model (such as the rumor spreading model) can be required. The DMP-based method, discussed in this chapter, can be straightforwardly adapted to such a setting. This remark justifies the name *dynamic message-passing algorithm* for the method.

The stochastic nature of infection propagation makes the estimation of the epidemic origin intrinsically hard: indeed, different initial conditions can lead to the same configuration at the observation time. Finding an estimator that locates the most probable origin, given observed configuration, is in general computationally intractable, except in very special cases such as the case where the contact network is a line or a regular tree [SZ10, SZ11, DZT13]. The methods that have been studied in the existing works are mostly based on various kinds of graph-centrality measures. Examples include the distance centrality or the Jordan center of a graph [SZ10, Cd11, SZ11, ZY13]. The problem was generalized to estimating a set of epidemic origins using spectral methods in [PVF12, FC12]. Another line of approach uses more detailed information about the epidemic than just a snapshot at a given time [PTV12]. The DMP algorithm, presented in the next section, estimates the probability that the observed snapshot resulted from a given patient zero in a way which is crucially different from the previous approaches: for

every possible origin of the epidemic, we use the DMP equations to estimate the probability that a given node in the network was in the observed state (S , I or R). We then use a mean-field-like approximation to compute the probability of the observed snapshot as a product of the marginal probabilities. We finally rank the possible origins according to that probability.

4.2 DMP algorithm

To define the problem of estimation of the epidemic origin, we consider the case where, at initial time $t = 0$, only one node is infected (the *patient zero*, i_0), and all other nodes are susceptible. After $t_0 > 0$ time steps (t_0 is in general unknown), we observe the state of a set of nodes $\mathcal{O} \subset V$, and the task is to estimate the location of the patient zero based on this snapshot, see Fig. 4.1.

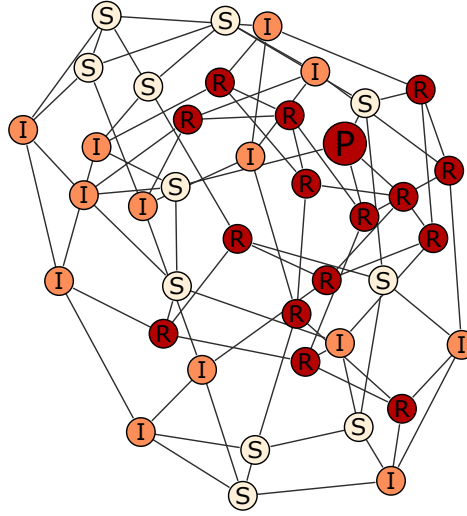


Figure 4.1: An example of a single instance of the inference problem on a random regular graph of degree $c = 4$ with $N = 40$ nodes. The patient zero is labeled by P and appears in the state R in the snapshot. The epidemic is generated for $\lambda = 0.5$ and $\mu = 0.5$, the snapshot is represented at time $t_0 = 5$.

Let us briefly explain two existing algorithms [SZ10, SZ11, ZY13] that we will use as benchmarks. The authors of [SZ10, SZ11, ZY13] considered only the case when all the nodes were observed, $\mathcal{O} = V$. In [P-2] we propose a generalization of these algorithms to a more general case. The most basic measure for node i to be the epidemic origin is the *distance centrality*, $D(i)$, which we define as

$$D(i) \equiv \sum_{j \in \mathcal{G}} d(i, j) (\delta_{\sigma_j(t_0), I} + \delta_{\sigma_j(t_0), R} / \mu_j), \quad (4.1)$$

where the graph \mathcal{G} is a connected component of the original graph G containing all infected and recovered nodes and only them, and $d(i, j)$ is the shortest path between node i and node j on the graph \mathcal{G} . The ad-hoc factor $1/\mu_j$ is introduced to distinguish recovered nodes that for small μ_j tend to be closer to the epidemic origin. In the existing works this factor was not present, because [SZ10, SZ11] treated only the SI model, and [ZY13] considered that susceptible and recovered nodes are indistinguishable. The

authors of [SZ10, SZ11] suggested the *rumor centrality* estimator and showed that for tree graphs, the rumor centrality and the distance centrality coincide. Another simple but well-performing estimator, Jordan centrality $J(i)$, was proposed in [ZY13] and corresponds to a node minimizing the maximum distance to other infected and recovered nodes:

$$J(i) \equiv \max_{j \in \mathcal{G}} d(i, j). \quad (4.2)$$

The node i for which $J(i)$ is minimal is known as the *Jordan center* of \mathcal{G} in the graph theory literature. Note that in [ZY13] the Jordan center was chosen only among the infected nodes, hence this implementation uses more information.

The core of the proposed algorithm is the DMP equations for the SIR model (2.51)-(2.56), which provides an estimate of the probabilities $P_S^j(t, i_0)$ (respectively $P_I^j(t, i_0)$, $P_R^j(t, i_0)$) that node j is in each of the three states S , I , or R , at time t , for a given patient zero i_0 . Let us first assume that the time t_0 is known. With the use of Bayes' rule, the probability that node i is the patient zero given the observed states is proportional to the joint probability of observed states given the patient zero, $P(i|\mathcal{O}) \sim P(\mathcal{O}|i)$. We also define an energy-like function of every node $E(i) \equiv -\log P(\mathcal{O}|i)$, such that nodes with lower energy are more likely to be the infection source. If one were able to compute $P(\mathcal{O}|i)$ exactly, finding i which minimizes $E(i)$ would be the optimal inference scheme for the patient zero problem. As there is no tractable way to compute exactly the joint probability of the observations, we approximate it using a mean-field-type approach as a product of the marginal probabilities provided by the dynamic message-passing

$$P(\mathcal{O}|i) \simeq \prod_{\substack{k \in \mathcal{O} \\ \sigma_k(t_0)=S}} P_S^k(t, i) \prod_{\substack{l \in \mathcal{O} \\ \sigma_l(t_0)=I}} P_I^l(t, i) \prod_{\substack{n \in \mathcal{O} \\ \sigma_n(t_0)=R}} P_R^n(t, i). \quad (4.3)$$

To estimate the value of t_0 , we compute the energy $E(i, t)$ for different possible values t , and choose the value that maximizes the “partition function” $Z(t) \equiv \sum_i e^{-E(i, t)}$. As mentioned previously, the algorithmic complexity for computing the energy $E(i)$ of a given vertex i (and therefore the probability that it is the epidemic origin) is $\mathcal{O}(t_0 N c)$, where c is the average degree of the graph.

4.3 Results and discussion

Let us briefly present the results that are obtained with the DMP algorithm. We first test our algorithm on random regular graphs (2.25), i.e. random graphs drawn uniformly from the set of graphs where every node has degree c . In all the simulations we consider uniform transmission and recovery probabilities $\lambda_{ij} = \lambda$ and $\mu_i = \mu$.

In the first illustrative example, inset of Fig. 4.2, we plot the energies $E(i)$ of the nodes for which the probability of being the epidemic origin is finite according to the DMP algorithm; the nodes are ordered according to the energy values. The true epidemic origin is marked with a red cross. We define the rank of candidates for the epidemic origin to be its position in this ranking (the lowest energy node having rank 0). The main graph of Fig. 4.2 shows the histogram of normalized ranks (i.e. the rank divided by the total number of nodes that were observed as recovered or infected) of the true epidemic origin as obtained from the DMP inference algorithm, compared to the rankings obtained by distance, rumor and Jordan centralities. The DMP algorithm considerably outperforms the three centrality measures, with a comparable computational cost.

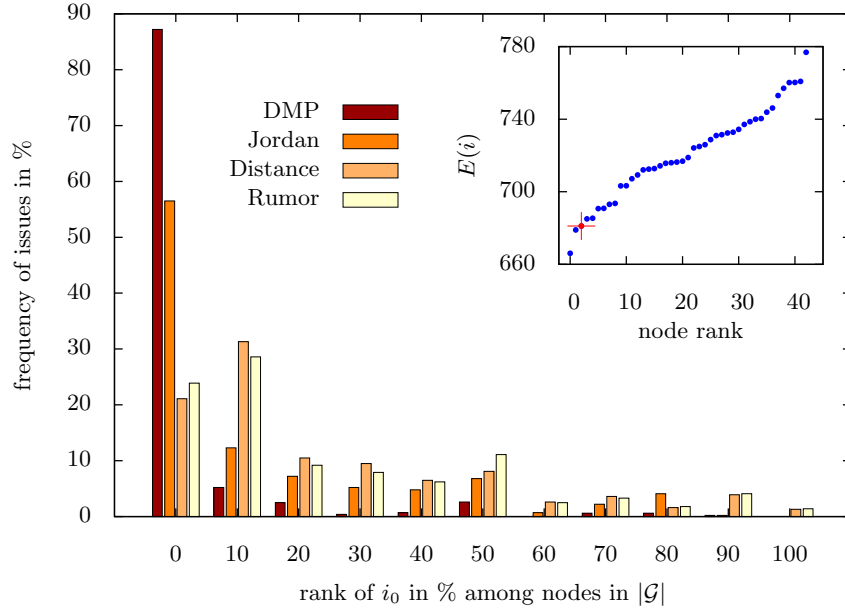


Figure 4.2: A test of inference of the epidemic origin on random regular graphs of degree $c = 4$, size $N = 1000$. Inset: An epidemic is generated with recovery probability $\mu = 1$, transmission probability $\lambda = 0.6$, a snapshot of all the nodes is taken at time $t_0 = 8$ (in this figure we assume we know the value of t_0), 242 nodes are observed to be in the I or R state. The dynamic message-passing is used to compute the energy of every node. This energy is finite for 43 nodes; it is plotted as a function of their rank r . The true patient zero is marked by a red cross, and its rank is $r(i_0) = 2$ in this case. Main figure: an epidemic is generated with $\mu = 1$, $\lambda = 0.5$, $t_0 = 5$. The histogram (over 1000 random instances) of the normalized rank (i.e. the rank divided by the number of R or I nodes in the snapshot) of the true patient zero is plotted for the dynamic message-passing (DMP) inference, as well as for the distance, rumor and Jordan centrality measures.

In Fig. 4.3 and Fig. 4.4 we present the average normalized rank of the true epidemic origin for random regular graphs for the whole range of the transmission probability λ , for different values of the recovery probability μ , and snapshot times t_0 . As an estimation for the spreading time t_0 , we take the one maximizing the “partition function” $Z(t) = \sum_i e^{-E(i,t)}$. The distribution of the estimated time is concentrated at the true spreading time t_0 . The Fig. 4.3 shows the dependence on the spreading time t_0 for fixed values of λ and μ . Note that DMP remains efficient even for relatively large t_0 , when the centrality algorithms fail to make a prediction. Importantly, in some range of parameters, the average normalized rank of the true epidemic origin is not so close to zero (note that the value $1/2$ of the normalized rank corresponds to a random guess of patient zero among all the infected or recovered nodes). The problem of estimating the epidemic origin with a good precision is very hard in these regions. In some cases the information about the epidemic origin was lost during the spreading process. For instance for $\lambda > \lambda_c = \mu/(c - 2 + \mu)$ [New02] the epidemic percolates at large times $t_0 \gg \log_c N$; then the information about the epidemic origin is lost. On the other hand for $t_0 < \log_c N$, the epidemic is confined to a tree network and in this case the inference of the origin is easier, cf. Fig. 4.3. In Fig. 4.4, we mostly focus on the intermediate case $t_0 \approx \log_c N$. We find that for different values of μ , DMP inference always outperforms the centrality

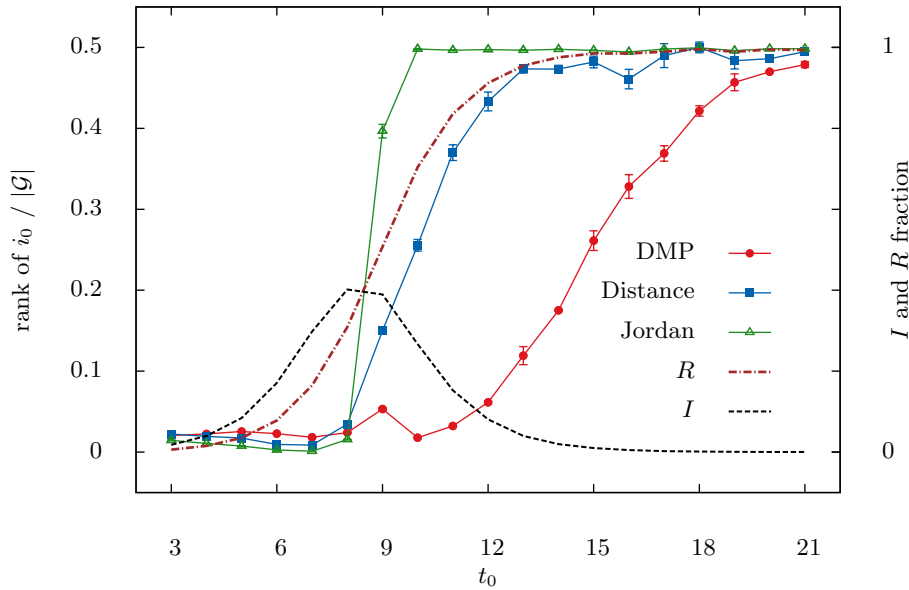


Figure 4.3: Average rank of the true epidemic origin on random regular graphs of size $N = 1000$ with degree $c = 4$. Each data point is averaged over 1000 instances. The plot shows the dependence of the average rank, normalized on the number of infected and recovered nodes in the snapshot $|\mathcal{G}|/N$, on the observation time t_0 for $\lambda = 0.7$ and $\mu = 0.5$. The DMP estimator (red circles) is compared to the Jordan centrality (green triangles) and the distance centrality (blue squares) estimators. The dashed line is the average fraction of nodes that were infected and the dash-dotted line is the average fraction of nodes that were recovered in the snapshot; both are normalized to N . In this figure t_0 is inferred by the algorithm.

measures (see, e.g., case (a)), except in a special case (b) ($\mu = 1$, corresponding to the deterministic recovery), in a range of $0.3 < \lambda < 0.58$ where Jordan center is a better estimation. In other cases, however, Jordan centrality is less performant. Note that for $\mu < 1$ Jordan centrality does not distinguish between recovered and infected nodes, which partly explains its rather bad performance in that case.

We have also studied the performance of the DMP algorithm in the case where the snapshot is incomplete, i.e. the state of a large fraction of nodes is not observed. We compare it to the generalizations of Jordan and distance centralities to this case that we propose in [P-2]. The idea behind this generalization consists in a careful construction of a connected component of infected, recovered and undefined nodes, for which the centrality algorithms can be applied. With incomplete snapshots, the DMP inference algorithm outperforms both centralities even in the particular case $0.3 < \lambda < 0.58$, $\mu = 1.0$, where for complete snapshots the Jordan centrality was better.

In [P-2], the systematic results for other families of random networks are also presented, like the Erdős-Rényi and scale-free random graphs, that can be qualitatively more relevant for applications, see Fig. 4.5 for cartoon examples of these networks. In both cases, the DMP algorithm considerably outperforms Jordan and distance centralities. In order to illustrate the method on non-randomly generated network, we studied the performance of DMP for synthetic data on a real network of the U.S. West-Coast power grid which contains $N = 4941$ nodes with a mean degree $\langle c \rangle = 2.67$ and a maximum degree 19 [WS98], also considered as an application to the patient zero problem in [SZ10]. This network is in fact a widely used example of a real network with the small-world property,

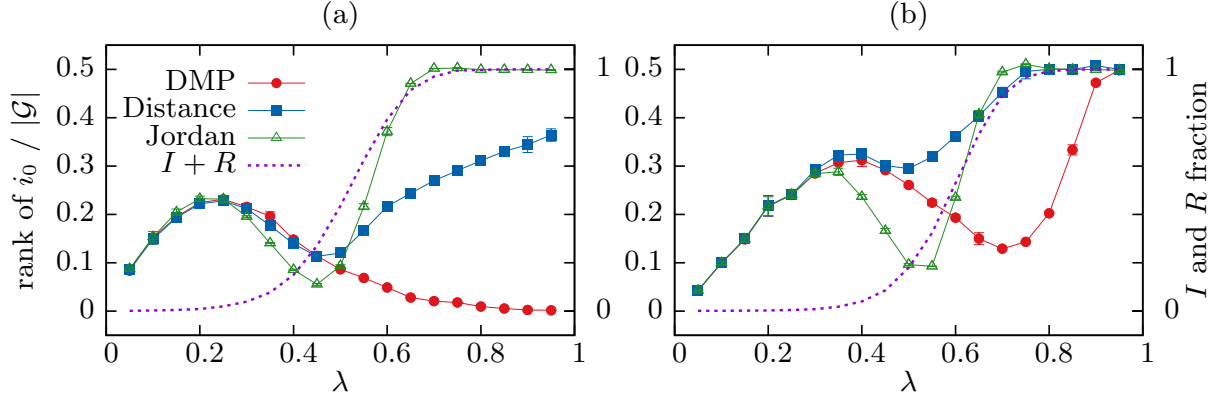


Figure 4.4: Average rank of the true epidemic origin on random regular graphs of size $N = 1000$ with degree $c = 4$. Each data point is averaged over 1000 instances. The plots (a) and (b) represent the dependences of the average rank on the infection rate λ , for the snapshot time $t_0 = 10$ and recovery probability μ : (a) $\mu = 0.5$ and (b) $\mu = 1$. In this figure t_0 is inferred by the algorithm. The DMP estimator (red circles) is compared to the Jordan centrality (green triangles) and the distance centrality (blue squares) estimators. The dotted line shows the average fraction of nodes that were infected or recovered in the snapshot, $|\mathcal{G}|/N$, we use this number to normalize the ranks of the epidemic origin.

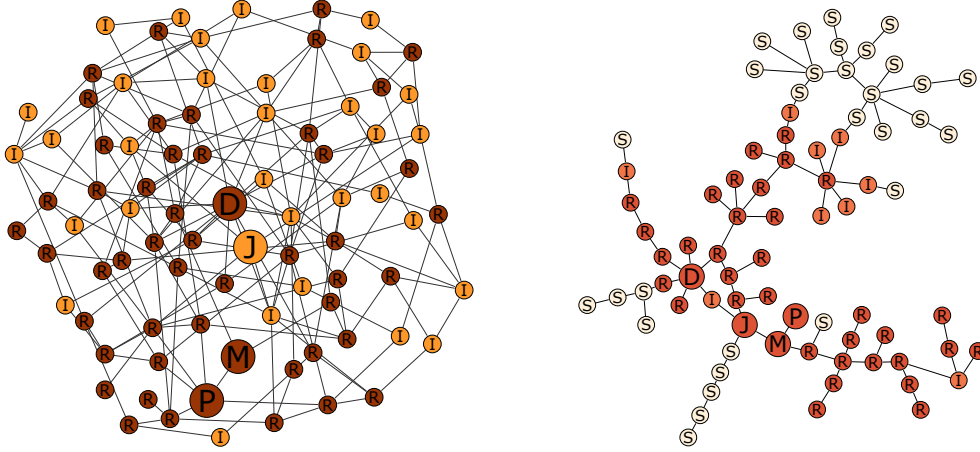


Figure 4.5: Left: An instance of inference problem on the Erdős-Rényi graph with average degree $\langle c \rangle = 4$ and $N = 84$. The epidemic is generated for $\lambda = 0.7$ and $\mu = 0.5$. In this example, only infected (light) and recovered (dark) nodes are present in the snapshot at time $t_0 = 5$. Right: An instance of inference problem on the scale-free graph with average degree $\langle c \rangle = 5/3$ and $N = 77$. The epidemic is generated for $\lambda = 0.7$ and $\mu = 0.5$. The snapshot is represented at time $t_0 = 10$. In both figures, the true patient zero is labeled by P , and the best-ranked nodes for DMP, Jordan and distance centralities are marked by M , J and D , correspondingly.

having a right-skewed degree distribution, and is quite different with respect to an Erdős-Rényi random graph of the same size and mean degree: its measure of cliquishness, the clustering coefficient $C = 0.08$, is much bigger than the transitivity of a corresponding random graph $C_{\text{rand}} = 0.005$ [WS98, New03]. The results are reported in Fig. 4.6: we see that the algorithm works well and DMP estimator gives better prediction for all range of λ .

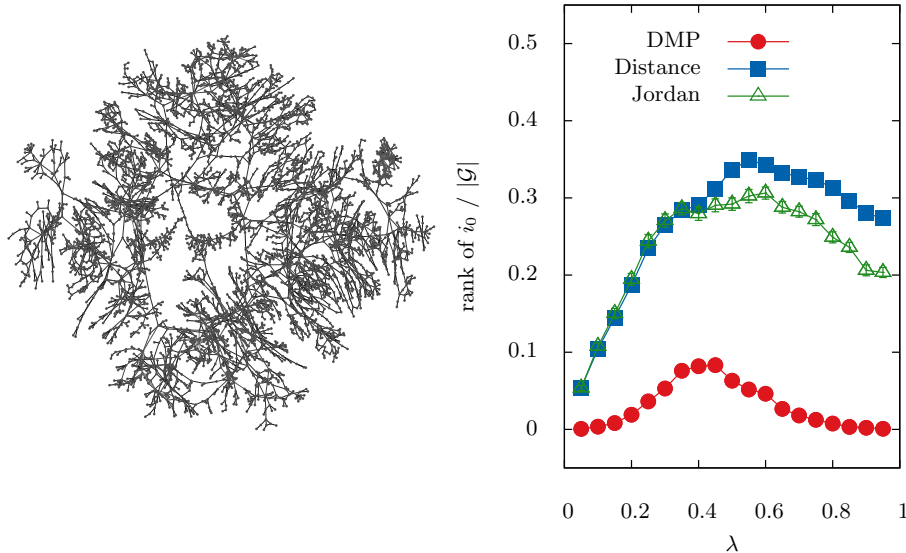


Figure 4.6: Left: A representation of the topology of the U.S. West-Coast power grid network, generated with Gephi [BHJ09]. Right: Normalized rank (averaged over 1000 instances) of the true epidemic origin for epidemic spreading with $\mu = 0.5$ and all nodes observed at time $t_0 = 10$, on the power grid network. DMP inference is significantly better than inference based on distance and Jordan centralities.

Because of its practical relevance, the patient zero problem continues to attract a significant attention. Unfortunately, the problem of finding the maximum likelihood (which would correspond to the Bayesian optimal inference) of a generic snapshot is a NP-hard problem, therefore all tractable algorithms are doomed to give only an approximate solutions. For example, there are two possible sources of sub-optimality for the DMP algorithm on real networks: first, the fact that the message-passing equations may lead to errors on loopy graphs; and second, the mean-field-like approximation (4.3) of the joint probability distribution. We have observed that taking into account the two-point correlation in this approximation does not lead to any improvement in our results. Hence, the performance of each algorithm can be only evaluated in a systematic comparison with other methods; for the DMP algorithm it means that there is no guarantee of best performance compared to other possible algorithms for the patient zero problem. Let us name, in particular, one interesting recent work [ABD⁺14b] which is closely related to the dynamic belief propagation approach, and outperforms the DMP algorithm for most values of the parameters. Let us briefly explain the approach of Altarelli et al., using our notations.

The starting point of [ABD⁺14b] is also the use of the Bayes' theorem, which states that the posterior probability of the initial configuration $\{\sigma_i^0\}_{i \in V}$ (containing the patient zero) given the observed snapshot $\mathcal{O} \equiv \{\sigma_i^{t_0}\}_{i \in V}$ is proportional to the probability of observing the snapshot \mathcal{O} given the initial condition; hence, the same problem remains: what is the best approximation of the probability $P(\mathcal{O} | \{\sigma_i^0\}_{i \in V})$ which appears in (4.3)? The work [ABD⁺14b] exploits the fact that the trajectories can be parametrized by pairs of flipping times $\{(\tau_i, \omega_i)\}_{i \in V}$, and establish the algorithm on the basis of the dynamic belief propagation, presented in the section 2.4: the nodes can be ranked by the values of the marginal probabilities to be in the state I at initial time, i.e. to have the zero susceptible-infected flipping time. In principle, the flipping times marginals $m^i(\tau_i, \omega_i)$ (for $i \in V$) of the forward dynamics can be computed via (2.65), using the belief propagation

equations (2.59) or (2.63) for the SIR dynamics. If this were indeed the case, the DMP algorithm and the approach of [ABD⁺14b] would be equivalent (except for possible convergence issues of the DBP equations on general graphs, as it is discussed by Altarelli and co-authors). Indeed, as we know from the chapter 3 in general, and from the equations (3.40)-(3.42) in particular, the DMP marginals can be expressed through the marginals of the trajectories, and the expression (4.3) can be rewritten as

$$\begin{aligned}
P(\mathcal{O}|i) \simeq & \prod_{\substack{k \in \mathcal{O} \\ \sigma_k(t_0)=S}} \left[\sum_{\tau_k > t} \sum_{\omega_k > \tau_k} m^i(\tau_k, \omega_k) \right] \\
& \times \prod_{\substack{l \in \mathcal{O} \\ \sigma_l(t_0)=I}} \left[\sum_{\tau_l \leq t} \sum_{\omega_l > t} m^l(\tau_l, \omega_l) \right] \prod_{\substack{n \in \mathcal{O} \\ \sigma_n(t_0)=R}} \left[\sum_{\omega_n \leq t} \sum_{\tau_n < \omega_n} m^n(\tau_n, \omega_n) \right], \quad (4.4)
\end{aligned}$$

being approximated by the marginals of the forward dynamics; when represented by a single sum (by opening all the brackets), each term corresponds to a set of flipping times that leads to the observed snapshot. However, the marginals used in [ABD⁺14b] are different: the computed marginals $m_*^i(\tau_i, \omega_i)$ (for $i \in V$) are restricted to be *fully consistent* with the observed configuration \mathcal{O} ; for example, if a node i is observed in the state S , then $m_*^i(\tau_i, \omega_i) = 0$ for all $\tau_i < t$. These modified marginals can be evaluated by explicitly adding extra factors (that ensure the consistency with the observed snapshot) in the standard factor graph, depicted in the Fig. 2.4. Similarly, the inference of the observation time is performed by adding the factious neighbors to each node of the original factor graph, which act as local external fields and may cause a spontaneous infection of nodes. A recent paper [ABD⁺14a] extends this approach to treat other interesting cases of incomplete information, such as confused and noisy observations, or unknown spreading parameters.

As it is argued in [ABD⁺14b], this consistency-restricted scheme is preferable since the mean-field-like approximation used in (4.3) may lead to the errors even on tree graphs. Still, surprisingly, the inferior DMP approach may still provide better results (with respect to the belief propagation approach and the version of the DMP algorithm, restricted to the connected component of infected and recovered nodes, both considered in [ABD⁺14b]) in loopy situations for a narrow range of parameters, for instance for large transmission probabilities λ . It would be interesting to search for better approximations of the likelihood in a general situation.

Chapter 5

Perspectives

In this final chapter of the part I, we will highlight the most important results that are related to the dynamic message-passing approach, and indicate some future directions.

Key results

In spite of some recent progress, the challenging question of efficient description of the non-equilibrium dynamics on networks remains open. In this part we have described a step in this direction, based on a generalization of the cavity method for dynamic problems, defined in terms of time trajectories of interacting variables. This framework makes it possible to develop a general procedure for deriving new dynamic message-passing equations for a large class of models with unidirectional dynamics. A nice property that these DMP equations share with the equilibrium cavity method is that they give asymptotically exact results if the contact network is sparse. The fact that DMP equations are applicable to single instances makes it possible to apply them to the inverse and optimization inference problems. Moreover, solving one instance with arbitrary initial conditions for many models (that involve one non-trivial transition) is as easy as running a single Monte-Carlo simulation, and has a polynomial complexity for more complex models.

Future directions

Let us mention some interesting, in our view, emergent directions that can be treated within the DMP approach, and the study of which are left for the future work. They can be roughly divided in three categories: dynamic on and of networks for existing models, and the development of new behavioral models.

Dynamics on networks

The DMP approach opens a way to a number of important applications aimed to a better control of the *dynamics on networks*. In particular, it would be interesting to explore the following natural applications:

1. Derivation of the DMP equations for the generalizations of the epidemiological models of interest that are adapted to the description of specific infectious diseases [AM91]. One of the possible generalizations consists in including additional states, such as immunized (M) or exposed (E), leading to the models like SEIR

(susceptible-exposed-infected-recovered). Another possibility is to introduce transitions that allow one to skip a certain state during the dynamic process, or to couple the model to external sources: birth, death, migration of agents, as well as a transition of the passive immunity to the newborn individuals. The range of possibilities is quite extended, as it is reflected by the title of the reference [Het94], “A thousand and one epidemic models”;

2. Optimization problems that require an inference of the optimal parameters of the model, e.g. determination of the most influential nodes in a network for an efficient information dissemination or development of the immunization strategies. A large amount of works exist on these topics; as an example, let us indicate a couple of recent works [ABDZ13b, ABDZ13a, GS14] that solve the inverse optimization problems with the belief propagation techniques using the irreversibility of the dynamics. The DMP method could also be used as a basis for the practical problems related to the optimal detection of cybernet attacks in computer networks, as well as for the control of failures propagation in infrastructures and power grids;
3. A very important direction that we will investigate consists in a development of the systematic approximations for the out-of-equilibrium models with general unrestricted dynamics. It turns out that the models defined on purely directed networks can be treated exactly within the dynamic cavity method, see (2.68). For undirected networks, the main idea would consist in approximating the general dynamics by a corresponding unidirectional process, for which the DMP equations can be derived, using the method presented in the chapter 3. The potential applications are numerous: dynamics of glassy systems with finite connectivity, modeling of the endemic diseases, dynamics of the algorithms in inference and constrained satisfaction problems, and the study of quantum problems.

Dynamics of networks

A second class of naturally emerging problems is related to the *dynamics of networks*: the study of models defined on dynamically-changing networks, and identification of the optimal rewiring strategies for adaptation and control of cascading processes in communication and electrical networks, see e.g. [SS11, VSP10, SwS13]. Note that the DMP equations can be readily applied to the contact networks that evolve in time: one only needs to encode the dynamics of the network into the time-changing transmission probabilities, e.g. $\lambda_{ij}(t)$ for the SIR model (2.51)-(2.56). The SIR model on dynamically changing networks has been already studied using the graph-averaged version of the DMP equations in [VM07, VM09]. We anticipate that the DMP approach to inference problems on a single graph will also be useful for studies where specific experimental data about the changing network, such as those of [SVB⁺11], can be used. The study of the networks that evolve on a timescale comparable to that of the dynamic process (as it generally happens in the real-world social networks) would allow to close the gap between the two descriptions, discussed in the section 2.3: quenched static networks, and rapidly changing networks that lead to a fully-mixed dynamics, in which the network effectively disappears from the mathematical description.

Behavioral models

Last but not least, one of the major challenges of actual research consists in understanding the co-evolution of networks with the dynamic process. The adaptive behavior of agents is an essential ingredient for an adequate description of a spreading process: the epidemic or opinions are spreading due to the contacts between individuals, but the individuals minimize, maximize or establish their interactions depending on the state of the infection or information spread. A correct characterization of the interplay between the dynamics and the behavioral feedback is therefore an important issue in the study of the spreading processes [FSJ10, Ves11]. Hence a necessity to construct new models of epidemic and information spreading that could include behavioral changes of individuals due to the propagation of awareness, or to the mass media effect: being widely distributed, media reports play a key role in the influence of people's behavior. Several attempts have been made in this direction in the context of epidemics, consisting in treating at the mean-field level the extensions of epidemiological models that incorporate the effect of mass media [TDB⁺11, TB12, MSS11], or make use of coupled disease-awareness spreading models [FGWJ09, FGJ10, PBGV11]. The awareness effects are usually taken into account by the changes in the couplings strength depending on the spread of the dynamic process, either on a local or on a global scales. The DMP equations can be useful for solving these emergent behavioral models.

Part II

Planar matching and optimal folding

Chapter 6

Introduction

Optimization is an omnipresent concept in nature in general, and in human activities in particular. A large number of problems of a very broad scope can be formulated as optimization problems, which consist in finding a configuration of variables that minimize some cost function. Recently, the discovered deep connections between statistical physics and optimization (see discussion in the subsection 2.1.3) have brought the understanding of the hardness appropriate to the optimization problems to a new level.

The *matching optimization problems* on graphs are frequently used as a playground for testing new ideas in the complexity science. In general, finding a maximum matching configuration is known to be a relatively easy problem of *polynomial complexity*, while computing the number of matchings is classified as an *exponential NP-hard* problem. In this part, we present a particular matching problem under a global planarity constraint, in which both *finding* and *counting* the number of maximum matchings is a polynomial problem. Despite this algorithmic simplicity, the analytical description of the easiest matching models appear to be rather involved. Although defined through only a few parameters, these models exhibit a non-trivial critical behavior. A particular interest for the study of the planar matching comes from its relevance to the statistical mechanics of the RNA secondary structures.

In this chapter, we will first define the simplest model for the disordered planar matching, the *Bernoulli model*, and present its relation to the toy model of the RNA molecules. Then, we will discuss the numerical results on the topological phase transition in this model, described in [P-3] and [P-4], and point out the difficulties of the analytical estimation of the critical point, arising from the quenched nature of disorder.

6.1 Disordered planar matching: definitions

The matching problems have attracted considerable attention in mathematics, physics and computer science communities, for a review see [LP09]. In the most general setting, they are defined as follows: given a graph $G = (V, E)$ with N nodes, find a collection of M edges on this graph, so that each node is incident with at most one edge in this set. When $M = N/2$, we say that the matching is *complete*, or *perfect*. If at most M_{\max} non-touching edges can be chosen at once, we define M_{\max} as the size of the *maximum matching*. Although finding a maximum matching on a general graph is a polynomial problem, counting the number of solutions might be a problem of an exponential complexity [MV80]. However, for some special graphs, the algorithmic complexity is reduced: for example, an equivalent problem of counting dimer covers on planar lattices has been

solved exactly by Kasteleyn back in 1961 [Kas61].

Deep connections between the spin-glass theory (cf. section 2.1) and optimization problems have allowed to apply the techniques developed for the study of disordered systems to the study of statistical properties of optimization problems. One of the pioneering examples is given by an application of the replica theory to bipartite weighted matching problem with random identically distributed independent weights [MP85, MP87b], computing the average minimal total weight. The belief propagation approach (described in 2.1.3) to this assignment problem has been studied in [BSS08]. The $\zeta(2)$ -limit has been rigorously proven by Aldous [Ald01] by means of the local weak convergence method.

An instance of the matching problem is usually defined by a matrix of weights A_{ij} between the N points. In addition, the matching assignments can be subject to some global constraint. In the case of planar matching models on a line that we consider in this part, this non-local topological constraint is given by the requirement to have a planar structure for the optimal matching configurations. The planar diagrams play a key role in many areas, including matrix and gauge theories [BIPZ78], many-body condensed matter physics [AG75], quantum spin chains [Sai90], random matrix theory [Meh04]. Another area where the planar matching appears naturally is the biophysics of secondary structures of RNA molecules, see next section 6.2. Probably the simplest model of this kind is the Bernoulli model [Neb04], defined as follows.

Take L points¹ $i = 1, \dots, L$ on a line, and define the entries of the symmetric contact matrix A as independent identically distributed random variables, generated by the distribution

$$\text{Prob}(A_{ij}) = p\delta(A_{ij} - 1) + (1 - p)\delta(A_{ij}), \quad (6.1)$$

where $\delta(x) = 1$ for $x = 0$, and $\delta(x) = 0$ otherwise. In other words, each element $A_{ij} = A_{ji}$ is independently either one with probability p for any $i \neq j$, or zero otherwise. Now we draw $L/2$ non-intersecting links between pairs of points allowed by the non-zero entries A_{ij} , such that each point is involved in one link only and the links form a planar diagram, see Fig. 6.1(a). If at least one such set of links exists, we say that the problem allows for the *perfect matching* solution. If the maximum matching structure includes some gaps, this configuration will be referred to as *imperfect matching*.

Let us point out an important one-to-one mapping between the L -point planar diagrams and the L -step Brownian excursions, known as Dyck paths [Lan03]. In this representation, also called mountain (or height) diagram, each monomers is represented by either an “up”-step (\nearrow) or a “down”-step (\searrow) with “up”-steps corresponding to opening arcs, and “down”-step to closing ones. An example is given in the Fig. 6.1, with the steps up and down at positions 2 and 9, corresponding to the arc between points 2 and 9 in the planar matching structure. The total number of Dyck paths of even length L is given by a Catalan number

$$C_{L/2} = \frac{L!}{(\frac{L}{2})!(\frac{L}{2} + 1)!} \sim \frac{2^L}{L^{3/2}} \sqrt{\frac{2^3}{\pi}}, \quad (6.2)$$

where the asymptotic expression is valid for $L \gg 1$. If $p = 1$ in our matching problem, all the planar configurations are solutions to the perfect matching problem, and their total number is then also given by (6.2).

¹Here and through the rest of this part, we will use the notation L for the number of nodes in order to avoid a confusion with a traditional notation for the 't-Hooft $1/N$ expansion that is used in the section 7.1.

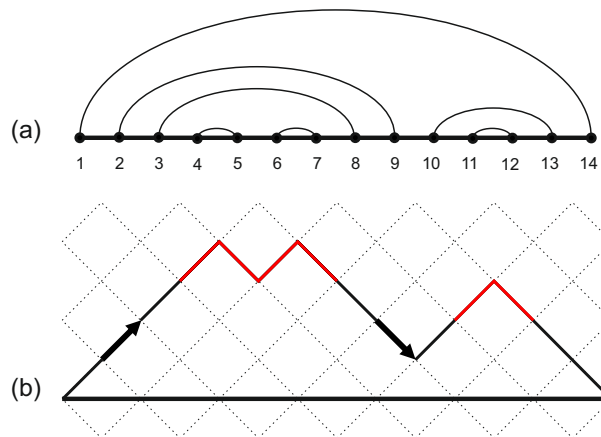


Figure 6.1: Example of (a) a perfect planar matching configuration, and (b) the corresponding mapping to a Dyck path. The arc is given by an “up” and “down” steps at the same height, shown by arrows \nearrow and \searrow . The part of the walk between arrows is a Dyck path itself. The shortest arcs correspond to the peaks of the Dyck path representation (marked with red).

For $0 < p < 1$, some planar diagrams in the fully-connected ensemble are forbidden. This reduces the number of possible planar configurations, which drops to zero below a certain value of link formation probability p_c , indicating an occurrence of the phase transition [P-3]. It can be equivalently thought of as a transition in a *constrained satisfaction problem* [Fri99]: as the number of constraints per node, imposed by the matrix A , is below a certain critical value, the problem exhibits a perfect matching solution, while otherwise no complete matching solution exists in the large L limit. Before presenting the details of this phase transition, given in 6.3 and 6.4, let us describe a connection of this model to the statistical physics of the RNA secondary structures.

6.2 Relation to the statistical mechanics of RNA secondary structures

A real RNA is a single-stranded biopolymer composed of four types of nucleotides: A, C, G, and U. The sequence of these monomers encodes the genetic information, and is called the *primary structure* of RNA. The polymer is not free, and is characterized by some compactification properties: unlike DNA that form a long double-helical structure with a complementary strand, a single-stranded RNA polymer folds onto itself in order to create local double-helical structures of stable Watson-Crick pairs A-U and G-C between complementary subsequences of the same strand. The set of all chemical bonds of this kind determines the *secondary structure* of RNA. The study of the secondary structures in RNA molecules is relatively easy compared to the investigation of the protein folding due to the separation of energy scales: the effective stacking energies (that include entropy change due to the loss of single-stranded degrees of freedom, as well as the expulsion of water molecules surrounding the hydrophobic parts of base pairs) of contacts are considerably larger than the energies associated with the *tertiary structure*, i.e. the spatial arrangement of the biopolymer.

Another restrictions that are generally assumed for the RNA secondary structures is the saturation of base pairings and the exclusion of the *pseudoknots*, depicted on

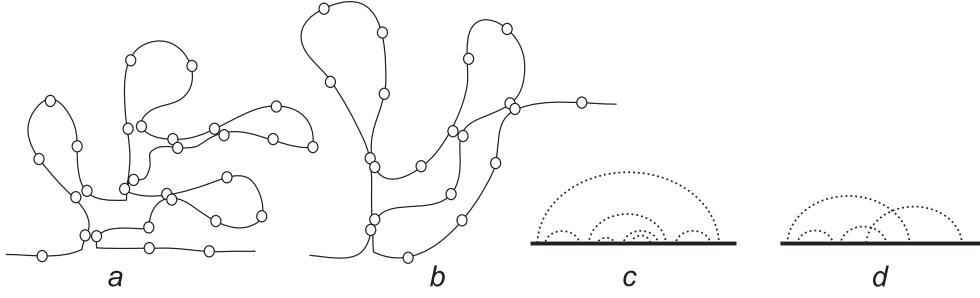


Figure 6.2: Examples of the RNA secondary structures and corresponding one-dimensional contacts representations: with [(b) and (d)] and without [(a) and (c)] pseudoknots.

the Fig. 6.2(b), and which are known to be rare in real RNAs [vB00]. If we number the monomers along the chain, this restriction means that an accepted configuration of bonds between monomers is subject to a constraint that two base pairings (i, j) and (k, l) are either *independent* $i < j < k < l$, or *nested* $i < k < l < j$. In other words, a valid secondary structure configuration, presented in the Fig. 6.2(a), leads to a planar diagram of contacts Fig. 6.2(c), already discussed the section 6.1, while the knotted configurations Fig. 6.2(b) and Fig. 6.2(d) are forbidden.

Although the primary sequence of a RNA molecule is obviously not random, the basic models of *random sequences* are used as a common playground for theoretical investigations of the RNA statistical properties. Given an alphabet of c monomer types, the nucleotides of different types are distributed at random in a chain, and the effective contact energies ϵ_{ij} between different monomers in a chain are defined. In practice, these energies are usually assumed to be taken from some probability distribution. Therefore, we see that the problem of optimal secondary structure (see Fig. 6.3(c)) is equivalent to the optimal matching problem, described in the previous section 6.1, if ϵ_{ij} are drawn from a bimodal probability distribution (6.1); continuous distribution of ϵ_{ij} would lead to a weighted formulation of the problem.

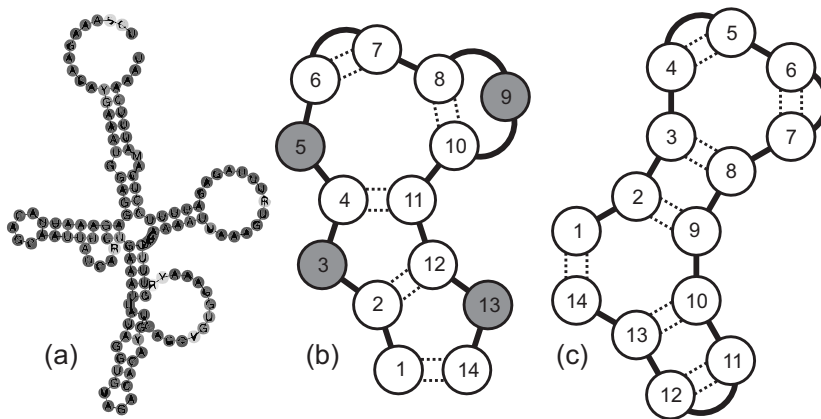


Figure 6.3: (a) Secondary structure of a real RNA gene HAR1F [PSL⁺06]; Examples of RNA secondary structures: with (b) and without (c) gaps, corresponding to the imperfect and perfect structures in the planar matching formulation.

Intuitively, it is clear that on the energy scale of order ϵ_{ij} the structure of the *heteropolymer* would play a significant role in the formation of the secondary structure,

while for much larger energies all the contacts will be more or less equivalent, and the RNA molecule would demonstrate a *homopolymer*-like behavior. This difference will be crucial for the molten-glass transition, discussed in the section 8.1. The physics of dAT homopolymers have been first studied by de Gennes [dG68], a recent review on the homopolymer models of RNA is given in [M03]. The parameter c is often taken equal to four, corresponding to the four-letter alphabet $\{A, C, G, U\}$ in real RNAs. However, for the modeling purposes, the parameter c , as well as particular matching rules can be used along with the temperature as tuning parameters for the *quenched sequence disorder* [BH02]. Hence, further (in the section 8.2) we will generalize the model to an alphabet with c arbitrary monomer types $\{A, B, C, \dots\}$, and assume for simplicity that the matching occurs between the monomers of the same type: A-A, B-B, C-C, *etc.* Although in the real RNA the matching rules involve instead pairings between the complementary monomer types, this difference becomes irrelevant for large sequences: what matters is the density of contacts in the contact matrix A . In Bernoulli model, these matching rules would correspond to a density $p = 1/c$ of ones. Note that $c = 1/p$ is not an integer for a general p , and one should hence give a physical sense to the notion of the non-integer alphabet; see section 8.2 for a detailed discussion. More realistic models of RNA folding include the condition of finite flexibility of the molecule, requiring a minimal length of a loop [MPRT02, KMM02]; we neglect this minimal hairpin condition since it is not natural *a priori* for the planar matching problem formulation, and is not crucial for the problem to be frustrated [BH02, SGS93].

The low-temperature properties of random RNAs have been studied in a number of works [BH99, PPRT00, Har01, PPRT01, KMM02, MPRT02, BH02]. These works have identified a finite-temperature transition to the glassy phase at sufficiently low temperatures, and have discussed the exponents in the glassy phase. Our formulation as an optimization problem corresponds instead to the zero-temperature limit of the problem, but for a varying alphabet; a connection to the freezing transition will be discussed in the section 8.1.

The aforementioned topological constraints and simplifications of the model for the secondary structures of RNA allows for an exact computation of the RNA's partition function [dG68, Hig96], leading to efficient algorithms for the prediction of the corresponding secondary structures based on the minimization of the corresponding free energy [NJ80, ZS81]. We use a version of this prediction algorithm based on dynamical programming procedure, and present the numerical results for the zero-temperature transition in the Bernoulli model in the next section.

6.3 Topological transition: numerical

Let us briefly discuss the numerical investigation of the topological transition in the planar matching problem, predicted in [VTN12] and described in [P-3]. It is based on an exact dynamical programming algorithm for a given quenched contact matrix. First of all, let us observe that the recursion relation for the partition function $Z_{i,i+k}$ of the part between nodes i and $i+k$ reads [dG68, Hig96]:

$$Z_{i,i+k} = Z_{i+1,i+k} + \sum_{s=i+1}^{i+k} \beta_{i,s} Z_{i+1,s-1} Z_{s+1,i+k} \quad (6.3)$$

where $\beta_{i,j} = e^{-A_{ij}/T}$ are statistical weights of bonds ($1 \leq i < j \leq n$), and A_{ij} are the elements of the contact matrix: $A_{ij} = 1$ if i and j match each other, and $A_{ij} = 0$ otherwise. In the zero-temperature limit, the equation (6.3) is reduced to the dynamical programming algorithm for the ground state free energy [NTV11]:

$$F_{i,i+k} = \lim_{T \rightarrow 0} T \ln Z_{i,i+k} = \max_{s=i+1, \dots, i+k} [F_{i+1,i+k}, \varepsilon_{i,s} + F_{i+1,s-1} + F_{s+1,i+k}]. \quad (6.4)$$

This expression can be interpreted as a growth of the optimal graph: at each time step a node is added to the sequence, and the new graph is rewired according to the non-crossing constraint, in order to minimize the number of gaps, see Fig. 6.4. The dynamic programming algorithm (6.4) has a cubic computational complexity in the length of the sample L .

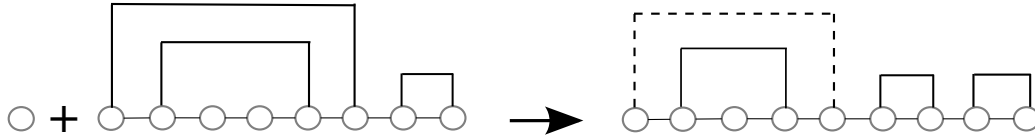


Figure 6.4: An interpretation of the algorithm (6.4) in terms of a growing graph: a newly added node creates an optimal bond with one of the existing nodes, and other links rewire in order to satisfy the non-crossing constraint.

Since the free energy F of the whole chain is proportional to the number of nucleotides involved in the planar bond formation, the combinatorial problem of planar matching can be regarded as a $T = 0$ optimization problem for the free energy of the RNA molecule with a given matrix of contacts, A . Therefore, the exact dynamical programming algorithm (6.4) allows to detect the phase transition by considering the fraction $f_L(p) = 2F/L$ of links, involved in planar binding, for different densities of contacts p in the limit $L \rightarrow \infty$: one expects $f_\infty(p) = 1$ for $p > p_c$, and $f_\infty(p) < 1$ for $p < p_c$.

Thus, looking for the fraction $\eta_L(p)$ of sequences, which allow perfect matchings, in the whole ensemble of random sequences, one has $\eta_\infty(p) = 1$ for $p > p_c$, and $\eta_\infty(p) = 0$ for $p < p_c$. The corresponding dependencies are shown in Fig. 6.5(a) for different polymer lengths, $L = 500, 1000, 2000$. As $L \rightarrow \infty$, the function $\eta_L(p)$ tends to a step function. Two different phases are observed: for $p > p_c$ one has a gapless perfect matching with all nodes involved in planar binding, while for $p < p_c$ there is always a finite fraction of gaps in the best possible matching.

The scaling analysis determines the phase transition at the critical point $p_c \approx 0.379$. The Fig. 6.5(b) shows that curves with different L collapse, demonstrating the scaling behavior $\eta((p - p_c)/L^\nu)$, giving the transition width in form of power-law decay $L^{-\nu}$, with $\nu = 0.5$. The convergence of the function f_L to a limiting value $f_\infty(p)$ (cf. Fig. 6.6) in the perfect and imperfect phases has, respectively, an exponential and a power-law tails:

$$\begin{cases} f_\infty(p) - f_L(p) \sim e^{-L/\ell(p)} & \text{for } p > p_c \\ f_\infty(p) - f_L(p) \sim L^{-\alpha(p)} & \text{for } p < p_c \end{cases} \quad (6.5)$$

where the screening length $\ell(p)$ diverges at the point $p = p_c$ (two examples for $p = 0.38$ and $p = 0.4$ are shown on Fig. 6.6(a) in the semi-logarithmic scale), and finite-size scaling analysis gives $0.8 \leq \alpha(p) \leq 1$ (see Fig. 6.6(b) for two examples, $p = 0.32$ and $p = 0.34$ on the log-log plot). Note that the exponential scaling in the perfect phase may not be universal (with respect to other models) and is likely to be a feature of the Bernoulli

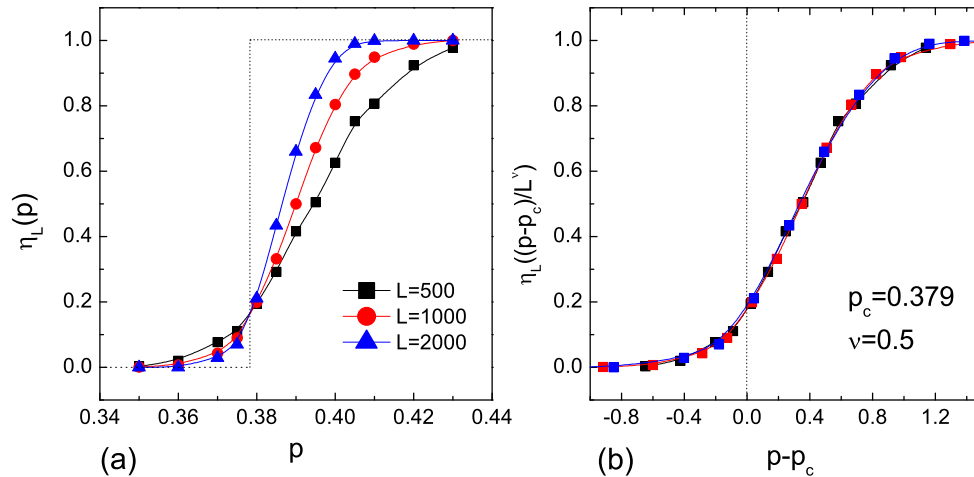


Figure 6.5: (a) The fraction of perfect matchings $\eta_L(p)$ as a function of the density p of ones in the contact matrix A for chain lengths $L = 500, 1000, 2000$, averaged over 10000 instances. The dashed line corresponds to the thermodynamic limit $L \rightarrow \infty$, yielding the critical value $p_c = 0.379$. (b) The scaling analysis of curves, corresponding to different chain lengths L . The fitting procedure gives the exponent of the transition width $\nu \approx 0.5$.

model (6.1), while the power-law behavior in the imperfect phase appears in other models, e.g. for integer-valued alphabet [VTN12].

6.4 Topological transition: naive mean-field and correlations

Let us indicate some immediate considerations with respect to the perfect-imperfect phase transition, reported in the section 6.3. First of all, let us mention the lower and upper bounds on the critical transition point that can be readily obtained using an explicit formulation in terms of integer-valued alphabets. It is easy to realize that a two-letter sequence always allows for a perfect-matching solution in the thermodynamic limit $L \rightarrow \infty$: indeed, it is not hard to show that the sequential pairing of neighbors of the same type and their removal from the chain ultimately gives at most two unpaired nodes, see an example in the Fig. 6.7. Hence, one immediately gets $p_c < 1/2$. A lower bound $p_c > 1/3$ for the RNA-type matching has been found in [Vla13] using the explicit construction in terms of a three-letter alphabet. A discussion on more elaborated rigorous bounds directly for the Bernoulli model is presented in the chapter 9.

A naive estimation of p_c can be easily obtained via the following mean-field-like argument. Since each arc in the diagram is present with the probability p , the probability that the whole configuration is allowed, is given by $p^{L/2}$. Assuming that planar diagrams in the fully-connected ensemble are *statistically independent*, we get the probability \mathcal{P} to have at least one perfect planar matching configuration:

$$\mathcal{P} = 1 - (1 - p^{L/2})^{C_{L/2}} = 1 - \exp(-p^{L/2} C_{L/2}), \quad (6.6)$$

where the last equality is valid for $L \rightarrow \infty$. In this limit, the probability \mathcal{P} is equal to one for $p > p_c$, and to zero for $p < p_c$. The perfect-imperfect *naive mean-field* threshold

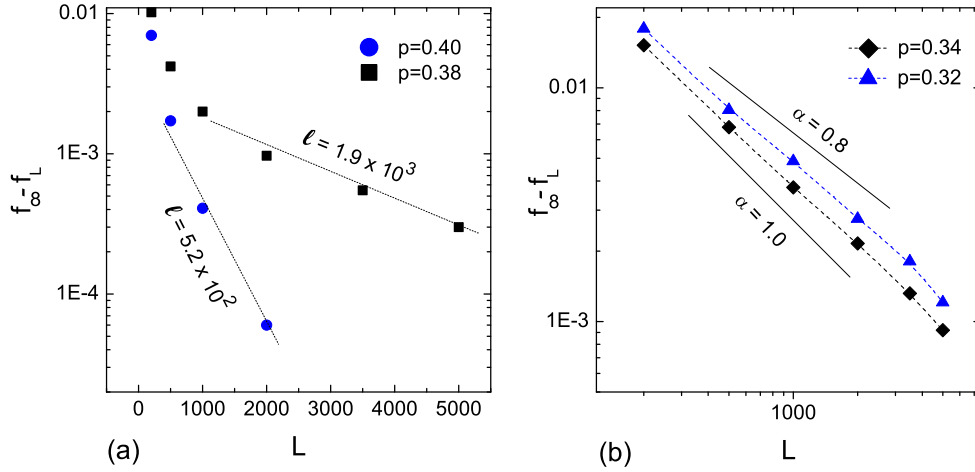


Figure 6.6: Convergence of fraction of links, involved in planar binding, f_L to the limiting value f_∞ in two regimes, $p > p_c$ and $p < p_c$. (a) In the perfect phase, the exponential convergence is demonstrated for $p = 0.38$ and $p = 0.4$ in the semi-logarithmic scale. The screening length $\ell(p)$ diverges as p approaches the critical value p_c . (b) In the imperfect phase, the power-law behavior is shown for $p = 0.32$ and $p = 0.34$ in the log-log scale. The exponent $\alpha(p)$ as a function of p takes values between 0.8 and 1. The data points are averaged over 1000 instances.

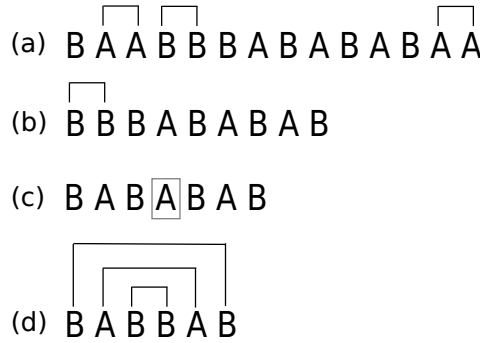


Figure 6.7: The *word reduction* algorithm for a two-letter sequence. At a first stage (a), we pair and remove all neighboring monomers of the same type. This procedure is repeated for a reduced sequence (b) until the irreducible sequence of the alternating type (c) is left. For this sequence, it is sufficient to remove one of the central monomers, counting it is a gap, and repeat the elimination procedure (a) and (b). At most 2 gaps are left overall at the final stage (d).

p_c is thus given by the condition

$$\lim_{L \rightarrow \infty} p_c [C_{L/2}]^{2/L} = 1, \quad (6.7)$$

yielding $p_c = 1/4$. However, here we have neglected the statistical correlations between different configurations in the fully-connected ensemble of planar configurations. For instance, let τ and ρ be the two arbitrarily chosen configuration of arcs out of the fully-connected ensemble of $C_{L/2}$ possible configurations. The probability that they both satisfy the constraints imposed by the contact matrix A is not simply equal to p^L , but instead is given by $p^{L/2} p^{L/2} p^{-n_{\tau \cap \rho} L/2}$, where $n_{\tau \cap \rho}$ is a fraction of common arcs in the

configurations τ and ρ .

This argument can be extended in order to introduce quantities $n_{1\cap 2\cap \dots \cap m}$ for the number of mutual arcs between m randomly picked configurations. Therefore, a more accurate expression for the probability \mathcal{P} of existence of at least one solution can be written using the inclusion-exclusion principle for the expectation values of overlap fractions:

$$\mathcal{P} = C_{L/2}p_1 - \frac{C_{L/2}(C_{L/2} - 1)}{2}p_2 + C_{C_{L/2}}^3p_3 + \dots, \quad (6.8)$$

where $p_1 = p^{L/2}$, $p_2 = p^L \mathbb{E} [p^{-n_{1\cap 2}L/2}]$, $p_3 = p^{3L/2} \mathbb{E} [p^{-C_3^2 n_{1\cap 2}L/2} p^{n_{1\cap 2\cap 3}L/2}]$, *etc.* However, the computation of these expectation values is very hard (a computation of the two-configuration overlap is sketched in the chapter 9), and the alternating series (6.8) is not easy to analyse. Moreover, a quick survey shows that the maximum contribution to this series comes from rather distant terms, and therefore it can not be truncated at a small number of terms. Of course, if we assume that all the diagrams in the fully-connected set are independent, we have $p_m = p_1^k$, and thus we recover the expression (6.6). Anyway, we see that the equation (6.7) provides only a crude estimation to the true value of p_c , and it has to be generalized to

$$\lim_{L \rightarrow \infty} \xi(p_c) [C_{L/2}]^{2/L} = 1, \quad \xi(p_c) = 1/4, \quad (6.9)$$

where $\xi(p)$ is some weight (due to correlations) to be determined. In the next chapter, we will see how to calculate the transition value analytically in a more accurate way.

Chapter 7

Analytical description of the topological phase transition in the Bernoulli model

In this chapter, we present analytical estimations of the perfect-imperfect transition point in Bernoulli model, described in the previous chapter. The difficulty in the calculations arise essentially from the quenched nature of disorder in the random contact matrix that has to be averaged out. We will follow two approaches. The first one is based on a field-theory generating function for planar diagrams, the corresponding estimation has been reported in [P-3]; in this chapter, we will reveal some technical details that were not present in the work [P-3]. The second approach relies on a combinatorial treatment of the quenched disorder, and opens a way to a systematic improvement of the estimation of the critical point. A key ingredient that makes the problem solvable is the fact that the *global* constraint satisfaction problem can be reduced to a set of *local* ones that are easier to solve. In this chapter, we will present the obtained results, and summarize the main features of the developed *combinatorial expansion* procedure.

7.1 Field-theory approach

In the previous chapter, we have highlighted a similarity between the perfect matching configurations of arcs and the planar diagrams that play an important role in the field theory. Indeed, the planar diagrams represent a dominant contribution to the perturbative computation of Green functions in the limit of large number of dimensions of the problem. Examples include the $1/N$ expansion in the $SU(N)$ theory of strong interactions [Col88], introduced for the first time by 't Hooft [’t 74], or the resolvent computation in the random matrix theory [Meh04]. In this section, we use this property for generating the appropriate configurations in the planar matching problem. As a possible application to the physics of RNA, this approach has been suggested by Orland and Zee in [OZ02], and further developed in [VOZ05] for the counting problem and in [POZ05] for the study of the tertiary structure. Although this approach is fruitful in a homogeneous setting, assuming that all the nodes of the biopolymer are equivalent, the problem becomes much harder in the heterogeneous case. In the following, we apply the $1/N$ -expansion technique to the disordered Bernoulli model. Let us start with the matrix model, previously suggested in [VOZ05], and adapted to our setting. For the set of L vertices, associate to vertex i an Hermitian matrix $(\phi_i)_{N \times N}$. The L -point generating function \mathcal{Z}_L can be

written as follows:

$$\begin{aligned}\mathcal{Z}_L(N, A) &= \frac{1}{W_L(N)} \int \prod_{k=1}^L d\phi_k e^{-\frac{N}{2} \sum_{i,j=1}^L (A^{-1})_{ij} \text{tr}(\phi_i \phi_j)} \frac{1}{N} \text{tr} \prod_{l=1}^L (1 + \phi_l) = \\ &= \frac{\int D\phi_i e^{H_0} \frac{1}{N} \text{tr} \prod_{l=1}^L (1 + \phi_l)}{\int D\phi_i e^{H_0}} \equiv \langle \prod_{l=1}^L (1 + \phi_l) \rangle_{H_0},\end{aligned}\quad (7.1)$$

where

$$H_0 = -\frac{N}{2} \sum_{i,j} (A^{-1})_{ij} \text{tr}(\phi_i \phi_j). \quad (7.2)$$

Since $\text{tr}(\phi_i \phi_j) = \sum_{a,b} \phi_{ab}^i \phi_{ba}^j = \sum_{a,b,c,d} \delta_{ad} \delta_{bc} \phi_{ab}^i \phi_{cd}^j$, every propagator enters with a $1/N$ factor, while every loop gives a factor N , and we have:

$$\langle \phi_i \phi_j \rangle_{H_0} = \frac{1}{N} \left(\frac{1}{N} A_{ij} \right) \sum_{a,b=1}^N \delta_{aa} \delta_{bb} = A_{ij}, \quad (7.3)$$

and it is easy to verify that due to the Wick theorem

$$\langle \phi_1 \dots \phi_L \rangle_{H_0} = \langle \sum \prod_{\text{pairs } k,k'} \phi_k \phi_{k'} \rangle_{H_0}, \quad (7.4)$$

one has

$$\mathcal{Z}_L(N, A) = 1 + \sum_{i < j} A_{ij} + \dots + \sum_{i < j < k < l} A_{ij} A_{kl} + \dots + \frac{1}{N^2} \sum_{i < j < k < l} A_{ik} A_{jl} + \dots \quad (7.5)$$

Hence, each non-planar configuration comes with a factor $1/N^2$ at some power and therefore vanishes in the large N limit. Since we are interested only in the subclass of complete diagrams, it will be easier to work with the generating function $Z_L(N; A)$

$$Z_L(N; A) = \frac{\int d\phi_1 \dots d\phi_L e^{H_0} \frac{1}{N} \text{tr}(\phi_1 \dots \phi_L)}{\int d\phi_1 \dots d\phi_L e^{H_0}} \equiv \langle \phi_1 \dots \phi_L \rangle_{H_0} \quad (7.6)$$

that counts in the limit $N \rightarrow \infty$ the number of planar diagrams with exactly $L/2$ arcs compatible with a specific realization of the disorder defined by the matrix A .

In the absence of any disorder, one can set $A_{ij} \equiv \alpha$ for any (i, j) , where α is some constant (it corresponds to the $p = 1$ limiting case in the Bernoulli model):

$$A = \begin{pmatrix} 0 & \alpha & \alpha & \dots & \alpha \\ \alpha & 0 & \alpha & \dots & \alpha \\ \alpha & \alpha & 0 & \dots & \alpha \\ \vdots & & & \dots & \alpha \\ \alpha & \dots & & \alpha & 0 \end{pmatrix}. \quad (7.7)$$

In this case the multi-dimensional integral (7.6) can be reduced by a series of Hubbard-Stratonovich transformations to a one-dimensional integral involving the spectral density of a Gaussian matrix, which is a well-known result of the random matrix theory [VOZ05,

GD09]. We will refer to this realization of A as to the fully-connected case. If we set $\alpha = 1$, we get [VOZ05]

$$\lim_{N \rightarrow \infty} Z_L(N; A) = C_{L/2}, \quad (7.8)$$

where $C_{L/2}$ is the Catalan number (6.2), as it should be. However, for a generic disordered matrix A , the calculations are intractable. Still, we show below that by averaging over the matrix distribution (6.1) and by applying the self-consistency arguments, we are able to treat the partially-connected system with $0 < p < 1$ as an effective fully-connected system with α different from one, thus obtaining a correction to the naive mean-field result (6.7). Following this arguments, the function $\xi(p)$ defined by the equation (6.9) can be computed by averaging $Z_L(N; A)$ over disorder:

$$\langle Z_L(N, A) \rangle_A \equiv \int dA_{ij} P(A_{ij}) Z_L(N, A). \quad (7.9)$$

Changing the order of integration, introducing new hermitian matrices $(h_i)_{N \times N}$ and performing the Hubbard-Stratonovich transformation, we get:

$$\begin{aligned} \langle Z_L(N, A) \rangle_A &= \int dA_{ij} P(A_{ij}) \frac{1}{W_L(N)} \int \prod_{k=1}^L d\phi_k e^{-\frac{N}{2} \sum_{ij} (A^{-1})_{ij} \text{tr}(\phi_i \phi_j)} \frac{1}{N} \text{tr}(\phi_1 \dots \phi_L) \\ &= \int \prod_{k=1}^L d\phi_k \frac{1}{N} \text{tr}(\phi_1 \dots \phi_L) \int dA_{ij} P(A_{ij}) \frac{1}{W_L(N)} \frac{1}{C} \\ &\quad \times \int \prod_{m=1}^L dh_m e^{-\frac{N}{2} \sum_{ij} A_{ij} \text{tr}(h_i h_j) + iN \sum_i \text{tr}(h_i \phi_i)} \end{aligned} \quad (7.10)$$

The Hubbard-Stratonovich transformation yields a normalization factor $1/C$, with

$$C = \int \prod_{m=1}^L dh_m e^{-\frac{N}{2} \sum_{ij} A_{ij} \text{tr}(h_i h_j)}. \quad (7.11)$$

We remind at this point that

$$W_L(N) = \int \prod_{m=1}^L dh_m e^{-\frac{N}{2} \sum_{ij} (A^{-1})_{ij} \text{tr}(h_i h_j)}, \quad (7.12)$$

and $C \propto 1/\sqrt{\det A}$, $W_L(N) \propto 1/\sqrt{\det(A^{-1})}$ with the same coefficients which are independent on A (typically $(\pi/N)^{N^2 L/2} 2^{NL/2}$). Since

$$\det(A^{-1}) = 1/\det A, \quad (7.13)$$

the product $W_L(N)C \equiv B$ is independent on A , and we have

$$\begin{aligned}
& \frac{1}{B} \int \prod_{k=1}^L d\phi_k \frac{1}{N} \text{tr}(\phi_1 \dots \phi_L) \prod_{m=1}^L dh_m e^{i \sum_i N \text{tr}(h_i \phi_i)} \int dA_{ij} P(A_{ij}) e^{-\frac{N}{2} \sum_{ij} A_{ij} \text{tr}(h_i h_j)} \\
&= \frac{1}{B} \int \prod_{k=1}^L d\phi_k \frac{1}{N} \text{tr}(\phi_1 \dots \phi_L) \prod_{m=1}^L dh_m e^{i \sum_i N \text{tr}(h_i \phi_i)} \prod_{ij} \left[(1-p) + p e^{-\frac{N}{2} \text{tr}(h_i h_j)} \right] \\
&= \frac{1}{B} \int \prod_{k=1}^L d\phi_k \frac{1}{N} \text{tr}(\phi_1 \dots \phi_L) \prod_{m=1}^L dh_m e^{i \sum_i N \text{tr}(h_i \phi_i)} \prod_{ij} \exp \log \left[1 + p \left(-\frac{N}{2} \text{tr}(h_i h_j) + \dots \right) \right] \\
&= \frac{1}{B} \int \prod_{k=1}^L d\phi_k \frac{1}{N} \text{tr}(\phi_1 \dots \phi_L) \int \prod_{m=1}^L dh_m e^{i N \sum_i \text{tr}(h_i \phi_i)} e^S, \tag{7.14}
\end{aligned}$$

where B is a numerical constant, $S = S_0 + V$, and

$$S_0 = -\frac{pN}{2} \sum_{ij} \text{tr}(h_i h_j), \tag{7.15}$$

$$V = \frac{p(1-p)N^2}{8} \sum_{ij} [\text{tr}(h_i h_j)]^2 - \frac{p(1-p)(1-2p)N^3}{48} \sum_{ij} [\text{tr}(h_i h_j)]^3 + \dots \tag{7.16}$$

Up to this point, no approximation has been made. The S_0 term (7.15) corresponds to a fully-connected matrix with an additional factor p behind. If this term only was present, then, performing the inverse Hubbard-Stratonovich transformation and returning to the functional of the type (7.6), we would get $\xi(p) = p$, recovering the naive mean-field value $p_c = 1/4$ given by (6.6). Note that adding the first term of the expansion (7.16) is equivalent to replacing the true bimodal distribution $P(A_{ij})$ (6.1) with a Gaussian, preserving the first and the second moments:

$$P(A_{ij}) \longrightarrow \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(A_{ij} - A_0)^2}{2\sigma^2} \right), \tag{7.17}$$

where $A_0 = p$ and $\sigma^2 = p(1-p)$. This is similar to some variational approaches that approximate the true Hamiltonian S by a trial Gaussian Hamiltonian [BMPY91]. Another remark is that using the Gaussian distribution instead of the bimodal distribution (6.1) from the beginning and assuming that the matrix A belongs to the Gaussian Orthogonal Ensemble (GOE)

$$P(A)dA \propto e^{-\frac{N}{2} \text{tr}(A^2)} dA, \tag{7.18}$$

something could be done directly in the integral (7.9), passing to the distribution over the inverse matrix of A . Indeed, if X is the inverse matrix of A , the distribution of $X = A^{-1}$ can be computed¹ using from the distribution of the eigenvalues $x_i = \lambda_i^{-1}$, where λ_i are the eigenvalues of A :

$$P(X)dX \propto e^{-\frac{N}{2} \text{tr}(X^{-2})} |\det X|^{-(N+1)} dX. \tag{7.19}$$

The correction to p_c due to the whole series V (7.16) can be estimated as follows. The series given by the action S can be thought of as a Gaussian theory with the interaction

¹Private communication by Yan V. Fyodorov.

V . Since V contains an infinite number of terms, it is impossible to treat it perturbatively. Still, we can use a self-consistent nonperturbative approach reminiscent of the Feynman's variational principle [Fey55, GZRF85] in the field theory: as all the fields $\{h_i\}_{i=1,\dots,L}$ in Eq.(7.16) are equivalent, we assume that the average of $N\text{tr}(h_i h_j)$, $N^2\langle h_i h_j \rangle_{S_0} \equiv U$ is independent on (i, j) . Within the adopted mean-field approximation, the replacement $e^S = e^{S_0} e^{\langle V \rangle}$ is supposed, where

$$\langle V \rangle = \frac{p(1-p)N}{8} U \sum_{ij} \text{tr}(h_i h_j) - \frac{p(1-p)(1-2p)N}{48} U^2 \sum_{ij} \text{tr}(h_i h_j) + \dots \quad (7.20)$$

Resumming the series (7.20), we obtain the following self-consistent equation for the “propagator” U :

$$\frac{1}{U} = -\frac{2}{U} \log \left[1 - p + p \exp \left(-\frac{U}{2} \right) \right]. \quad (7.21)$$

The equation (7.21) yields $U = -2 \log \left[1 - \frac{1-1/\sqrt{e}}{p} \right]$. Hence, finally, we can write

$$S = -\frac{\xi(p)N}{2} \sum_{ij} \text{tr}(h_i h_j) \quad (7.22)$$

where

$$\xi(p) = \left(-2 \log \left[1 - \frac{1-1/\sqrt{e}}{p} \right] \right)^{-1}. \quad (7.23)$$

Substituting (7.23) into (6.9), we get an estimation for the critical value $p_c^* = 0.455$. Although the self-consistent approximation (7.20) leads to the correct direction of the shift of p_c from the naive mean-field value $p_c = 0.25$, it is rather crude, and we would like to have a better control on the estimated value. This can be achieved in a different approach, presented in the next section.

7.2 Combinatorial approach

As we have seen in the previous section, the field-theory-based estimation requires a use of uncontrollable, to a certain extent, approximation. In this section we describe a procedure for a detailed treatment of quenched disorder at the level of shortest arcs in the complete planar matching problem. In particular, we show how to get successive estimations to the value of the transition point via arcs expansion, explicitly calculating the contributions of shortest and next-to-shortest arcs, and treating the contribution of the rest in a mean-field manner.

To this purpose, we exploit the formulation of the problem in terms of Dyck paths, presented in the section 6.1, by combining exact combinatorial and mean-field techniques. The method is based on the observation that the arcs with smaller lengths are more likely to appear in the complete matching structure than those with higher lengths. Indeed, locally, in the complete matching configuration, the arc opened at i and closed at j corresponds to the part of a Dyck path, starting by an “up”-step \nearrow in position i and ending by the first “down”-step \searrow at the same height in position j , cf. Fig. 6.1(b). Hence, this random walk between i and j is a Dyck path itself, and the probability to find an arc connecting i and j , reads

$$P(i, j) = \frac{C_{(k-1)/2}}{2^{k+1}}, \quad (7.24)$$

where $k = j - i$; the nominator represents the total number of Dyck paths of length k , given by (6.2), and the denominator is the total number of possible random walks of this length.

From (7.24) we see that short links play an exceptional role in the formation of planar configurations: $P(i, j)$ is non-zero for odd k only, and a few first values are $P(i, i + 1) = 1/4$, $P(i, i + 3) = 1/16$, $P(i, i + 5) = 1/32$, *etc.* In particular, in the large L limit, about a half of all $L/2$ arcs are the shortest ones (“ S -arcs”) of length two (corresponding to the red peaks in the Fig. 6.1(b)).

7.2.1 First order of the expansion

In the work [P-3], the dominance of S -arcs is used to provide an estimate for the perfect-imperfect transition point by considering the following approximation:

$$\xi^{L/2}(p) = \underbrace{p^{L/4}}_{\text{long arcs}} \underbrace{\mathcal{P}_S^{(1)}(p)}_{S\text{-arcs}}, \quad (7.25)$$

that is, explicitly accounting for the correlations coming from the shortest arcs, and assuming that all longer arcs give a mean-field contribution $p^{L/4}$. Thus, the problem is reduced to placing $L/4$ shortest arcs on the line of L points, representing positions $(i, i + 1)$, each position being allowed or forbidden as dictated by the contact matrix values $A_{i, i+1}$, see Fig. 7.1. Note that since the arcs can not share the same node, the S -arcs can not occupy neighboring positions $(i, i + 1)$ and $(i + 1, i + 2)$ in such a placement.

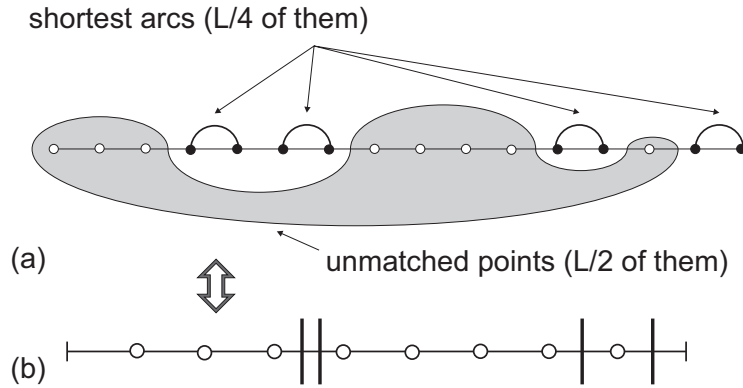


Figure 7.1: Computation of $B_S^{(1)}(p)$: (a) Selection of $L/4$ non-touching arcs on the set of L points ($L/2$ open dots remain unmatched); (b) the same problem reformulated as a partitioning of vertical segments (arcs) between open dots (unmatched points). A certain number of partitions are forbidden by the matrix of contacts A .

We can express $\mathcal{P}_S^{(1)}(p)$ as follows:

$$\mathcal{P}_S^{(1)}(p) = \frac{B_S^{(1)}(p)}{B_S^{(1)}(1)}, \quad (7.26)$$

where $B_S^{(1)}(p)$ is the number of ways to put uniformly $L/4$ S -arcs, allowed by the contact matrix A of density p , on a line of L points, according to the non-touching constraint. It is easy to see that in the fully-connected case $p = 1$ (all the positions are allowed), $B_S^{(1)}(1) = C_{3L/4}^{L/4}$, corresponding to the placement of $L/4$ objects among $L/4$ S -arcs and

$L/2$ unmatched vertices. In [P-3], the approximation $B_S^{(1)}(p) = C_{3pL/4}^{L/4}$ is used, assuming that each position of arcs $(i, i+1)$ out of L available *a priori* is allowed with probability p . Strictly speaking, it is not true, and gives only an upper-bound on the value of critical point computed at this level. Indeed, the density of ones in the diagonal $A_{i,i+1}$ is equal to p in the thermodynamic limit, but the ones are distributed independently, meaning that they may correspond to incompatible neighboring positions $(i, i+1)$ and $(i+1, i+2)$; at the same time, the “circle-and-stick” representation automatically incorporates the non-touching constraint. In the following, we derive a contribution of the S -arcs (or peaks in the Dyck path representation, see Fig. 6.1), via an exact procedure [P-4].

Let us first introduce some useful notations. We denote an allowed ($A_{i,i+1} = 1$) position for the S -arc by a square with a dot $(i, i+1) \equiv \square$, and the forbidden position ($A_{i,i+1} = 0$) by an empty square $(i, i+1) \equiv \square$. In the general, $p \neq 1$ case, the contact matrix partitions the length- L chain of all possible shortest arcs positions into pieces, representing the sequences of allowed positions, surrounded by forbidden ones: $(\dots \square \square \square \square \square \dots)$. Let us denote q_k the density (in the large L limit) of sequences of allowed positions of length k : $\underbrace{\square \square \square \dots \square}_k$. We have

$$q_k = p^k (1 - p)^2, \quad (7.27)$$

where the two factors $(1 - p)$ come from the forbidden positions nearest to the first and to the last positions of the sequence, and each factor p is the probability of an allowed position. It is easy to check that the overall density of allowed positions for the shortest arcs must be equal to p :

$$\sum_k k q_k = p(1 - p)^2(1 + 2p + 3p^2 + \dots) = p. \quad (7.28)$$

Given the physical sense of $B_S^{(1)}(p)$, we need to solve the following *constrained independent set* (CIS) problem: count the number of ways to distribute $L/4$ arcs so that they do not touch each other, on the ensemble of these allowed partitions. For each sequence it means that if a certain position is chosen, other arcs can not be placed in the neighboring positions, even if these last are allowed by the contact matrix A . Note, however, that this *global* CIS problem is reduced to a set of *local* ones on the sequences with densities q_k : since they are separated by at least one forbidden position, the distribution of S -arcs happens independently on each sequence.

As a by-product, here comes a non-trivial strict bound on the value of p_c . It is easy to see that for a piece of length k , at most $\lfloor (k+1)/2 \rfloor \equiv r_k$ positions can be occupied under the non-touching constraint. Therefore, the maximum fraction of shortest arcs is

$$\sum_k \left\lfloor \frac{k+1}{2} \right\rfloor q_k = p(1+p)(1-p)^2(1 + 2p^2 + 3p^4 + 4p^6 + \dots) = \frac{p}{1+p}. \quad (7.29)$$

Since we need to place at least $L/4$ arcs, we immediately get that $p_c > 1/3$. It coincides with the lower bound discussed in the section 6.4.

Now, the solution of the local CIS problem is easy, and is given by

$$R_{m,k} = C_{k-m+1}^m, \quad (7.30)$$

where $R_{m,k}$, the number of ways to put m S -arcs on the allowed sequence of positions of length k (see [P-4] for details of the derivation). Given the solution of the local CIS

problem (7.30), it is easy to construct the solution of the global problem. Let us introduce a generating function for a piece of length k :

$$Q_k(s) = \sum_{m=0}^{r_k} R_{m,k} s^m. \quad (7.31)$$

Then the generating function for the whole chain of L a priori available positions for the S -arcs reads

$$Q(s) = \prod_{k=1}^L (Q_k(s))^{Lq_k}, \quad (7.32)$$

or, explicitly,

$$Q(s) = (1+s)^{Lp(1-p)^2} (1+2s)^{Lp^2(1-p)^2} (1+3s+s^2)^{Lp^3(1-p)^2} (1+4s+3s^2)^{Lp^4(1-p)^2} \dots \quad (7.33)$$

Since we want to place $L/4$ shortest arcs, we are interested in the coefficient behind the $s^{L/4}$: this is exactly the quantity $B_S^{(1)}(p)$. This coefficient is given by the integration of $Q(s)/s^{L/4}$ around zero:

$$B_S^{(1)}(p) = \frac{1}{2\pi i} \oint ds \exp [L((1-p)^2 f_s(p) - 1/4 \log s)], \quad (7.34)$$

where

$$f_s(p) = \sum_{k=1}^L p^k \log \left(\sum_{m=0}^{[(k+1)/2]} C_{k-m+1}^m s^m \right). \quad (7.35)$$

Using explicit re-summation, this result can be written as

$$f_s(p) = \sum_{k=1}^L p^k \log \left(\frac{(1 + \sqrt{1+4s})^{k+2} - (1 - \sqrt{1+4s})^{k+2}}{2^{k+2} \sqrt{1+4s}} \right). \quad (7.36)$$

Each term in the sum (7.35) is decreasing, so in numerical calculations we can approximate this function by partial sums to some order k_0 . The integral (7.34) can be treated by the steepest descent method. The saddle-point equation reads

$$(1-p)^2 \frac{\partial f_s(p)}{\partial s} = \frac{1}{4s}. \quad (7.37)$$

Given the solution s^* of the equation (7.37), one gets the expression for $B_S^{(1)}(p)$:

$$B_S^{(1)}(p) = \exp [L((1-p)^2 f_{s^*}(p) - 1/4 \log s^*)]. \quad (7.38)$$

Approximating the large deviation function (7.35) by partial sums up to the fifteenth order, and combining with (7.26), (7.25) and (6.9), we get a fast convergence to the prediction of the critical point $p_c^* = 0.3376$, see Fig. 7.2, providing an expected shift to a lower value from the result $p_c^* = 0.35$, found at this level of expansion if one assumes a mean-field approximation $B_S^{(1)}(p) = C_{3pL/4}^{L/4} [P-3]$.

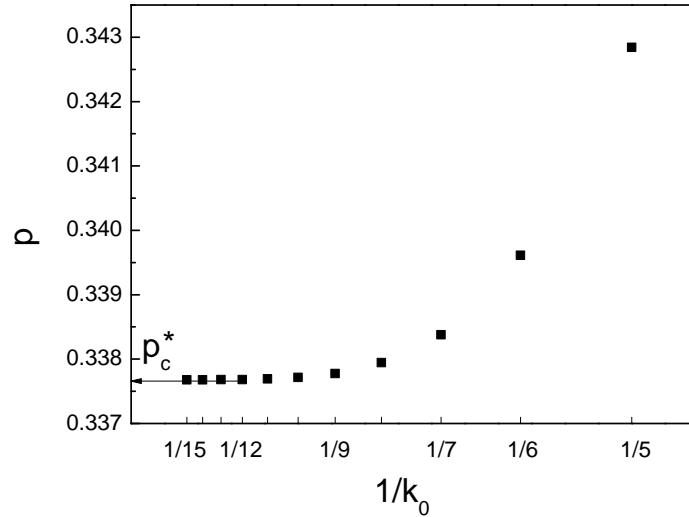


Figure 7.2: Estimation of the prediction p_c^* for the critical value p_c at the first order of arcs expansion. Each point represents a corresponding estimation when (7.35) is approximated by a partial sum up to the lengths k_0 . The estimations demonstrate a fast convergence with k_0 to the value $p_c^* = 0.3376$.

7.2.2 Beyond the first order

The first-order estimation can be systematically improved by considering correlations arising from the higher-length arcs. At the second order, the idea is to use not only the information from the diagonal $A_{i,i+1}$ of the contact matrix, but also from the diagonal $A_{i,i+3}$, i.e. to take into account the constraints on the placement of the shortest (S), length-2 arcs, that come from the placement of the length-4, or next-to-shortest (NS) arcs. Therefore, we can write, as previously

$$\xi^{L/2}(p) = \underbrace{p^{L/4}}_{\text{longer arcs}} \underbrace{\mathcal{P}_S^{(2)}(p)}_{S\text{-arcs}}, \quad (7.39)$$

but now the influence of the $L/16$ NS -arcs is accounted in the factor $\mathcal{P}_S^{(2)}(p)$ representing $L/4$ S -arcs. As before, we will compute the contributions of the shortest arcs under the correlations arising from the placement of the NS -arcs, and treat the contribution arising from the longer arcs in a mean-field manner.

In other words, the problem is now reduced to the placement of both S and NS arcs that respect the constraints imposed by the contact matrix A . Obviously, the placement of arcs of one type introduces additional constraints on the placement of those of other type. First, some of the places will become forbidden because of the non-crossing constraints. Second, if we are interested in the complete matching configurations, placing a length-4 arc automatically means placing a length-2 arc underneath. Therefore, our goal is to place altogether $L/16$ NS -arcs, each covering a S -arc, $L/4 - L/16 = 3L/16$ remaining S -arcs and $L - 4 \times L/16 - 2 \times 3L/16 = 3L/8$ unmatched vertices. The placements are subject to the non-touching constraints. Proceeding in the same way as in the previous section, we can write

$$\mathcal{P}_S^{(2)}(p) = \frac{B_S^{(2)}(p)}{B_S^{(2)}(1)}, \quad (7.40)$$

where the factor $B_S^{(2)}(p)$ represents the contributions of the S -arcs under the correlations arising from the presence of the NS -arcs. The denominator of the product (7.40) is given by a multinomial coefficient

$$B_S^{(2)}(1) = \frac{\frac{5L}{8}!}{\frac{L}{16}! \frac{3L}{16}! \frac{3L}{8}!} = C_{5L/8}^{L/16} C_{9L/16}^{3L/16}. \quad (7.41)$$

The multinomial coefficient has the following physical sense: it counts the number of ways to place length-4 constructions, length-2 arcs and unmatched vertices when all the places are available by the contact matrix. It can be factorized into two binomial coefficients that have the following sense: first place $L/16$ length-4 constructions among $L/16 + 3L/16 + 3L/8 = 5L/8$ objects, and then place $3L/16$ remaining S -arcs among $3L/16 + 3L/8 = 9L/16$ available places.

The computation of $B_S^{(2)}(p)$ follows the similar ideas as for the $B_S^{(1)}(p)$. First of all, the problem is localized on independent subsequences of a special form, and the local problem of distribution of S and NS arcs is solved on these subsequences. Unfortunately, it is hard to write a closed combinatorial formula, similar to (7.30), since the solution inside each block depends on the distribution of forbidden and allowed positions. However, if we truncate the series at some length k_0 (as it has been done at the first order of arcs expansion), these solutions can be computed for each sequence via explicit enumeration, see [P-4] for details. Again, going in maximum length up to $k_0 = 15$, the global problem demonstrates a fast convergence to the value $p_c^* = 0.3743$ which is very close to the value $p_c = 0.379$ found in numerical simulation, see Fig. 7.3.

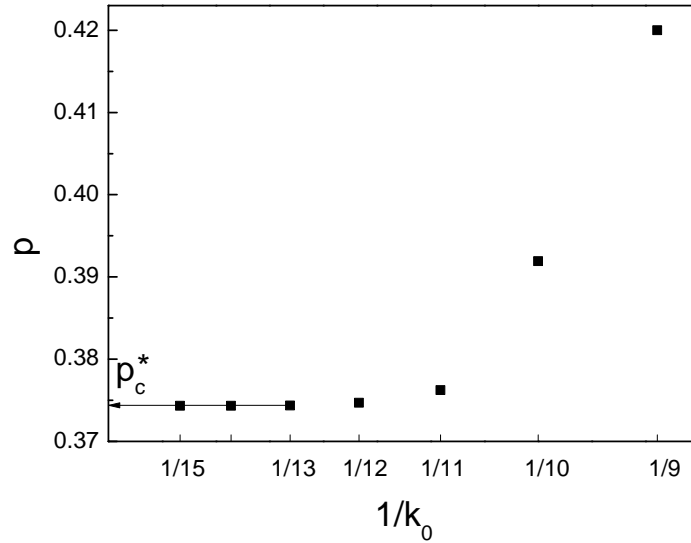


Figure 7.3: Estimation of the prediction p_c^* for the critical value p_c at the second order of arcs expansion. Each point represents a corresponding estimation when the lengths up to k_0 of independent subsequences are taken into account. The estimations demonstrate a fast convergence with k_0 to the value $p_c^* = 0.3743$.

In principle, this estimation can be improved further by considering the higher arcs contribution. For example, for the next order, including length-6, or next-to-next-to-shortest (NNS) arcs, we can write

$$\xi^{L/2}(p) = \underbrace{p^{L/4}}_{\text{longer arcs}} \underbrace{\mathcal{P}_S^{(3)}(p)}_{\text{S-arcs}}, \quad (7.42)$$

where now the S -arcs are placed according to the restrictions imposed by $L/32$ NNS -arcs and $L/16$ NS -arcs. A subtlety here is that if we are interested in the fully-matched configurations, each of the NNS -arcs may hide either 2 S -arcs, either a nested structure of NS and S arcs. Hence, placing $L/32$ NNS -arc on a certain position (allowed with the probability $p(1 - (1 - p^2)^2)$) automatically means placing $L/64$ NS -arcs and $3L/64$ S -arcs; at the same time, $L/16 - L/64 = 3L/64$ of NS -arcs are still remaining outside the placed NNS -arcs, and $L/4 - 3L/64 - 3L/64 = 5L/32$ of S -arcs are still remaining outside both NNS and NS . We have to place them altogether with $5L/16$ unmatched vertices. Therefore, if we write, as usual,

$$\mathcal{P}_S^{(3)}(p) = \frac{B_S^{(3)}(p)}{B_S^{(3)}(1)}, \quad (7.43)$$

the factor $B_S^{(3)}(1)$ will be given by

$$B_S^{(3)}(1) = \frac{\frac{35L}{64}!}{\frac{L}{32}! \frac{3L}{64}! \frac{5L}{32}! \frac{5L}{16}!} = C_{35L/64}^{L/32} C_{33L/64}^{3L/64} C_{15L/32}^{5L/32}. \quad (7.44)$$

The calculation of the factor $B_S^{(3)}(p)$ would involve, as previously, the partitioning of the problem to a set of local CIS problems. The solution to the global problem could then be obtained by imposing that the overall number of S -arcs is fixed to $L/4$.

Chapter 8

Application: optimal folding in RNA secondary structures

Aiming at the application to the toy models of random RNA, in this chapter we establish the relation between the zero-temperature perfect-imperfect matching transition in random planar diagrams, and a temperature-dependent molten-glass transition in random RNAs, widely discussed in the literature. We find that the perfect-imperfect phase transition point lies on the critical line, separating molten and glassy regions, and coincides with the freezing transition at zero temperature [P-3]. Therefore, while on the corresponding phase diagram the molten phase exists in both perfect and imperfect regions, the glassy phase is present only in the region with gaps. As a second application, we introduce two new models involving explicit representation of an instance of the problem as a string of letters, and study numerically the perfect-imperfect phase transition in these models [P-4].

8.1 Connection to the molten-glass transition

The investigation of thermodynamic properties of RNA secondary structures has been addressed in a number of papers [BH99, PPRT00, BH02, KMM02, MPRT02, LW06, HT06, DW07, HIL⁺14]. Many of them provided numerical and analytical evidence for the existence of a low-temperature glassy phase. In [BH02] it was shown that in the high-temperature phase the system remains in the molten phase, characterized by a homopolymer-like behavior. In the molten phase the disorder is irrelevant, and the binding matrix elements A_{ij} can be replaced by some effective value α . Carrying out the two-replica calculation, the authors were able to show that the system exhibits a phase transition from a high-temperature regime, in which the replicas are independent, to a low-temperature phase, in which the disorder is relevant and replicas are strongly coupled. The authors numerically characterized the transition to a glassy phase by imposing a pinch between two bases and measuring the corresponding energy cost.

Several other works [KMM02, MPRT02] used an alternative, the so-called ε -coupling method, to investigate the nature and the scaling laws of the glassy phase, observing the effect of typical excitations imposed by a bulk perturbation. The authors argued that for the models with non-degenerate ground states, the low-temperature phase is not marginal, but is governed by a scaling exponent, close to $\theta = 1/3$. The field-theory approach to the freezing transition has been developed in [LW06, DW07]. The explicit numerical studies of the specific heat demonstrate that molten-glass transition is only a

fourth order phase transition [PPRT00].

Regardless of particular details of models considered in all these works, it is clear that the existence of the glassy phase is possible only in a sufficiently disordered and frustrated system. Besides the planarity constraint, shared by all simple models of random RNA, the Bernoulli model is described by a unique disorder parameter, p , that controls the density of allowed contacts. In this model, the appearance of the glassy phase is impossible above a certain threshold, p^* . Indeed, it is well-known that for $p = 1/2$ (corresponding to an effective alphabet $c = 2$), there is no transition to the glassy phase at all, and the system remains always in the molten phase [PPRT00, BH02]. Below, we present arguments, supporting the hypothesis that p^* is equal to the critical value of perfect-imperfect matching transition, p_c .

To identify the dependence of the molten-glass transition temperature on the effective alphabet (defined as $c = 1/p$), we follow the procedure suggested in [BH02]. In the high-temperature regime the disorder is irrelevant (this corresponds to a homopolymer-like behavior in polymer language) and one can put $A_{ij} = \alpha$. In this regime the free energy of the chain of length L scales linearly with L , up to a logarithmic correction, which is just the logarithm of the power-law multiplier in the Catalan number (6.2) enumerating all possible structures: $F(L, T) = f(T)L - (3T/2) \ln L$, where $f(T)$ is some (non-universal) function of the temperature. In particular, the energy cost of imposing a bond connecting two monomers at distance $L/2$ from each other equals in the high temperature regime

$$\Delta F(L, T) = F(L, T) - 2F(L/2, T) = \frac{3}{2}T \ln \frac{L}{4}. \quad (8.1)$$

The violation of this behavior indicates [BH02] the appearance of the glassy phase. This fact can be used to detect the transition temperature in the Bernoulli model. Namely, we use the following fit for $\Delta F(L, T)$ (where $F(L, T)$ is to be determined via recursion relations (6.4))

$$\Delta F(L, T) = a(T) \ln L + b(L), \quad (8.2)$$

and plot the T -dependence of $a(T)$, see Fig. 8.1. We interpret the deviation of the $a(T)$ from the high-temperature value $3T/2$ as appearance of the glass transition. Note that the logarithmic fit (8.2) for the free energy does not give a correct asymptotics at low temperatures (indeed, the true asymptotics is known to include power-law and logarithm-squared terms [HT06]).

As it follows from Fig. 8.1, the expected behavior (8.1) is indeed observed at high temperatures, and is violated at a certain temperature T_c . Following [BH02] we identify this regime change with the molten-glass transition. We see that with the increase of p , the critical temperature T_c shifts to lower values, approaching zero for some $0.35 < p^* < 0.5$. At low temperatures, the numerical computations become very time consuming, leading to the loss of precision in the vicinity of p^* . However, it seems that the hypothesis $p^* = p_c$ still holds: the sequences corresponding to $p > p_c$ remain in the molten phase, the pinching free energy (8.2) has the same dependence even for very low temperatures.

The results presented in this section suggest the generic phase diagram shown in the Fig. 8.2 for the Bernoulli model of random RNA chains. The perfect-imperfect transition at zero temperature, separates two matching regions: with and without gaps. In the previous chapters, we have proved analytically the existence of the transition from the perfect matching region to the imperfect one, and provided estimates for the values of the transition point, p_c . Using the exact dynamical programming algorithm (6.4), we

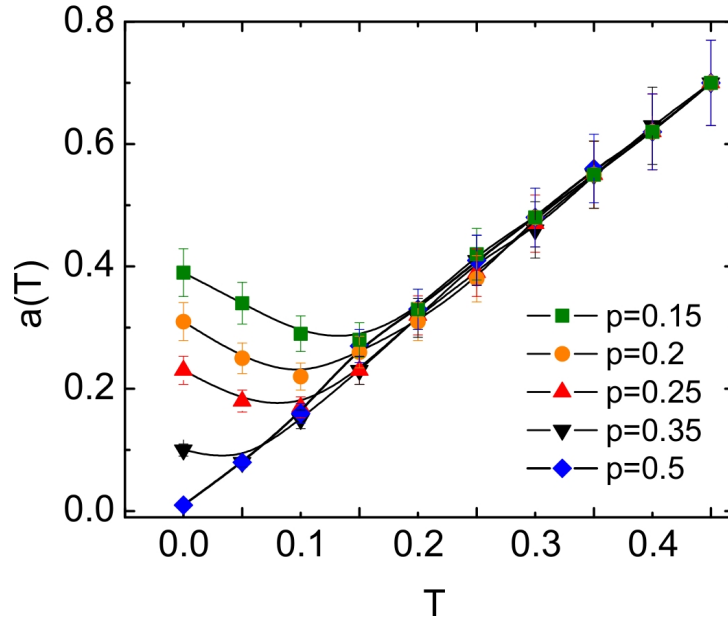


Figure 8.1: The dependence of the coefficient $a(T)$ in (8.2) on the temperature for $p = 0.15, 0.2, 0.25, 0.35, 0.5$. For $p > p^*$ ($0.35 < p^* < 0.5$), the coefficient $a(T)$ seems to follow the $a(T) = 3T/2$ law, typical for the molten phase, up to very low temperatures. For $p < p^*$, the $a(T)$ -dependence deviates from the high-temperature behavior at some temperature, which we identify as a critical temperature of transition to the glassy phase. The data points are averaged over 10000 samples.

found this critical value to be $p_c \approx 0.379$, highlighted by a thick dashed line (B-C) in Fig. 8.2. The previous studies have been mostly concentrated on the description of the finite-temperature molten-glass transition for a sufficiently frustrated model with a fixed alphabet (corresponding to a fixed p in the Bernoulli model). An example of such a phase transition point is marked by a thin dashed line in the Fig. 8.2, and corresponds to an intensively studied case of the 4-letter alphabet ($p = 0.25$). The ensemble of critical points for different values of p gives a critical curve (A-B) in the (T, p) plane.

The computational cost increases drastically for temperatures close to zero (and, hence, in the vicinity of p_c), and the recursive relations (6.3) are no more applicable. However, we can still try to carry out the analysis of the pinching free energy $\Delta F(L, T)$ at zero temperature, using the exact dynamical programming algorithm (6.4). Indeed, the glassy phase does not exist if $\Delta F(\infty, 0) = 0$. This happens for $p > p^*$, where p^* is defined as the density of constrains, for which the critical temperature is zero: $T_c(p^*) = 0$. The corresponding plot is shown in the inset of Fig. 8.2. According to (6.5), the pinching free energy (8.1) decreases with growth of L in the imperfect matching phase, while increases (with growth of L) in the perfect matching regime. Hence, the value of p^* in the large L limit can be identified as a crossing point of $\Delta F(L, 0)$ curves for different L . The crossing point for $L = 1000$ and $L = 2000$ is indeed found to be very close to the value $p_c = 0.379$, strongly supporting the hypothesis $p^* = p_c$. The aforementioned results indicate that the critical curve $T_c(p)$ crosses zero at the critical value p_c . Hence, the perfect-imperfect transition point seems to lie at the critical line, separating molten and glassy regions, and coincides with its limiting $T = 0$ value. We see that although the glassy phase exists only in the region where the gaps are present, the molten phase lies in both, perfect and imperfect, matching regions.

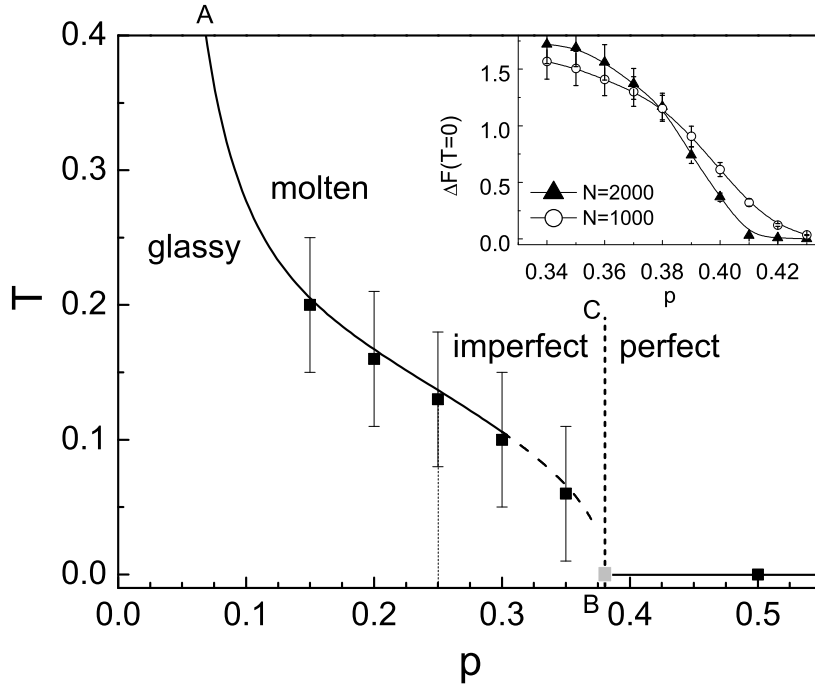


Figure 8.2: Main figure: the phase diagram of Bernoulli model on the (T, p) plane. The data points correspond to the critical temperature T_c of the molten-glass transition for different values $p = 0.15, 0.2, 0.25, 0.3, 0.35, 0.5$. A 4-letter alphabet ($p = 0.25$), is highlighted by a thin dashed line. The critical curve (A-B) separates glassy and molten phases. We conjecture that at zero temperature, the endpoint B, giving p^* , coincides with the critical point p_c for the perfect-imperfect transition. The thick dashed line (B-C) separates the perfect and imperfect matching cases. The glassy phase lies entirely inside the region, characterized by gaps. Inset: an evidence for the conjecture $p^* = p_c$. Study of the pinching free energy $\Delta F(L, T)$ at zero temperature. In the limit of large L , the glassy phase is absent for $p > p^*$, characterized by $\Delta F(\infty, 0) = 0$. The point p^* can be identified as a crossing point for different $\Delta F(L, 0)$ curves, presented here for $L = 1000$ and $L = 2000$, and its value is found to be very close to $p_c = 0.379$. The data points are averaged over 1000 samples.

Because of the one-parameter dependence, the Bernoulli model is probably the simplest model for modelling the secondary structure of the RNA, that captures the essential physical properties of the process. Being applied to the studies of the thermodynamic properties of random RNAs, it provides some enlightenment on the nature of molten-glass transition at zero temperature. In principle, one could generalize the present approach to investigate more sophisticated and realistic models of the RNA secondary structure, for example, by introducing the minimal allowed hairpin length [PPRT00, BH02, KMM02], taking into accounts the pseudoknots [VOZ05] and different binding probabilities [VOZ05, MPRT02]. In the next section, however, we will instead discuss the models that allow for an explicit sequence representation.

8.2 Other models for random sequences with non-integer alphabets

Although it is clear that, given the pairing complementarity rules, one can always build a contact matrix from a given primary sequence, the opposite in general is not true. Indeed, in the Bernoulli model (6.1), each element of the matrix is generated *independently* according to the probability distribution (6.1), hence, it lacks the transitivity: even if the elements A_{ij} and A_{jk} appear to be equal to one in the contact matrix, the element A_{ik} might be zero.

It would be interesting to understand whether there is a way to construct explicit random primary sequence that could model the primary sequences with non-integer alphabets. In the context of the phase transition described in the section 6.3, we have seen that there is a critical value $p_c \approx 0.379$ of the bond formation probability that separates the regions of optimal and non-optimal structures; this critical probability corresponds to the critical alphabet $c_{cr} \approx 2.64$ in this generalized primary sequence setting. In this section, we address the following questions: i) Is it possible to construct explicitly a random sequence with transitive or partially transitive matching rules that would correspond to a non-integer alphabet c , i.e. have a density $p = 1/c$ of ones in the contact matrix, generated according to this sequence? ii) Do these sequences exhibit an analogous critical behavior as the Bernoulli model with the same parameter p , and what is the relation to the behavior of the Bernoulli model?

8.2.1 Construction of the non-integer alphabets

For the models of random sequences, we consider a set of monomers of different types, that we will call A, B, C, etc. Perhaps the most natural way to think of the non-integer alphabet $2 < c < 3$ is to consider three types of monomers: A, B and C, mixed together. For the sake of simplicity, we will assume that the links can be established between the monomers of the same type, A-A, B-B and C-C. It is clear that if three types of monomers are distributed randomly and independently along the sequence, this corresponds to an alphabet $c = 3$. However, the effective non-integer alphabet $c < 3$ can be modeled if one assumes that the distribution of monomers along the chain is correlated. Suppose that starting from the first randomly chosen monomer, each next monomer in the sequence is generated according to the Markov-like process, with the probabilities that depend on the monomer at the previous step:

	A	B	C
A	$1 - 2\epsilon$	ϵ	ϵ
B	ϵ	$1 - 2\epsilon$	ϵ
C	ϵ	ϵ	$1 - 2\epsilon$

This probability matrix is chosen to be symmetric with respect to all monomer types. Each monomer type appears in subsequences unless it is changed to another type: $(\dots A A A B B B B A C C C \dots)$. Using the word reduction algorithm, described in the section 6.4 (see Fig. 6.7), we see that without any loss of generality, the repeated monomers can be matched along the chain; this way, each subsequence of a certain type of even or odd length is reduced to one or zero monomers of this type, respectively: $(\dots B A A A C \dots) \rightarrow (\dots B A C \dots)$.

The variation of the parameter ϵ from 0 to $1/3$ then gives a sequence that corresponds to an effective alphabet c in a range from 1 to 3. The relation between ϵ and c can be estimated as follows:

$$c = \left(\frac{1}{\epsilon} - 2 \right)^{2\epsilon} \frac{1}{1 - 2\epsilon}. \quad (8.3)$$

The rational behind this estimation is based on the concept of Shannon information entropy [Sha48]. The entropy rate of this markovian sequence is given by

$$S = - \sum_{a=A,B,C} P(a) \sum_{b=A,B,C} P(b | a) \log P(b | a), \quad (8.4)$$

where $P(a) = 1/3$ is an *a priori* probability for the monomer of a certain type, and $P(b | a)$ is a conditional probability that the monomer of the type a is followed by the monomer of the type b ; this probability is given by the probability matrix of the considered Markov process. On the other hand, if one assumes that the sequences constructed in this way are described by an effective alphabet with c equivalent monomers, we simply have

$$S = - \sum_{a=1}^c \hat{P}(a) \log \hat{P}(a) \quad (8.5)$$

with $\hat{P}(a) = 1/c$. The combination of (8.4) and (8.5) gives us the relation (8.3). Thus constructed alphabet will be referred to as the “correlated” alphabet.

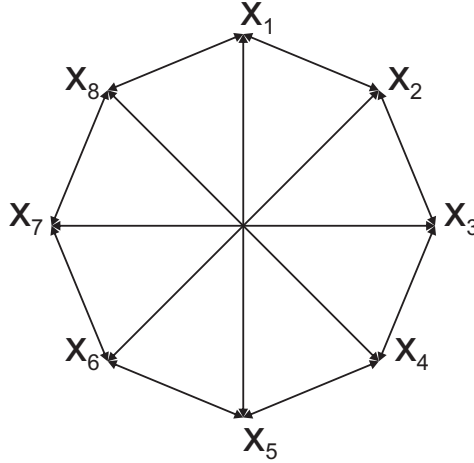


Figure 8.3: An example of the matching rules in the $(8, 3)$ rational alphabet model. In this representation, a double arrow between the monomers of types X_i and X_j means that they can potentially form a bond in the matching structure.

Another model that can be suggested for non-integer alphabets can be obtained using the observation that each non-integer alphabet c can be approximated by a rational fraction $c = P/Q$. Imagine a random polymer with P different monomer types X_1, \dots, X_P , but now allow each of them to bind only with Q other monomer types. The complementary rules can be depicted as a P -polygon with $Q - 2$ additional links, where each link means a possible matching between two monomers, see Fig. 8.3 for an example with

$P = 8$ and $Q = 3$. The “commutation relations” for the monomers read

$$\{X_i, X_{i \pm j}\} = 1 \text{ for } 1 \leq j \leq [Q/2], \quad (8.6)$$

$$\{X_i, X_i\} = 1 \text{ if } Q \text{ and } P \text{ odd}, \quad (8.7)$$

$$\{X_i, X_{i+P/2}\} = 1 \text{ if } Q \text{ odd and } P \text{ even}, \quad (8.8)$$

$$\{X_i, X_{i+j}\} = 0 \text{ otherwise}, \quad (8.9)$$

where $\{X_i, X_k\}$ represents a presence (one) or absence (zero) of possible matching between the two monomers X_i, X_k ; the periodic condition $X_{i+P} \equiv X_i$ is understood. We will call this model a (P, Q) “rational alphabet”. Note that by construction this alphabet is non-transitive. A particularity of this model is that there is an infinite number of ways to represent c as a fraction. Let us call P^* and Q^* as the minimal P and Q that give $c = P/Q$. Then $P = nP^*$ and $Q = nQ^*$ for an arbitrary integer n give the same value of c , although involving a different number of monomer types. In the thermodynamic limit $L \rightarrow \infty$ it will make no difference since the density of ones in the contact matrix will be exactly $p = Q/P$, but for finite L it may result in a different behaviors for the models with $c = P^*/Q^*$, $c = 2P^*/2Q^*$, etc. In order to minimize this effect, we place ourselves in the context of the urn model, in which the number of monomers of different sorts in the sequence are restricted to be equal.

8.2.2 Perfect matching transition for non-integer alphabets

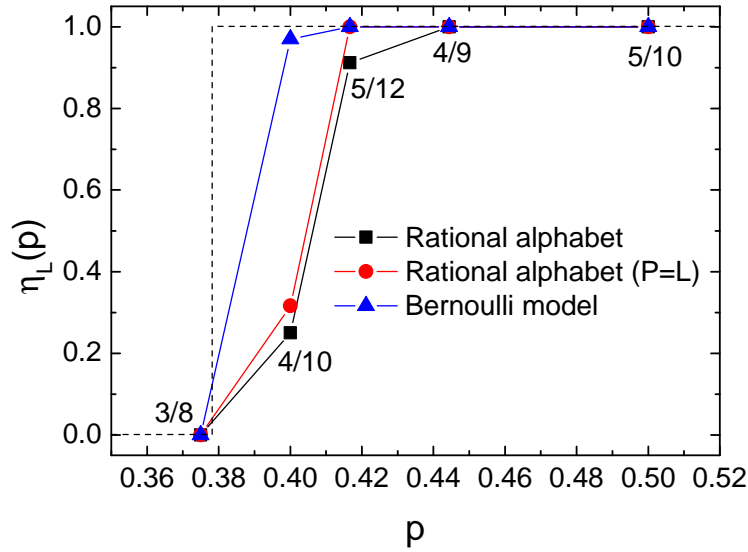


Figure 8.4: The fraction of perfect matchings $\eta_L(p)$ as a function of the density p of possible contacts in the model of (P, Q) rational alphabet (squares, the respective values of $p = P/Q$ are indicated on the plot), fluctuation-free Bernoulli model (rational alphabet model with $P = L$, circles) and Bernoulli model (triangles). The simulations have been performed for $L = 2000$ and averaged over 10000 instances.

We have investigated the behavior of both correlated and rational alphabets with respect to the perfect matching transition. To this purpose, we start by drawing random sequences corresponding to a particular c , then construct the contact matrix A according

to the matching rules defined in both models, and, finally solve the matching problem for each instance by the dynamical programming algorithm.

Surprisingly, we have not observed any transition with variation of c for the model of the correlated alphabet. In fact, if $c > 2$ (or $\epsilon > 0.1135$) in this model, there is always a non-zero fraction of sequences that do not allow for the complete matching solutions. A possible reason for this is that due to the structure of the sequence, the matching on each subsequence is easy, but then the sequence is reduced to the primary structure of length $\mathcal{O}(L)$ which corresponds effectively to the alphabet $c = 3$, while it is known that for this alphabet (for $p = 1/3$) in general there is no solution to the perfect matching problem [Vla13].

On the contrary, the rational alphabet model clearly exhibits a critical behavior in the matching problem. In the Fig. 8.4, we present the numerical results for the fraction, $\eta_L(p)$, of contact matrices that allow perfect matchings for different p . To avoid the sensitivity on the value of P due to the finite size effects, we have chosen simple test values $p = Q/P$ with similar P in the range $P \in [8, 12]$. The number of perfect matching in these points are compared to the special case of the limit $P = L$, i.e. when all the L randomly distributed monomers in the chain are distinct, however being able to match $Q = pL$ other monomers in the chain. This limit corresponds to the fluctuations-free Bernoulli model, in which every line of the matrix A contains *exactly* pL of ones, without fluctuations of order \sqrt{L} that appear in the model defined by (6.1). The rational alphabets give similar predictions, which are however very different with respect to the predictions of the Bernoulli model. This difference illustrates the “positive” role of fluctuations of the number of contacts in the Bernoulli matrix from the viewpoint of the matching problem.

Chapter 9

Perspectives

This final chapter of the part II is devoted to some open questions and perspectives for the problems that emerge from the planar-matching results.

Key results

There is a particular interest for studying the planar matching models considered in this part. On one hand, their statistical properties are fully determined by only a few parameters: one for the Bernoulli model (p) and for the model of correlated alphabet (c , or ϵ), or two for the rational alphabet model (P and Q). On the other hand, these disordered models exhibit a non-trivial critical behavior. Although the matching problem can be solved by the dynamical programming algorithm with a polynomial complexity $\mathcal{O}(L^3)$, where L is the size of an instance of the problem), the analytical estimation of the critical point is hard due to the quenched nature of the disorder. We have seen that the analytical investigation has led to a number of interesting problems and results, including an expansion procedure that benefits from the observation that the arcs of small length play an exceptional role in the complete matching structures. A key ingredient that makes the problem solvable is the fact that the global constraint satisfaction problem can be reduced to a set of local ones that are easier to solve. The developed method hence provides an insight into the fundamental structural properties of the fully-matched configurations.

Future directions

As it often happens, finding a solution to some previous problems bring up new questions and ideas. In this section, we will describe some open questions related to the topics discussed in this part, and shape the directions of the future research.

Strict bounds and second-moment method

In the sections 6.4 and 7.2, we have mentioned the bounds to the phase transition point in the Bernoulli model, $p_c < 1/2$ and $p_c > 1/3$. It would be interesting to obtain rigorous bounds on this critical value, using methods from the optimization science. One of the powerful methods that could be used to this purpose is the *second-moment method*

[ANP05]. It starts from the fact that for any integer-valued variable X , the following identities hold:

$$\Pr[X > 0] \leq \mathbb{E}[X], \quad (9.1)$$

$$\Pr[X > 0] \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}. \quad (9.2)$$

In order to introduce useful notations for the second-moment method, let us rewrite the derivation of the naive mean-field value (it will follow from the identity (9.1)). Let X be the number of solutions to the perfect matching problem, and let τ be a randomly chosen configuration of arcs out of the fully-connected ensemble of $C_{L/2}$ possible configurations. We will say that the configuration τ (or an arc (ij) in this configuration) is SAT if it satisfies the constraints imposed by the contact matrix A , and we note $\mathbb{1}_\tau = 1$ if τ is SAT, and $\mathbb{1}_\tau = 0$ otherwise. We have

$$\mathbb{E}[X] = \sum_{\tau} \mathbb{E}[\mathbb{1}_\tau] = \sum_{\tau} \Pr[\tau \text{ is SAT}] = \sum_{\tau} \prod_{(ij) \in \tau} \Pr[(ij) \text{ is SAT}] = C_{L/2} p^{L/2}. \quad (9.3)$$

The first equality follows from the linearity of the expectation value. We see from (9.1) that $\Pr[X > 0]$ is non-zero only if $p > 1/4$, which is the naive mean-field value.

In order to use the expression (9.2), one needs to compute the second moment of X , and hence to account for the statistical correlations between a pair of configurations. As we have seen in the section 6.4, the probability that two randomly chosen configurations τ and ρ are SAT is given by $p^{L/2} p^{L/2} p^{-n_{\tau \cap \rho} L/2}$, where $n_{\tau \cap \rho}$ is a fraction of common arcs between τ and ρ . Precisely, we have

$$\mathbb{E}[X^2] = \sum_{\tau} \sum_{\rho} \mathbb{E}[\mathbb{1}_\tau \mathbb{1}_\rho] = \sum_{\tau, \rho} \Pr[\tau \& \rho \text{ are SAT}] = \sum_{\tau, \rho} p^{L/2} p^{L/2} \mathbb{E}[p^{-n_{\tau \cap \rho} L/2}]. \quad (9.4)$$

Remark 1. Ultimately, we will be interested in p that will lead to $\Pr[X > 0]$. Hence, in the first approximation, we could try to find the lower bound to the right hand-side of the equation (9.2), in particular provided by the upper bound on the $\mathbb{E}[X^2]$. Directly from (9.4), we see that we can use the dominance of S (shortest) arcs in the fully-matching configurations and approximate $n_{\tau \cap \rho}$ by $n_{\tau \cap \rho}^S$, where $n_{\tau \cap \rho}^S$ is the number of common S arcs in the configurations τ and ρ .

We will follow the approach to the two-replica problem, suggested in [BH02]. First of all, let us rewrite the expression (9.4) by introducing an auxiliary probability \tilde{p} for the common arcs:

$$\mathcal{G}(L+1, \tilde{p}) = \sum_{\tau} \sum_{\rho} p^{L/2} p^{L/2} \mathbb{E}[(\tilde{p})^{-n_{\tau \cap \rho} L/2}]. \quad (9.5)$$

Obviously, we have $\mathbb{E}[X^2] = \mathcal{G}(L+1, \tilde{p} = p)$. In what follows, we will use the following observation: the common arcs, shared by two configurations τ and ρ , form a (non fully-matched) valid matching structure by themselves. Therefore, we rearrange the sum in (9.5) in a different way: first, we sum over all possible structures of common arcs; for a given structure of common arcs, we sum over all remaining possibilities of

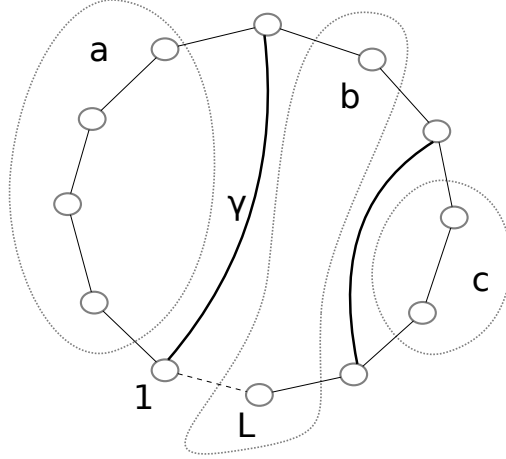


Figure 9.1: Illustration of the equation (9.6) for the two-replica problem. The summation runs over the common link structure (thick lines) of two configurations, denoted by γ . Inside each bubble (in this example, three of them are drawn: a , b and c) we count the number pairs of non-overlapping complete-matching pairings.

arcs placement under the constraint that *no new common arcs are created*. The last sum factorizes into the independent contribution of loops, or bubbles, and each contribution depends solely on the number of sites inside the bubble, see Fig. 9.1. Moreover, since we are interested in the complete matching configurations, we need to count only the fully-matched contributions of the bubbles. Taking into account all these considerations, we finally get

$$\mathcal{G}(L+1, \tilde{p}) = \sum_{\substack{\gamma \\ \text{common arcs}}} p^L (\tilde{p})^{-|\gamma|} \prod_{\substack{i \\ \text{bubbles}}} \mathcal{Q}(l_i + 1), \quad (9.6)$$

where $|\gamma|$ is the number of common arcs in the structure γ , l_i is the number of sites in the bubble i , and $\mathcal{Q}(l+1)$ has a sense of the number of pairs (τ_l, ρ_l) of fully-matching configurations in the bubble with l sites that do not share common arcs:

$$\mathcal{Q}(l+1) = \sum_{\tau_l, \rho_l} \mathbb{1}[n_{\tau_l \cap \rho_l} = 0]. \quad (9.7)$$

Remark 2. If we are interested in the computation of $n_{\tau \cap \rho}^S$ only, the equation (9.6) is greatly simplified: since the structures of common arcs involve only the S arcs and thanks to the word reduction property (see example in the Fig. 6.7), there is always exactly one bubble left.

From the definitions (9.5) and (9.7) it is clear that

$$p^L \mathcal{Q}(L+1) = \mathcal{G}(L+1, \tilde{p} \rightarrow \infty). \quad (9.8)$$

Another obvious remark is that for $\tilde{p} = 1$, we have

$$\mathcal{G}(L+1, \tilde{p} = 1) = \mathbb{E}[X]^2 = (C_{L/2})^2 p^L. \quad (9.9)$$

Using two consistency relations (9.8) and (9.9), it is possible to write a self-consistency equation for the Laplace-transformed function $\mathcal{G}(L+1, \tilde{p})$ (the details with straightforward correspondences are presented in [BH02]):

$$\widehat{\mathcal{G}}(z, \tilde{p}) = \frac{1}{p} \widehat{Q} \left(\frac{z}{p} - p^2(\tilde{p})^{-1} \widehat{\mathcal{G}}(z, \tilde{p}) \right). \quad (9.10)$$

This equation is valid for any value of \tilde{p} , and in particular the value $\tilde{p} = 1$ can be used. The z-transform of $\mathcal{G}(L+1, \tilde{p} = 1) = \mathbb{E}[X]^2$ is given by

$$\widehat{\mathbb{E}[X]^2} = c_0 \text{Li}_3 \left[\frac{4p}{z} \right], \quad (9.11)$$

where $c_0 = 2^3/\pi$, and the trilogarithm $\text{Li}_3[x]$ is defined as

$$\text{Li}_3[x] = \sum_{k=1}^{\infty} \frac{x^k}{k^3}. \quad (9.12)$$

The further progress can be done by analyzing the singularities in the equations (9.10) and (9.11). Since the singularity of $\widehat{\mathbb{E}[X]^2}$ occurs at the point $z_0 = 4p$, we are able to locate the singularity ξ_0 of the function $\widehat{Q}(\xi)$: $\xi_0 = 4 - pc_0\zeta(3)$. Finally, this value is used to determine the singularity z_1 of $\widehat{\mathcal{G}}(z, \tilde{p})$ for arbitrary \tilde{p} :

$$\frac{z_1}{p} - p(\tilde{p})^{-1} c_0 \zeta(3) = 4 - pc_0 \zeta(3), \quad (9.13)$$

or

$$z_1(\tilde{p}) = 4p + p^2 c_0 \zeta(3) \left[\frac{1}{\tilde{p}} - 1 \right]. \quad (9.14)$$

Performing the inverse Laplace transform, we have $\mathcal{G}(L+1, \tilde{p}) \propto z_1^L(\tilde{p})$, and, finally,

$$\mathbb{E}[X^2] = \mathcal{G}(L+1, \tilde{p} = p) \propto (4p + p(1-p)c_0\zeta(3))^L. \quad (9.15)$$

We see that $\mathbb{E}[X^2]$ is exponentially larger than $\mathbb{E}[X]^2$, and (9.2) always gives a trivial relation, meaning that the vanilla second-moment method fails for this problem. In fact, this result could be predicted in advance: since we know that the S -arcs represent a half of all links, it is clear that we will always have non-vanishing overlaps $n_{\tau \cap \rho}^S$ and $n_{\tau \cap \rho}$ between two configurations τ and ρ , and hence an additional exponentially small factor in (9.4). Still, it is known [ANP05] that the situation can be improved by considering, instead of X , a new variable Y (still related to the existence of solution of the optimization problem), which could correspond in our case to reweighed configurations. It would be interesting to find a proper weight that would lead to a non-trivial bound from (9.2). An insight could come from our knowledge for the structure of the complete-matching configurations: a possible idea would be to assign an exponential weight to the arcs depending on their length in order to eliminate the dominating contribution of the small arcs to the value of the overlap $n_{\tau \cap \rho}$. However, this question remains open, and is left for a future work.

Narayana statistics

Let us have a second look on the first order of expansion procedure in the computation of $B_S^{(1)}(p)$ in the section 7.2. When solving the problem of placing $L/4$ S -arcs on the line of L positions, we have assumed that the short arcs are uncorrelated apart from the non-overlap constraints. In real arc structures it is not the case: indeed, the total number of available structures with exactly k shortest arcs in the absence of disorder is known to be given by the so-called Narayana number $N(2L, k)$ [Deu99], and this is the quantity that should in principle be used instead of $C_{3L/4}^{L/4}$. Still, there is no danger for the general expansion procedure. Indeed, the correlations between short arcs are induced by the positions of longer ones, so assuming them to be uncorrelated seems to be a natural first approximation, while the correlations will arise naturally as one takes into account arcs of length 3, 5, *etc.* Still, the estimation of the critical value p_c can be made more accurate already at the first order of the expansion procedure, correctly accounting for the correlations between the S -arcs, corresponding to the peaks in the Dyck path mapping, see Fig. 6.1. As far as we know, these correlations have not yet been quantified; it would be useful to compute the correlation function $C(i_1, i_2)$ of two peaks located at the points i_1 and i_2 , and understand more in detail the Narayana statistics of the Dyck paths.

Localization in Cayley graphs

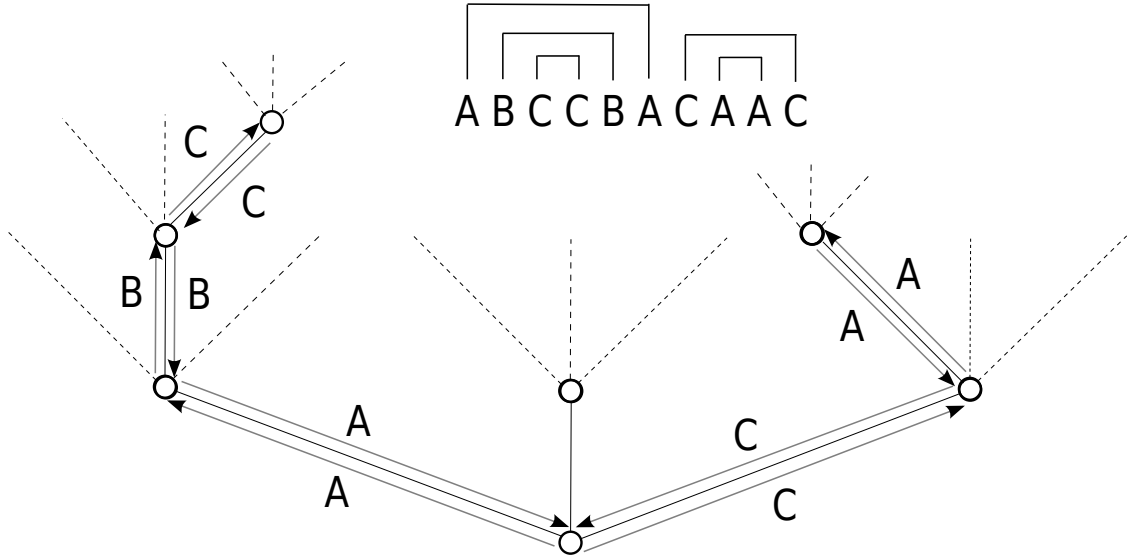


Figure 9.2: Example of the mapping of the planar matching problem to a walk on a Cayley tree with $c = 3$ branches. A sequence allows for a perfect matching if there exist a walk that returns to the origin at the step L .

Apart from the mapping to the Brownian excursions, the RNA-type matching, considered in the sections 6.2 and 8.2, allows for another mapping: a random walk on a Cayley tree, represented in the Fig. 9.2. It is constructed as follows. Associate to each letter of the alphabet a vector (representing a particular branch of the tree at each generation), and use the following rule: for each letter in the sequence, a move can be done forward or backward only in the direction of the corresponding branch. It is clear that the complete-matching solution corresponds to a walk that returns to the origin after L

steps. In a random setting, a generalization of these rules would correspond to a random walk on such a tree; obviously, under these simple rules, a simple random walk has no chance to return to the origin. Starting from this point, it would be interesting to understand whether this tree structure can be modified, for example by “gluing” together some of the branches in order to get a localization close to the origin; see e.g. [VNB00] for an example of a modification of this kind. This procedure is interesting in itself, because such a Cayley graph corresponds to a structural representation of some locally-free groups with partially commutative relations between the generators. Preliminary results on the path counting on such topologies indicate an existence of a localized behavior due to the “entropic trap” in the origin. But this is a starting point for a completely different ongoing story, and we finish at this point.

List of publications

- [P-1] Andrey Y. Lokhov, M. Mézard, and Lenka Zdeborová. Dynamic message-passing equations for models with unidirectional dynamics. *Phys. Rev. E* **91**, 012811, 2015.
- [P-2] Andrey Y. Lokhov, M. Mézard, Hiroki Ohta, and Lenka Zdeborová. Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Phys. Rev. E* **90**, 012801, 2014.
- [P-3] Andrey Y. Lokhov, Olga V. Valba, Mikhail V. Tamm, and Sergei K. Nechaev. Phase transition in random planar diagrams and RNA-type matching. *Phys. Rev. E* **88**, 052117, 2013.
- [P-4] Andrey Y. Lokhov, Olga V. Valba, Sergei K. Nechaev, and Mikhail V. Tamm. Topological transition in disordered planar matching: combinatorial arcs expansion. *J. Stat. Mech.*, P12004, 2014.

Bibliography

- [AB02] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.
- [ABD⁺14a] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall’Asta, Alessandro Ingrosso, and Riccardo Zecchina. The zero-patient problem with noisy observations. *arXiv:1408.0907*, 2014.
- [ABD⁺14b] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall’Asta, Alejandro Lage-Castellanos, and Riccardo Zecchina. Bayesian Inference of Epidemics on Networks via Belief Propagation. *Physical Review Letters*, 112(11):118701, 2014.
- [ABDZ13a] F. Altarelli, A. Braunstein, L. Dall’Asta, and R. Zecchina. Large deviations of cascade processes on graphs. *Physical Review E*, 87(6):062115, 2013.
- [ABDZ13b] Fabrizio Altarelli, Alfredo Braunstein, L Dall’Asta, and Riccardo Zecchina. Optimizing spread dynamics on graphs by message passing. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(09):P09011, 2013.
- [AG75] Aleksei A Abrikosov and Lev P Gorkov. *Methods of quantum field theory in statistical physics*. Courier Dover Publications, 1975.
- [Ald01] David J Aldous. The $\zeta(2)$ limit in the random assignment problem. *Random Structures & Algorithms*, 18(4):381–418, 2001.
- [AM91] R. M. Anderson and R. M. May. *Infectious diseases of humans (Vol. 1)*. Oxford University Press, 1991.
- [AM12] Erik Aurell and Hamed Mahmoudi. Dynamic mean-field and cavity methods for diluted Ising systems. *Physical Review E*, 85(3):031119, 2012.
- [ANP05] Dimitris Achlioptas, Assaf Naor, and Yuval Peres. Rigorous location of phase transitions in hard optimization problems. *Nature*, 435(7043):759–64, 2005.
- [AS04] David Aldous and J Michael Steele. The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, pages 1–72. Springer, 2004.
- [BA99] A Barabási and R Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

- [BBV08] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge University Press Cambridge, 2008.
- [BCKM96] Jean-Philippe Bouchaud, Leticia Cugliandolo, Jorge Kurchan, and Marc Mézard. Mode-coupling approximations, glass theory and disordered systems. *Physica A: Statistical Mechanics and its Applications*, 226(3-4):243–273, 1996.
- [BCKM98] Jean-Philippe Bouchaud, Leticia F. Cugliandolo, Jorge Kurchan, and Marc Mézard. Out of Equilibrium Dynamics in Spin-Glasses and Other Glassy Systems. In *Spin Glasses and Random Fields*, pages 161–223. World Scientific, 1998.
- [Bet35] H. A. Bethe. Statistical Theory of Superlattices. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 150(871):552–575, 1935.
- [BG96] C. Berrou and A. Glavieux. Near optimum error correcting coding and decoding: turbo-codes. *IEEE Transactions on Communications*, 44(10):1261–1271, 1996.
- [BH99] R. Bundschuh and T. Hwa. RNA Secondary Structure Formation: A Solvable Model of Heteropolymer Folding. *Physical Review Letters*, 83(7):1479–1482, 1999.
- [BH02] R. Bundschuh and T. Hwa. Statistical mechanics of secondary structures formed by random RNA sequences. *Physical Review E*, 65(3):031903, 2002.
- [BH13] Dirk Brockmann and Dirk Helbing. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342(6164):1337–42, 2013.
- [BHJ09] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362, 2009.
- [BIPZ78] E. Brézin, C Itzykson, G Parisi, and J B Zuber. Planar diagrams. *Communications in Mathematical Physics*, 59(1):35–51, 1978.
- [Bir99] Giulio Biroli. Dynamical TAP approach to mean field glassy systems. *Journal of Physics A: Mathematical and General*, 32(48):8365–8388, 1999.
- [BLLS71] Dick Bedeaux, Katja Lakatos-Lindenberg, and Kurt E. Shuler. On the Relation between Master Equations and Random Walks and Their Solutions. *Journal of Mathematical Physics*, 12:2116, 1971.
- [BLM⁺06] S Boccaletti, V Latora, Y Moreno, M Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006.
- [BM97] Jean-Philippe Bouchaud and Marc Mézard. Universality classes for extreme-value statistics. *Journal of Physics A: Mathematical and General*, 30(23):7997–8015, 1997.

- [BMPY91] J-P Bouchaud, M Mezard, G Parisi, and J S Yedidia. Polymers with long-range self-repulsion: a variational approach. *Journal of Physics A: Mathematical and General*, 24(17):L1025–L1030, 1991.
- [Bol80] Béla Bollobás. A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs. *European Journal of Combinatorics*, 1(4):311–316, 1980.
- [BSS08] Mohsen Bayati, Devavrat Shah, and Mayank Sharma. Max-Product for Maximum Weight Matching: Convergence, Correctness, and LP Duality. *IEEE Transactions on Information Theory*, 54(3):1241–1251, 2008.
- [BY86] K Binder and AP Young. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Reviews of Modern Physics*, 58(4), 1986.
- [BZ99] A. Barrat and R. Zecchina. Time scale separation and heterogeneous off-equilibrium dynamics in spin models over random graphs. *Physical Review E*, 59(2):R1299–R1302, 1999.
- [CC06a] Michael Chertkov and Vladimir Chernyak. Loop calculus in statistical physics and information science. *Physical Review E*, 73(6):065102, 2006.
- [CC06b] Michael Chertkov and Vladimir Y Chernyak. Loop series for discrete statistical models on graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(06):P06009–P06009, 2006.
- [Cd11] Cesar Henrique Comin and Luciano da Fontoura Costa. Identifying the starting point of a spreading process in complex networks. *Phys. Rev. E*, 84:56105, 2011.
- [CDMM03] S. Cocco, O. Dubois, J. Mandler, and R. Monasson. Rigorous Decimation-Based Construction of Ground Pure States for Spin-Glass Models on Random Lattices. *Physical Review Letters*, 90(4):047205, 2003.
- [CFL09] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, May 2009.
- [CH10] Reuven Cohen and Shlomo Havlin. *Complex networks: structure, robustness and function*. Cambridge University Press, 2010.
- [CHS93] A. Crisanti, H. Horner, and H. J. Sommers. The spherical p-spin interaction spin-glass model. *Zeitschrift für Physik B Condensed Matter*, 92(2):257–271, 1993.
- [CK93] L. Cugliandolo and J. Kurchan. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Physical Review Letters*, 71(1):173–176, 1993.
- [CK94] L F Cugliandolo and J Kurchan. On the out-of-equilibrium relaxation of the Sherrington-Kirkpatrick model. *Journal of Physics A: Mathematical and General*, 27(17):5749–5772, 1994.

- [CK08] Leticia F Cugliandolo and Jorge Kurchan. The out-of-equilibrium dynamics of the SherringtonKirkpatrick model. *Journal of Physics A: Mathematical and Theoretical*, 41(32):324018, 2008.
- [Col88] Sidney Coleman. *Aspects of symmetry: selected Erice lectures*. Cambridge University Press, 1988.
- [Coo01] A.C.C. Coolen. Statistical mechanics of Recurrent Neural networks II: Dynamics. In *Handbook of Biological Physics Vol 4*, pages 597–662. Elsevier Science, 2001.
- [CPS10] Claudio Castellano and Romualdo Pastor-Satorras. Thresholds for Epidemic Spreading in Networks. *Physical Review Letters*, 105(21):218701, 2010.
- [CS88] A. Crisanti and H. Sompolinsky. Dynamics of spin systems with randomly asymmetric bonds: Ising spins and Glauber dynamics. *Physical Review A*, 37(12):4865–4874, 1988.
- [CS92] A. Crisanti and H. J. Sommers. The spherical p-spin interaction spin glass model: the statics. *Zeitschrift für Physik B Condensed Matter*, 87(3):341–354, 1992.
- [CS93] A. C. C. Coolen and D. Sherrington. Dynamics of fully connected attractor neural networks near saturation. *Phys. Rev. Lett.*, 71(23):3886–3889, 1993.
- [CS94] A C C Coolen and D Sherrington. Order-parameter flow in the SK spin glass. I. Replica symmetry. *Journal of Physics A: Mathematical and General*, 27(23):7687–7707, 1994.
- [CS95] A. Crisanti and H.-J. Sommers. Thouless-Anderson-Palmer Approach to the Spherical p-Spin Spin Glass Model. *Journal de Physique I*, 5(7):805, 1995.
- [Cug03] L F Cugliandolo. Dynamics of glassy systems. In *Slow Relaxations and nonequilibrium dynamics in condensed matter*, pages 367–521. Springer Berlin / Heidelberg, 2003.
- [CWW⁺08] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security*, 10(4):1–26, 2008.
- [dAT78] J R L de Almeida and D J Thouless. Stability of the Sherrington-Kirkpatrick solution of a spin glass model. *Journal of Physics A: Mathematical and General*, 11(5):983–990, 1978.
- [DCLN07] Ian Dobson, Benjamin A Carreras, Vickie E Lynch, and David E Newman. Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization. *Chaos (Woodbury, N.Y.)*, 17(2):026103, 2007.
- [De 78] C. De Dominicis. Dynamics as a substitute for replicas in systems with quenched random impurities. *Physical Review B*, 18(9):4913–4919, 1978.
- [Der81] Bernard Derrida. Random-energy model: An exactly solvable model of disordered systems. *Physical Review B*, 24(5):2613–2626, 1981.

- [Deu99] Emeric Deutsch. Dyck path enumeration. *Discrete Mathematics*, 204(1-3):167–202, 1999.
- [dG68] P G de Gennes. Statistics of branching and hairpin helices for the dAT copolymer. *Biopolymers*, 6(5):715–29, January 1968.
- [DG06] C De Dominicis and I Giardinà. *Random fields and spin glasses: a field theory approach*. Cambridge University Press, 2006.
- [DGG⁺87] Alan Demers, Mark Gealy, Dan Greene, Carl Hauser, Wes Irish, John Larson, Sue Manning, Howard Sturgis, Dan Swinehart, Don Woods, Scott Shenker, and Doug Terry. Epidemic Algorithms for Replicated Database Maintenance. *ACM Proceedings of the sixth annual ACM Symposium on Principles of distributed computing.*, pages 1–12, 1987.
- [DGM08] S. Dorogovtsev, A. Goltsev, and J. Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics*, 80(4):1275–1335, 2008.
- [DGZ87] B Derrida, E Gardner, and A Zippelius. An Exactly Solvable Asymmetric Neural Network Model. *Europhysics Letters (EPL)*, 4(2):167–173, 1987.
- [DK64] D. J. Daley and D. G. Kendall. Epidemics and rumors. *Nature*, 204(1118), 1964.
- [DK65] D. J. Daley and D. G. Kendall. Stochastic rumours. *IMA Journal of Applied Mathematics*, 1(1):42–55, 1965.
- [DM87] C De Dominicis and P Mottishaw. Replica symmetry breaking in weak connectivity systems. *Journal of Physics A: Mathematical and General*, 20(18):L1267–L1273, 1987.
- [DNAW00] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 03(01n04):87–98, 2000.
- [DSS97] Deepak Dhar, Prabodh Shukla, and James P Sethna. Zero-temperature hysteresis in the random-field Ising model on a Bethe lattice. *Journal of Physics A: Mathematical and General*, 30(15):5259–5267, 1997.
- [Dur10] Rick Durrett. Some features of the spread of epidemics and information on a random graph. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10):4491–8, 2010.
- [DW07] Francois David and Kay Wiese. Systematic Field Theory of the RNA Glass Transition. *Physical Review Letters*, 98(12):128102, 2007.
- [DZT13] Wenxiang Dong, Wenyi Zhang, and Chee Wei Tan. Rooting out the rumor culprit from suspects. In *2013 IEEE International Symposium on Information Theory*, pages 2671–2675. IEEE, July 2013.
- [EA75] S F Edwards and P W Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965–974, 1975.

- [ER60] P Erdős and A Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5:17–61, 1960.
- [FA86] Y Fu and P W Anderson. Application of statistical mechanics to NP-complete problems in combinatorial optimisation. *Journal of Physics A: Mathematical and General*, 19(9):1605–1620, 1986.
- [FC12] Vincenzo Fioriti and Marta Chinnici. Predicting the sources of an outbreak with a spectral technique. *arXiv:1211.2333*, 2012.
- [Fey55] R. Feynman. Slow Electrons in a Polar Crystal. *Physical Review*, 97(3):660–665, 1955.
- [FGJ10] S. Funk, E. Gilad, and V.A.A. Jansen. Endemic disease , awareness, and local behavioural response. *Journal of Theoretical Biology*, 264(2):501–509, 2010.
- [FGWJ09] Sebastian Funk, Erez Gilad, Chris Watkins, and Vincent A A Jansen. The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16):6872–6877, 2009.
- [FH93] Konrad H Fischer and John A Hertz. *Spin Glasses*. Cambridge University Press, 1993.
- [Fri99] Ehud Friedgut. Necessary and sufficient conditions for sharp thresholds and the k-sat problem. *J. Amer. Math. Soc.*, 12(20):1017–1054, 1999.
- [FSJ10] Sebastian Funk, Marcel Salathé, and Vincent A A Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society, Interface*, 7(50):1247–56, 2010.
- [Gal62] R. Gallager. Low-density parity-check codes. *IEEE Transactions on Information Theory*, 8(1):21–28, 1962.
- [Gal02] S. Galam. Minority opinion spreading in random geometry. *The European Physical Journal B*, 25(4):403–406, 2002.
- [Gar85] E. Gardner. Spin glasses with p-spin interactions. *Nuclear Physics B*, 257:747–765, 1985.
- [GD09] I Garg and N Deo. A random matrix approach to RNA folding with interaction. *Pramana*, 73(3):533–541, 2009.
- [GDGP⁺02] X. Guardiola, A. Díaz-Guilera, C. Pérez, A. Arenas, and M. Llas. Modeling diffusion of innovations in a social network. *Physical Review E*, 66(2):026121, 2002.
- [Gla63] Roy J. Glauber. Time-Dependent Statistics of the Ising Model. *Journal of Mathematical Physics*, 4(2):294, 1963.
- [GM84] D.J. Gross and M Mezard. The simplest spin glass. *Nuclear Physics B*, 240(4):431–452, 1984.

- [GM88] H. Gutfreund and M. Mezard. Processing of Temporal Sequences in Neural Networks. *Physical Review Letters*, 61(2):235–238, 1988.
- [GS14] Alberto Guggiola and Guilhem Semerjian. Minimal contagious sets in random regular graphs. *arXiv:1407.7361*, 2014.
- [Gue03] Francesco Guerra. Broken Replica Symmetry Bounds in the Mean Field Spin Glass Model. *Communications in Mathematical Physics*, 233(1):1–12, 2003.
- [GZRF85] S. N. Gorshkov, A. V. Zabrodin, C. Rodriguez, and V. K. Fedyanin. Feynman’s variational principle for a polaron in a magnetic field. *Theoretical and Mathematical Physics*, 62(2):205–210, 1985.
- [Har74] T. E. Harris. Contact Interactions on a Lattice. *The Annals of Probability*, 2(6):969–988, 1974.
- [Har01] A. Hartmann. Comment on Glassy Transition in a Disordered Model for the RNA Secondary Structure. *Physical Review Letters*, 86(7):1382–1382, February 2001.
- [HCCS05] J. Hatchett, I. Castillo, A. Coolen, and N. Skantzos. Dynamical Replica Analysis of Disordered Ising Spin Systems on Finitely Connected Random Graphs. *Physical Review Letters*, 95(11):117204, 2005.
- [Het94] Herbert W Hethcote. A Thousand and One Epidemic Models. In *Frontiers in Mathematical Biology*, pages 504–515. Springer Berlin, 1994.
- [Het00] Herbert W. Hethcote. The Mathematics of Infectious Diseases. *SIAM Review*, 42(4):599–653, 2000.
- [Hig96] Paul Higgs. Overlaps between RNA Secondary Structures. *Physical Review Letters*, 76(4):704–707, 1996.
- [HIL⁺14] G. N. Hayrapetyan, F. Iannelli, J. Lekscha, V. F. Morozov, R. R. Netz, and Y. Sh. Mamasakhlisov. Reentrant Melting of RNA with Quenched Sequence Randomness. *Physical Review Letters*, 113(6):068101, 2014.
- [HK14] Haiping Huang and Yoshiyuki Kabashima. Dynamics of asymmetric kinetic Ising systems revisited. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(5):P05020, 2014.
- [HP14] Kathleen E. Hamilton and Leonid P. Pryadko. Tight lower bound for percolation threshold on a quasi-regular graph. *arXiv:1405.0050*, 2014.
- [HT06] S. Hui and L.-H. Tang. Ground state and glass transition of the RNA secondary structure. *The European Physical Journal B*, 53(1):77–84, 2006.
- [HWC⁺04] J P L Hatchett, B Wemmenhove, I Pérez Castillo, T Nikolettopoulos, N S Skantzos, and A C C Coolen. Parallel dynamics of disordered Ising spin systems on finitely connected random graphs. *Journal of Physics A: Mathematical and General*, 37(24):6201–6220, June 2004.

- [Kas61] P.W. Kasteleyn. The statistics of dimers on a lattice. *Physica*, 27(12):1209–1225, 1961.
- [KCAL05] R. Kinney, P. Crucitti, R. Albert, and V. Latora. Modeling cascading failures in the North American power grid. *The European Physical Journal B*, 46(1):101–107, 2005.
- [KFL01] F.R. Kschischang, B.J. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001.
- [KGV83] S Kirkpatrick, C D Gelatt, and M P Vecchi. Optimization by simulated annealing. *Science (New York, N.Y.)*, 220(4598):671–80, 1983.
- [KH08] Martin Kiemes and Heinz Horner. Dynamics of an Ising spin glass on the Bethe lattice. *Journal of Physics A: Mathematical and Theoretical*, 41(32):324017, 2008.
- [Kik51] Ryoichi Kikuchi. A Theory of Cooperative Phenomena. *Physical Review*, 81(6):988–1003, 1951.
- [Kim04] Allan J. Kimmel. Rumors and the Financial Marketplace. *Journal of Behavioral Finance*, 5(3):134–141, 2004.
- [KM27] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics. part I. *Proceedings of the Royal society of London*, 115(700), 1927.
- [KM11] Yashodhan Kanoria and Andrea Montanari. Majority dynamics on trees and the dynamic cavity method. *The Annals of Applied Probability*, 21(5):1694–1748, 2011.
- [KMM02] F Krzakala, M Mézard, and M Müller. Nature of the glassy phase of RNA secondary structure. *Europhysics Letters (EPL)*, 57(5):752–758, 2002.
- [KN10] Brian Karrer and M. E. J. Newman. Message passing approach for general epidemic models. *Physical Review E*, 82(1):016101, 2010.
- [KNZ14] Brian Karrer, M. E. J. Newman, and Lenka Zdeborova. Percolation on sparse networks. *arXiv:1405.0483*, 2014.
- [Kos98] Michael Kosfeld. Rumours and Markets. *Journal of Economic Literature*, 1998.
- [KPV93] J. Kurchan, G. Parisi, and M. A. Virasoro. Barriers and metastable states as saddle points in the replica approach. *Journal de Physique I*, 3(8):1819–1838, 1993.
- [KR03] P. Krapivsky and S. Redner. Dynamics of Majority Rule in Two-State Interacting Spin Systems. *Physical Review Letters*, 90(23):238701, 2003.
- [KR07] Eben Kenah and James Robins. Second look at the spread of epidemics on networks. *Physical Review E*, 76(3):036113, 2007.

- [KS78] S Kirkpatrick and D Sherrington. Infinite-ranged models of spin-glasses. *Physical Review B*, 17(11):4384, 1978.
- [KS87] I. Kanter and H. Sompolinsky. Mean-field theory of spin-glasses with finite coordination number. *Physical Review Letters*, 58(2):164–167, 1987.
- [KS98] Y Kabashima and D Saad. Belief propagation vs. TAP for decoding corrupted messages. *Europhysics Letters (EPL)*, 44(5):668–674, 1998.
- [KYK12] Boris Kholodenko, Michael B Yaffe, and Walter Kolch. Computational approaches for analyzing information flow in biological networks. *Science signaling*, 5(220):re1, 2012.
- [Lan03] Sergei K Lando. *Lectures on generating functions*. Amer Mathematical Society, 2003.
- [LCS96] S N Laughton, A C C Coolen, and D Sherrington. Order-parameter flow in the SK spin-glass: II. Inclusion of microscopic memory effects. *Journal of Physics A: Mathematical and General*, 29(4):763–786, 1996.
- [LGLDG03] Mateu Llas, Pablo M. Gleiser, Juan M. López, and Albert Díaz-Guilera. Nonequilibrium phase transition in a model for the propagation of innovations among economic agents. *Physical Review E*, 68(6):066101, 2003.
- [LP09] László Lovász and Michael D Plummer. *Matching theory*. American Mathematical Society, 2009.
- [Luc12] Andrew Lucas. Exact mean field dynamics for epidemic-like processes on heterogeneous networks. *arXiv:1206.6294*, 2012.
- [LW06] Michael Lässig and Kay Wiese. Freezing of Random RNA. *Physical Review Letters*, 96(22):228101, 2006.
- [Mö3] M. Müller. Statistical physics of RNA folding. *Physical Review E*, 67(2):021914, 2003.
- [MC08] A Mozeika and A C C Coolen. Dynamical replica analysis of processes on finitely connected random graphs: I. Vertex covering. *Journal of Physics A: Mathematical and Theoretical*, 41(11):115003, 2008.
- [MC09a] Kazushi Mimura and A C C Coolen. Parallel dynamics of disordered Ising spin systems on finitely connected directed random graphs with arbitrary degree distributions. *Journal of Physics A: Mathematical and Theoretical*, 42(41):415001, 2009.
- [MC09b] A Mozeika and A C C Coolen. Dynamical replica analysis of processes on finitely connected random graphs: II. Dynamics in the Griffiths phase of the diluted Ising ferromagnet. *Journal of Physics A: Mathematical and Theoretical*, 42(19):195006, 2009.
- [Meh04] Madan Lal Mehta. *Random matrices*. Academic press, 2004.

- [Mil11] Joel C Miller. A note on a paper by Erik Volz: SIR dynamics in random networks. *Journal of mathematical biology*, 62(3):349–58, 2011.
- [MM09] Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, January 2009.
- [MP85] M. Mézard and G. Parisi. Replicas and optimization. *Journal de Physique Lettres*, 46(17):771–778, 1985.
- [MP86a] M. Mézard and G. Parisi. A replica analysis of the travelling salesman problem. *Journal de Physique*, 47(8):1285–1296, 1986.
- [MP86b] M Mézard and G Parisi. Mean-Field Equations for the Matching and the Travelling Salesman Problems. *Europhysics Letters (EPL)*, 2(12):913–918, 1986.
- [MP87a] M Mézard and G Parisi. Mean-Field Theory of Randomly Frustrated Systems with Finite Connectivity. *Europhysics Letters (EPL)*, 3(10):1067–1074, 1987.
- [MP87b] M. Mézard and G Parisi. On the solution of the random link matching problems. *Journal de Physique*, 48(9):1451–1459, 1987.
- [MP01] M. Mézard and G Parisi. The Bethe lattice spin glass revisited. *The European Physical Journal B*, 20(2):217–233, 2001.
- [MPRT02] Enzo Marinari, Andrea Pagnani, and Federico Ricci-Tersenghi. Zero-temperature properties of RNA secondary structures. *Physical Review E*, 65(4):041919, 2002.
- [MPSV02] Y. Moreno, R. Pastor-Satorras, and A. Vespignani. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B*, 26(4):521–529, 2002.
- [MPWZ02] R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina. Coloring Random Graphs. *Physical Review Letters*, 89(26):268701, 2002.
- [MPZ02] M Mézard, G Parisi, and R Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297(5582):812–5, 2002.
- [MR95] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995.
- [MR05] Andrea Montanari and Tommaso Rizzo. How to compute loop corrections to the Bethe approximation. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(10):P10011–P10011, 2005.
- [MRT03a] A. Montanari and F. Ricci-Tersenghi. Microscopic Description of Aging Dynamics: Fluctuation-Dissipation Relations, Effective Temperature, and Heterogeneities. *Physical Review Letters*, 90(1):017203, 2003.

- [MRT03b] Andrea Montanari and Federico Ricci-Tersenghi. Aging dynamics of heterogeneous spin models. *Physical Review B*, 68(22):224429, 2003.
- [MRTZ03] M. Mézard, F. Ricci-Tersenghi, and R. Zecchina. Two Solutions to Diluted p-Spin Models and XORSAT Problems. *Journal of Statistical Physics*, 111(3-4):505–533, 2003.
- [MS11] M Mézard and J Sakellariou. Exact mean-field inference in asymmetric kinetic Ising systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(07):L07001, 2011.
- [MS13] Hamed Mahmoudi and David Saad. Generalized mean field approximation for parallel dynamics of the Ising model. 2013.
- [MSR73] P. Martin, E. Siggia, and H. Rose. Statistical Dynamics of Classical Systems. *Physical Review A*, 8(1):423–437, 1973.
- [MSS11] A K Misra, A Sharma, and J B Shukla. Modeling and analysis of effects of awareness programs by media on the spread of infectious diseases. *Mathematical and Computer Modelling*, 53(5-6):1221–1228, 2011.
- [MSV12] Joel C Miller, Anja C Slim, and Erik M Volz. Edge-based compartmental modelling for infectious disease spread. *Journal of the Royal Society, Interface*, 9(70):890–906, 2012.
- [MT73] D. P. Maki and M. Thompson. *Mathematical models and applications: with emphasis on the social, life, and management sciences*. Englewood Cliffs: Prentice-Hall, 1973.
- [Mur89] J D Murray. *Mathematical biology*. Springer-Verlag, 1989.
- [MV80] Silvio Micali and Vijay V. Vazirani. An $O(v^2 \log E)$ algorithm for finding maximum matching in general graphs. In *21st Annual Symposium on Foundations of Computer Science (sfcs 1980)*, pages 17–27. IEEE, October 1980.
- [MVP87] Marc Mézard, Miguel Angel Virasoro, and Giorgio Parisi. *Spin glass theory and beyond*. World Scientific, 1987.
- [MZ96] Rémi Monasson and Riccardo Zecchina. Entropy of the K-Satisfiability Problem. *Physical Review Letters*, 76(21):3881–3885, 1996.
- [MZ97] Rémi Monasson and Riccardo Zecchina. Statistical mechanics of the random K-satisfiability model. *Physical Review E*, 56(2):1357–1370, 1997.
- [MZ02] Marc Mézard and Riccardo Zecchina. Random K-satisfiability problem: From an analytic solution to an efficient algorithm. *Physical Review E*, 66(5):056126, 2002.
- [MZK⁺99] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky. Determining computational complexity from characteristic phase transitions. *Nature*, 400(6740):133–137, 1999.

- [NB09] I Neri and D Bollé. The cavity approach to parallel dynamics of Ising spins on a graph. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(08):P08009, 2009.
- [Neb04] Markus E Nebel. Investigation of the Bernoulli model for RNA secondary structures. *Bulletin of mathematical biology*, 66(5):925–64, 2004.
- [New02] M. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):016128, 2002.
- [New03] M E J Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167–256, 2003.
- [NJ80] Ruth Nussinov and Ann B Jacobson. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proceedings of the National Academy of Sciences*, 77(11):6309–6313, 1980.
- [NTV11] S K Nechaev, M V Tamm, and O V Valba. Statistics of noncoding RNAs: alignment and secondary structure prediction. *Journal of Physics A: Mathematical and Theoretical*, 44(19):195001, 2011.
- [NY96] Hidetoshi Nishimori and Michiko Yamana. Dynamical Probability Distribution Function of the SK Model at High Temperatures. *Journal of the Physical Society of Japan*, 65(1):3–6, 1996.
- [OCK13] Reuben O’Dea, Jonathan J Crofts, and Marcus Kaiser. Spreading dynamics on spatially constrained complex brain networks. *Journal of the Royal Society, Interface*, 10(81):20130016, 2013.
- [OP09] Tore Opsahl and Pietro Panzarasa. Clustering in weighted networks. *Social Networks*, 31(2):155–163, 2009.
- [OS10] H. Ohta and S. Sasa. A universal form of slow dynamics in zero-temperature random-field Ising model. *EPL (Europhysics Letters)*, 90(2):27008, 2010.
- [OZ02] H. Orland and A. Zee. RNA folding and large N matrix theory. *Nuclear Physics B*, 620(3):456–476, 2002.
- [Par80a] Giorgio Parisi. A sequence of approximated solutions to the SK model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980.
- [Par80b] Giorgio Parisi. Magnetic properties of spin glasses in a new mean field theory. *Journal of Physics A: Mathematical and General*, 13(5):1887, 1980.
- [Par80c] Giorgio Parisi. The order parameter for spin glasses: A function on the interval 0-1. *Journal of Physics A: Mathematical and General*, 13(3):1101, 1980.
- [PBGV11] Nicola Perra, Duygu Balcan, Bruno Gonc, and Alessandro Vespignani. Towards a Characterization of Behavior-Disease Models. *PLoS ONE*, 6(8):e23084., 2011.

- [Pea88] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.
- [Per84] P. Peretto. Collective properties of neural networks: A statistical physics approach. *Biological Cybernetics*, 50(1):51–62, 1984.
- [POZ05] M. Pillsbury, H. Orland, and A. Zee. Steepest descent calculation of RNA pseudoknots. *Physical Review E*, 72(1):011911, 2005.
- [PPRT00] A Pagnani, G Parisi, and F. Ricci-Tersenghi. Glassy Transition in a Disordered Model for the RNA Secondary Structure. *Physical Review Letters*, 84(9):2026–2029, February 2000.
- [PPRT01] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi. Reply. *Physical Review Letters*, 86(7):1383–1383, February 2001.
- [PSL⁺06] Katherine S Pollard, Sofie R Salama, Nelle Lambert, Marie-Alexandra Lambert, Sandra Coppens, Jakob S Pedersen, Sol Katzman, Bryan King, Courtney Onodera, Adam Siepel, Andrew D Kern, Colette Dehay, Haller Igel, Manuel Ares, Pierre Vanderhaeghen, and David Haussler. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, 443(7108):167–72, 2006.
- [PSV01a] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6):066117, 2001.
- [PSV01b] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic Spreading in Scale-Free Networks. *Phys. Rev. Lett.*, 86:3200–3203, 2001.
- [PTV12] Pedro C Pinto, Patrick Thiran, and Martin Vetterli. Locating the Source of Diffusion in Large-Scale Networks. *Phys. Rev. Lett.*, 109:68702, 2012.
- [PVF12] B A Prakash, J Vreeken, and C Faloutsos. Spotting Culprits in Epidemics: How many and Which ones? In *ICDM'12; Proceedings of the IEEE International Conference on Data Mining*, 2012.
- [RH11] Yasser Roudi and John Hertz. Mean Field Theory for Nonequilibrium Network Reconstruction. *Physical Review Letters*, 106(4):048702, 2011.
- [Rog10] E. M. Rogers. *Diffusion of innovations*, 2010.
- [RTWZ01] Federico Ricci-Tersenghi, Martin Weigt, and Riccardo Zecchina. Simplest random K-satisfiability problem. *Physical Review E*, 63(2):026702, 2001.
- [RTZ00] Federico Ricci-Tersenghi and Riccardo Zecchina. Glassy dynamics near zero temperature. *Physical Review E*, 62(6):R7567–R7570, 2000.
- [Sai90] Riichiro Saito. A Proof of the Completeness of the Non Crossed Diagrams in Spin 1/2 Heisenberg Model. *Journal of the Physics Society Japan*, 59(2):482–491, 1990.

- [SC03] G Semerjian and L. F Cugliandolo. Dynamics of dilute disordered models: A solvable case. *Europhysics Letters (EPL)*, 61(2):247–253, 2003.
- [SCM04] G. Semerjian, L. F. Cugliandolo, and A. Montanari. On the Stochastic Dynamics of Disordered Spin Models. *Journal of Statistical Physics*, 115(1/2):493–530, 2004.
- [SDS⁺93] James P Sethna, Karin Dahmen, Kartha Sivan, James A Krumhansl, Bruce W Roberts, and Joel D Shore. Hysteresis and hierarchies: Dynamics of disorder-driven first-order phase transformations. 70(21):3347–3350, 1993.
- [SGS93] C. Sfatos, A. Gutin, and E. Shakhnovich. Phase diagram of random copolymers. *Physical Review E*, 48(1):465–475, 1993.
- [Sha48] Claude E Shannon. A Mathematical Theory of Communication. *Bell System Tech. Journal*, 27:623, 1948.
- [SK75] D Sherrington and S Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35(26):1792, 1975.
- [SM14] Munik Shrestha and Cristopher Moore. Message-passing approach for threshold models of behavior in networks. *Physical Review E*, 89(2):022805, 2014.
- [Som81] H. Sompolinsky. Time-Dependent Order Parameters in Spin-Glasses. *Physical Review Letters*, 47(13):935–938, 1981.
- [Som87] Hans-Jürgen Sommers. Path-integral approach to Ising spin-glass dynamics. *Physical Review Letters*, 58(12):1268–1271, 1987.
- [SRMH12] Jason Sakellariou, Yasser Roudi, Marc Mezard, and John Hertz. Effect of coupling asymmetry on mean-field solutions of the direct and inverse SherringtonKirkpatrick model. *Philosophical Magazine*, 92(1-3):272–279, 2012.
- [SS98] David Strang and Sarah A. Soule. Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills. *Annual Review of Sociology*, 24(1):265–290, 1998.
- [SS11] Ira B Schwartz and Leah B Shaw. Rewiring for adaptation. *Physics*, 3(17), 2011.
- [SVB⁺11] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Corinne Régis, Jean-François Pinton, Nagham Khanafer, Wouter Van den Broeck, and Philippe Vanhems. Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC medicine*, 9:87, 2011.
- [SW04] Guilhem Semerjian and Martin Weigt. Approximation schemes for the dynamics of diluted spin models: the Ising ferromagnet on a Bethe lattice. *Journal of Physics A: Mathematical and General*, 37(21):5525–5546, 2004.

- [SWS00] Katarzyna Sznajd-Weron and Josef Szanajd. Opinion Evolution in closed community. *International Journal of Modern Physics C*, 11(06):1157–1165, 2000.
- [SwS13] Katarzyna Sznajd-weron and Janusz Szwabi. Rewiring the network . What helps an innovation to diffuse? 2013.
- [SZ81] H. Sompolinsky and A. Zippelius. Dynamic Theory of the Spin-Glass Phase. *Physical Review Letters*, 47(5):359–362, 1981.
- [SZ82] H. Sompolinsky and A. Zippelius. Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses. *Physical Review B*, 25(11):6860–6875, 1982.
- [SZ10] Devavrat Shah and Tauhid Zaman. Detecting sources of computer viruses in networks. In *Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems - SIGMETRICS '10*, page 203, New York, 2010. ACM Press.
- [SZ11] Devavrat Shah and Tauhid Zaman. Rumors in a Network: Who's the Culprit? *IEEE Transactions on Information Theory*, 57(8):5163–5181, 2011.
- [Sza98a] Grzegorz Szamel. Glauber dynamics of the SK model: theory and simulations in the high-temperature phase. *Journal of Physics A: Mathematical and General*, 31(50):10045–10052, 1998.
- [Sza98b] Grzegorz Szamel. Glauber dynamics of the SK model: theory and simulations in the low-temperature phase. *Journal of Physics A: Mathematical and General*, 31(50):10053–10063, 1998.
- [’t 74] G. ’t Hooft. A planar diagram theory for strong interactions. *Nuclear Physics B*, 72(3):461–473, 1974.
- [Tal06] Michel Talagrand. The Parisi formula. *Annals of Mathematics*, 163(1):221–263, 2006.
- [TAP77] DJ Thouless, PW Anderson, and RG Palmer. Solution of ’solvable model of a spin glass’. *Philosophical Magazine*, 35(3):593, 1977.
- [TB12] Jean M Tchuenche and Chris T Bauch. Dynamics of an Infectious Disease Where Media Coverage Influences Transmission. *ISRN Biomathematics*, 2012(581274):9–11, 2012.
- [TDB⁺11] Jean M Tchuenche, Nothabo Dube, Claver P Bhunu, Robert J Smith, and Chris T Bauch. The impact of media coverage on the transmission dynamics of human influenza. *BMC Public Health*, 11(Suppl 1):S5, 2011.
- [VB85] L Viana and A J Bray. Phase diagrams for dilute spin glasses. *Journal of Physics C: Solid State Physics*, 18(15):3037–3051, 1985.
- [vB00] F. H. D. van Batenburg. PseudoBase: a database with RNA pseudoknots. *Nucleic Acids Research*, 28(1):201–204, January 2000.

- [Ves11] Alessandro Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8(1):32–39, 2011.
- [Vla13] A. A. Vladimirov. Non-crossing matchings. *Problems of Information Transmission*, 49(1):54–57, April 2013.
- [VM07] Erik Volz and Lauren Ancel Meyers. Susceptible-infected-recovered epidemics in dynamic contact networks. *Proceedings. Biological sciences / The Royal Society*, 274(1628):2925–33, 2007.
- [VM09] Erik Volz and Lauren Ancel Meyers. Epidemic thresholds in dynamic contact networks. *Journal of the Royal Society, Interface*, 6(32):233–41, 2009.
- [VNB00] A. M. Vershik, S. Nechaev, and R. Bikbov. Statistical Properties of Locally Free Groups with Applications to Braid Groups and Growth of Random Heaps. *Communications in Mathematical Physics*, 212(2):469–501, 2000.
- [Vol08] Erik Volz. SIR dynamics in random networks with heterogeneous connectivity. *Journal of mathematical biology*, 56(3):293–310, 2008.
- [VOZ05] Graziano Vernizzi, Henri Orland, and A. Zee. Enumeration of RNA Structures by Matrix Models. *Physical Review Letters*, 94(16):168103, 2005.
- [VRB03] Werner Vogels, Robbert Van Renesse, and Ken Birman. The Power of Epidemics : Robust Communication for Large-Scale Distributed Systems. *ACM SIGCOMM Computer Communication Review*, 33(1):131–135, 2003.
- [VSP10] Sven Van Segbroeck, Francisco C Santos, and Jorge M Pacheco. Adaptive contact networks change effective disease infectiousness and dynamics. *PLoS computational biology*, 6(8), 2010.
- [VTN12] O V Valba, M V Tamm, and S K Nechaev. New Alphabet-Dependent Morphological Transition in Random RNA Alignment. *Physical Review Letters*, 109(1):018102, 2012.
- [WS98] D J Watts and S H Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–2, 1998.
- [YFW00] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Generalized belief propagation. In *Advances in Neural Information Processing Systems (NIPS)*, pages 689–695. 2000.
- [YFW03] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 8:236–239, 2003.
- [YFW05] J.S. Yedidia, W.T. Freeman, and Y. Weiss. Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, 2005.
- [Zac77] W W Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33:452–473, 1977.

- [ZS81] Michael Zuker and Patrick Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(1):133–148, 1981.
- [ZY13] K Zhu and L Ying. Information source detection in the SIR model: A sample path based approach. In *2013 Information Theory and Applications Workshop (ITA)*, pages 1–9. IEEE, 2013.