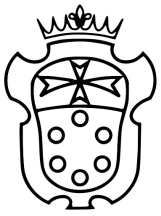
Dynamic network models with applications to finance

by

Piero Mazzarisi



SCUOLA
NORMALE
SUPERIORE

PhD Thesis
in
Financial Mathematics

Supervisors

Prof. Fabrizio Lillo
Prof. Stefano Marmi

A Maria

Acknowledgements

First of all, I would like to thank my supervisor Fabrizio Lillo for his guidance and patience during these years. He has been an inspiring model of enthusiasm and professionalism in academic research. And I would like to thank Fabrizio for his endless support to my growth.

I also would like to thank my internal supervisor in Scuola Normale Superiore Stefano Marmi, his support and enthusiasm has always encourage me to do the best. And I also would like to thank Vito Latora for opening my eyes to the study of networks during my second year at University of Catania and for allowing me to be a visiting student in Queen Mary University under his supervision during my PhD.

I would like to thank Paolo Barucca, Daniele Tantari, and Mateusz Wilinski for their support, suggestions, and infinitely valuable discussions.

Then, I would like to thank the people who made these years full of life, joy, and peace with one word, long talks, or simply one beer.

Finally, I am immeasurably grateful to my family. You are my persistent stability in this dynamic and complex world.

List of publications

Published papers

- Barucca, P., Lillo, F., Mazzarisi, P., and Tantari, D. (2018) Disentangling group and link persistence in dynamic stochastic block models. *Journal of Statistical Mechanics: Theory and Experiment*, 12:123407
- Mazzarisi, P., Lillo, F., and Marmi, S. (2019). When panic makes you blind: A chaotic route to systemic risk. *Journal of Economic Dynamics and Control*, 100:176-199
- Wilinski, M., Mazzarisi, P., Tantari, D., and Lillo, F. (2019). Detectability of macroscopic structures in directed asymmetric stochastic block model. *Physical Review E*, 99:042310
- Mazzarisi, P., Barucca, P., Lillo, F., and Tantari, D. (2019). A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market. *European Journal of Operational Research* (in press).

Other works

- Mazzarisi, P. and Lillo, F. (2017). Methods for reconstructing interbank networks from limited information: A comparison. In *Econophysics and Sociophysics: Recent Progress and Future Directions* (Chapter), pages 201-215. Springer.

Contents

Introduction

Networks are ubiquitous. In recent years there has been a growing interest in the study of complex networks [Newman, 2010, Holme and Saramäki, 2012, Latora et al., 2017]. One of the reasons is that many natural and artificial systems are characterized by the presence of a sparse structure of interactions, *i.e.* only a small fraction of the possible pairs of elements mutually interact (at least at each time). Thus the topology of the network of interactions plays an important role in understanding the aggregate behavior of many complex systems. Moreover most of the investigated systems evolve over time and the structure of the network is generically not constant but new links are formed and old ones are destroyed at each time. Understanding and modeling network dynamics is therefore of paramount importance.

"More is different" [Anderson, 1972]: the whole becomes not only more complex than but also very different from the sum of its parts. In fact, in the passage from the reductionist analysis, focused on the study of the elementary unit of a system, to the constructionist one, focused on the study of an extensive system, entirely new properties appear. This is the case of systemic risk in financial networks: the aggregate of the interacting subunits, which form the financial system, governs the stability of the entire system and the topology of the network of interactions may facilitate the propagation of risk among the subparts. Systemic risk is in effect an emergent phenomenon.

The goal of networks modeling is thus capturing how the microscopic behaviors of single subparts give rise to the linkage structure of the network, then studying how changes at the micro level affect the aggregate behavior of the system. At the same time, flipping perspective, the statistical inference of a network model uses data to deduce the underlying mechanisms which govern the process of network formation and evolution. Learning the linkage mechanisms which have likely produced a real-world network allows also to disentangle a range of microscopic behaviors which may coexist in forming the aggregated structure of the network. Finally, modeling how connections are formed and describing the dynamics governing their evolution, help us, not only understanding, but also predicting the future behavior of a system, possibly at any level of aggregation.

Contribution

This thesis provides new contributions to the field of network models, in two directions. On one hand, we study statistical models of static networks, in particular by contributing to the problem of community detection when link direction is taken into account, thus identifying what are the macroscopic structures of interest for the problem and the conditions for detectability [Wilinski et al., 2019]. Then, we introduce novel statistical models of dynamic networks which are able to capture simultaneously latent dynamics for node-specific characteristics together with link-specific persistence patterns. While the latent dynamics drives the evolution of the network topologies, such as the node degree, *i.e.* the number of incident links to the node, or the community structure, *i.e.* how nodes connect each other in forming groups, link persistence preserves the past structure of the network. Within this context, the contribution of the thesis is twofold, both theoretical and empirical [Mazzarisi et al., 2019a, Barucca et al., 2018]. We develop novel methodologies to disentangle the two linkage mechanisms in order to learn correctly both latent variables and static parameters of the models. And we consider also applications to financial data to reveal genuine patterns of persistence, which reflects the role both nodes and links have in the process of network formation and evolution.

On the other hand, with a focus on the systemic risk of financial systems, we present a theoretical study of the expectation feedback mechanism which governs the dynamics of a financial network, thus determining its dynamical stability [Mazzarisi et al., 2019b]. Any financial system is an expectation feedback system: the current decisions of financial agents depend on what they expect will occur in the future. Agents' decisions affect the price dynamics in illiquid markets. Then, when expectations are formed by using models of past observations, the price dynamics itself feeds back on agents' expectations. This is in effect a feedback dynamics. Interestingly, the process of expectation formation by agents and the price dynamics act on different time scales. In our modeling, it is slow for the agents' expectations and fast for the price dynamics. Moreover, the agents' decisions, given the expectations formed on the basis of the *random* price dynamics, is to some extent *deterministic*, because they represent the optimal portfolio choice in a heavily regulated market. This separation of time scales is crucial and we are able to characterize analytically the feedback dynamics in the asymptotic limit of one time scale infinitely larger than the other one. Hence, we contribute to the research field of systemic risk with the first analytical proof (to the best of our knowledge) of how expectation feedbacks in relation to the estimation of investments' risk and dependencies determine the dynamical instability of a financial system.

In line with the two research directions, the thesis is divided in two parts.

Latent dynamics and persistence patterns in networks

The first part of the thesis is about statistical modeling of both static and dynamic networks. The statistically-oriented literature on the analysis of networks has an outstanding history, starting from the seminal work of [Wolff, 1950] which pointed out the importance of the statistical approach in social networks, the paper of [Moreno, 1934] which invented the sociogram, *i.e.* a diagram of points and lines used to represent relations among people, a precursor to the graph representation for networks, and the empirical studies of Milgram and Travers [Milgram, 1967, Travers and Milgram, 1977], which first recognized the "Small World" phenomenon, *i.e.* short paths of connections linking most people in social networks. Interestingly, most of the early examples of networks were relatively small (in terms of the number of nodes) and involved the study of a single instance of the networked system at a fixed point in time, because of the lack of data. With one exception. In 1968, [Sampson, 1969] collected and studied a "big" dataset of five networks describing the evolution of social relations between novice monks in a New England monastery, with a particular attention on the groups which were formed and their evolution. In effect, this is the first study on the dynamic community structure in temporal networks.

However, the first probabilistic formulation of static networks in terms of random graphs is represented by the Erdös-Rényi model [Erdös and Rényi, 1959] and all the subsequent literature was built upon this milestone work. Erdös and Rényi worked on random graphs with fixed number of nodes and links, and studied the properties of this model with increasing link density, *i.e.* the number of links with respect to all possible connections. An Erdös-Rényi random graph (in one of the two possible variants of the model) is constructed by sampling links independently for any possible couple of nodes with a fixed link probability. Despite the fact of nice mathematical properties and tractability, the Erdös-Rényi model shows some drawbacks for applications, for instance homogenous node degree. For this reason, many generalizations have followed, most of them to capture some characteristics observed in real-world networks, such as power-law distribution of the node degree, effects due to reciprocation of links, community structures, and so on. Here, we focus mainly on two well known models, namely the fitness model (or β-model) [Caldarelli et al., 2002, Chatterjee et al., 2011] and the Stochastic Block Model (SBM) [Holland et al., 1983, Snijders and Nowicki, 1997].

1. The fitness model describes static networks with latent node-specific variables, termed the node *fitnesses*, each one governing the probability for all links incident to the node. Hence, the fitness is a latent feature which controls the connectivity of the node. In that sense, not all nodes are equal to each other and the fitness model represents one possible mechanism to explain heterogeneous degree distributions in networks. There exist many applications, for instance as a possible solution of the problem of network reconstruction in the presence of missing information in economic and financial networks [Garlaschelli and Loffredo, 2004a, Mazzarisi and Lillo, 2017].

2. Stochastic block models have beed introduced to describe the community structure of a network. SBM generalizes the Erdös-Rényi model by giving each pair of nodes a connection probability depending on the communities they belong to, thus describing a large variety of macroscopic structures observed in real-world networks. SBM is specified by the latent group memberships of nodes together with the matrix of connection probabilities, which is termed affinity matrix. From the point of view of statistical inference, one challenging problem is the detection of unobserved communities (generated according to SBM), *i.e.* inferring the node memberships which display a positive overlap with the original assignments, where the overlap is a measure of the intersection between the inferred and the original community memberships, defined such that it is one for exact labeling, zero for random guess. Communities are defined as *detectable* when we are able to find a labeling correlated with the original one, thus resulting in a positive overlap. In fact, SBM has been extensively studied to find the conditions of detectability of communities, starting from the observation of the linkage structure of a network. This problem, known as *detectability threshold* problem, refers to a sharp transition in the parameters space separating two distinct regimes: (i) when it is possible, at least in part, to infer communities of nodes, and (ii) when inference cannot work better than random assignments. The problem of community detection has been solved for static undirected graphs by [Decelle et al., 2011b], which have introduced an optimal inference method, called *Belief Propagation* (BP) algorithm, to show how the detectability threshold depends on the level of assortativity of the network, roughly speaking the difference in link density within the same community and between different communities: below a given threshold for the level of assortativity, the linkage structure is so much homogeneous that any method of community detection is equivalent to a random guess.

Interestingly, both the fitness model and stochastic block models, as well as the Erdös-Rényi model, can be described within the same modeling framework represented by the Exponential Random Graphs (ERG), statistical models of network ensembles[1] specified by an exponential probability distribution for network metrics [Park and Newman, 2004].

Moreover, networks evolve in time, and some links appear and others disappear as time goes on [Holme and Saramäki, 2012]. The same system observed at different times reveals temporal patterns which cannot be captured by studying a single network snapshot, *i.e.* a realization of the network at a particular point in time.

Link formation in dynamic systems is a complex phenomenon. Present links, *e.g.* credit transactions in interbank markets, may depend on present node properties but also on previous network states, *e.g.* the existence of previous transactions between two counterparties. For example, in the interbank market, banks have *preferential* credit relations with some specific counterparties, for several reasons, but, first of all, because they trust them. And preferential credit relations can induce memory

[1] A statistical network ensemble is a probability distribution for the states of the networked system, which are represented by a set of graphs.

effects in link formation, in particular link persistence. Nevertheless, the characteristics of the subparts of a system may change over time, thus completely new connections would appear, never observed in the past. This is the case, *e.g.* , of social networks organized in groups of friends. People meet new people and establish friendships, too. As a consequence, link persistence reveals link-specific mechanisms in the process of network formation as well as node-specific latent variables drive the evolution of the network topology. Interestingly, the different connection mechanisms may coexist in the network dynamics.

When considering dynamic networks in discrete time as time series of graph snapshots, statistical models have been generalized in two directions: (i) modeling explicitly how network metrics evolve by describing some p-th order Markov chain for links, or (ii) introducing a latent process for node-specific characteristics which determine how the topology changes over time.

1. The first class of models tries to capture persistence patterns for temporal network metrics, *i.e.* metrics involving consecutive network snapshots. For instance, link stability is a measure of the frequency of links (or no-links) repeated in time. It is in effect the definition of link persistence. ERG models have been extended to the dynamic case by [Hanneke et al., 2010], to describe such persistence patterns in dynamic network ensembles specified by exponential probability distributions for temporal metrics, the so-called Temporal Exponential Random Graphs (TERG).
   Link persistence in dynamic networks can be also modeled by a discrete autoregressive DAR process [Jacobs and Lewis, 1978b], which describes the mechanism of copying a link from the past. Interestingly, we prove analytically in Chapter 1 the ensemble equivalence for the DAR(p) network model and TERG for link stability with p-order Markov dependence.
2. The second class of dynamic models, on the contrary, describes single-snapshot metrics, such as the node degree at that snapshot, but evolving in time by following some latent dynamics for the node-specific characteristics. For instance, we can capture how the degree of the node changes over time with a dynamic fitness. In our research, among different proposals, we introduce the dynamic fitness model by specifying a latent autoregressive process for the fitness. The dynamic fitness model is in effect a possible specification of a large class represented by Varying-Coefficient Exponential Random Graph (VCERG) models [Lee et al., 2017], which extend ERG to dynamic networks by describing a latent dynamics for the parameters of the exponential random graphs. Finally, within this second class, dynamic stochastic block models capture how the community structure of a network evolves in time. Again, the problem of community detectability can be defined for time-varying communities, in a similar fashion of the static case. In the case of Markovian evolution of communities, the problem has been solved by [Ghasemian et al., 2016], which have quantified the detectability threshold in terms of the level of assortativity (similarly to the static case) and the persistence of communities.

In this stream of research (which is extensively reviewed in Chapter 1), most models have focused either on the statistical description of temporal network metrics, disregarding the role of time-varying topology, or on the dynamical evolution of latent variables, disregarding the role of previous network states in determining the future ones. Instead, we follow a more general rationale dealing with the statistical description of time series of network snapshots, depending on a set of latent variables and on the previous network states.

This thesis provides novel contributions in many directions: (i) we solve the problem of community detection for static networks when link direction is considered; (ii) we propose novel models of dynamic networks which combine the two different approaches in literature, in particular by coupling the Markovian behavior for link persistence with the latent dynamics of node-specific characteristics, for both the fitness model and the stochastic block model. Hence, we propose novel inference methods which are able to disentangle the two linkage behaviors; (iii) we apply our novel methodology to the e-MID interbank market, thus revealing genuine patterns of preferential lending; (iv) we study how link persistence affects the problem of community detection in dynamic stochastic block model by leading to the identification of past communities rather than present ones. Hence, we propose a novel computationally efficient algorithm, that is named Lagged Snapshot Dynamic (LSD) algorithm, which is based on BP but corrected for the bias induced by link persistence.

More specifically,

Chapter 2 we introduce a novel model of dynamic networks where two mechanisms control the probability of a link between two nodes: (i) the existence or absence of this link in the past, and (ii) node-specific latent variables describing the propensity of each node to create links. The first behavior is described by a link-specific Markovian mechanism of copying from the past, similarly to the DAR(1) process, and it is combined with a dynamic generalization of the fitness model, where the latent node fitness evolves in time by following an autoregressive process. We show how the persistence of node fitnesses can give rise to link persistence, also in the absence of the explicitly modeled mechanism of copying from the past. Hence, we propose an Expectation-Maximization (EM) algorithm for model estimation, based on Maximum A Posteriori inference approach, which is able to disentangle between the two linkage mechanisms, thus leading to the unbiased estimation of both latent dynamics and model parameters. We show that EM outperforms the naive Single Snapshot Inference (SSI), a method not accounting for the full information of the time series of network snapshots, see the top left panel of Figure A. Finally, EM is in effect a statistical filtering algorithm which can be used also for link prediction.

We apply our methodology to the e-MID interbank network and we empirically show that the two linkage mechanisms are associated with two different trading behaviors in the process of network formation, namely preferential trading and trading driven by node-specific characteristics. We use the statistical test introduced in [Hatzopoulos et al., 2015] as a measure of preferential trading. Hence,

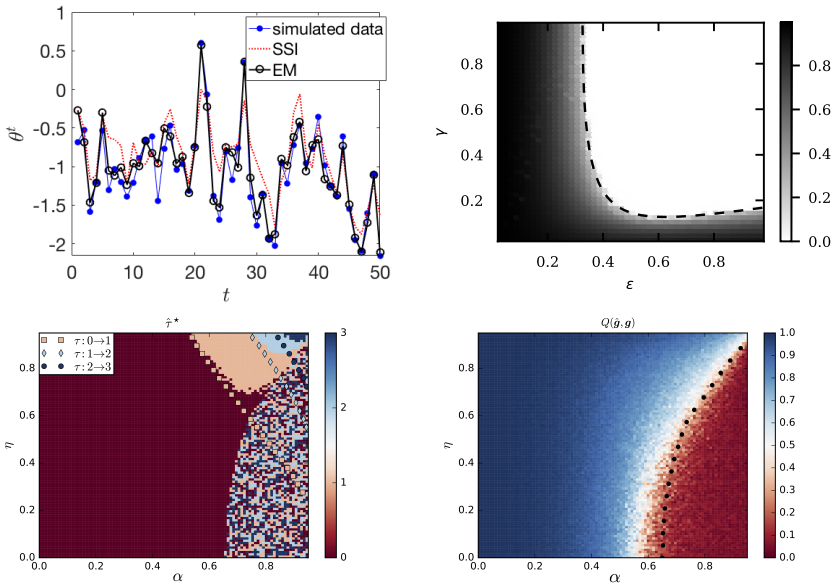


Fig. A Top left: the tracking of the simulated autoregressive fitness dynamics (blue dots) by the EM estimation method (black circles) compared with Single Snapshot Inference SSI (red dots). Top right: overlap between original and inferred community memberships in directed SBM as a function of the assortative parameter ε and the asymmetry parameter γ: increasing the level of assortativity (asymmetry) corresponds to decreasing ε (γ). Bottom left: optimal time lag $\tau^\star$ in time-lagged inference of the dynamic SBM in the presence of link stability as a function of group and link persistence parameters η and α, respectively. Bottom right: overlap in dynamic SBM as a function of η and α.

on one hand, the results show a significant correlation between the persistence features of links and the couples of banks statistically validated as preferential lending relations. On the other hand, fitness dynamics displays a positive correlation with the total bank exposure, thus revealing that the degree of the nodes is driven by the dynamics of the balance sheet aggregates (within the market).
Finally, we perform an out-of-sample link prediction exercise to show that the network topology captured by the latent dynamics is more important than preferential linkages in determining the average characteristics of the e-MID network, with the exception of a small set of links reflecting the preferential relations.

Chapter 3 We study the problem of community detectability for stochastic block models when link direction is considered. With respect to the undirected case, the addition of directions make the detection even easier, since nodes can be classified according to their in(out)-degree even when the groups have on average the same number of links. As a consequence, the problem becomes non trivial only if both in- and out-degrees are equal (on average). This condition reveals thus novel macroscopic structures of interest, characterized by an affinity matrix that is a multiple of a doubly stochastic matrix, *i.e.* with constant sums along rows and

columns. To solve the problem of detectability threshold, we consider a special directed SBM, termed asymmetric planted partition model, that is the generalization of the assortative planted partition model (*i.e.* the benchmark model used in the community detection literature) when links have directions and whose affinity matrix is doubly stochastic and circulant. By using the BP algorithm generalized for directed networks and by exploiting the properties of circulant matrices, we find closed form solutions and show how the detectability threshold (line) depends on the level of assortativity as well as the asymmetry between the number of links from one community to another one, and the number of links in the opposite direction. In fact, a positive overlap between the inferred and original community memberships can be achieved not only by increasing assortativity but also by increasing asymmetry, see the top right panel of Figure A. Both of them, above a given threshold, lead to positive detectability regardless of the value of the other. We further describe the phase transition in relation to the number of groups. For small number of groups we observe a second order phase transitions in both assortativity and asymmetry. When the number of groups is above five, then the phase transitions becomes of first order.

Chapter 4 We study the problem of community detection for dynamic networks whose evolution is governed by both a latent dynamics for the group memberships and a Markovian behavior for the link persistence. In particular, we study the inference of the assortative planted partition model in which both node memberships and links evolve by following a DAR(1) process. Thus, the dynamics is governed by two parameters, namely group and link persistence parameters. By considering BP inference from single snapshot observations of the network, we show that link persistence makes the inference of communities harder, decreasing the detectability threshold, while community persistence tends to make it easier. Moreover, link persistence tends to preserve the past linkage structure of the network, thus entangling links generated according to the present communities with past links which take memory of the old ones. We analytically show that communities inferred from single network snapshot can share a maximum overlap with the underlying communities of a specific previous instant in time. This leads to time-lagged inference: the identification of past communities rather than present ones. Hence, we quantify this effect in terms of the optimal time lag to be considered for correct labeling, see the bottom left panel of Figure A and propose a corrected algorithm, the Lagged Snapshot Dynamic (LSD) algorithm, for community detection in dynamic networks with link persistence. We analytically and numerically characterize the detectability threshold of such algorithm as a function of the group and link persistence parameters of the model, see the bottom right panel of Figure A. Finally, we notice that a suitable modification of the dynamic BP algorithm of [Ghasemian et al., 2016] outperforms LSD by using the information of the whole time series of network snapshots, but at the expense of higher computational costs.

Feedback dynamics in financial networks

The second part of the thesis is a theoretical study about expectation feedbacks governing the dynamics of portfolio decisions of large financial institutions in markets where the buying and selling move the prices because of illiquidity of investments. As the financial crisis of 2007-2008 made clear, illiquidity combined with a finite number of available assets in creating a diversified portfolio, can have destabilizing effects when one or several large institutions begin a deleveraging cycle. We study this phenomenon by investigating the evolution of a financial bipartite system of institutions investing in some risky and illiquid assets in the presence of risk contagion mediated by overlapping portfolios and fire sale spillovers.

The milestone work of [Allen and Gale, 2000] made evident that any financial system exhibit a high degree of interdependence because of both direct and common exposures between counterparties, the presence of many financial instruments creating high overlap among portfolios, and many interconnections between different markets, which contribute to a unique global market as recent financial crises have evidenced. The description of financial systems in terms of networks may help in understanding the process of formation of systemic risk and how the contagion mechanism works.

In particular, financial networks can significantly amplify distress propagation in the presence of fire sale spillovers and overlapping portfolios. Indeed, the contagion of risk may be propagated through overlapping portfolios, in which the falling of the price of an 'external' asset will affect simultaneously all institutions having that investment in their portfolio [Caccioli et al., 2014, Corsi et al., 2016, Gualdi et al., 2016]. The shock can be further amplified by, *e.g.* , distressed selling of one institution [Cont and Wagalath, 2013] or the strategy of leverage targeting adopted by large investors [Adrian and Shin, 2010]. The mechanisms of contagion of risk together with the drivers governing the dynamics of a financial systems, in particular the role of expectation feedbacks [Hommes, 2013], are reviewed in Chapter 5.

In the presence of distress propagation, financial institutions are active players which take decisions and react when exposed to risk. In particular, portfolio decisions depend on the expectations institutions have about the future risk. We study how risk expectations affect the systemic risk of a financial system where the contagion is mediated by overlapping portfolios.

More specifically,

Chapter 6 we combine all stylized facts about systemic risk and we propose an analytical model to study the impact of expectation feedbacks on systemic dynamical stability of a financial system.

In fact, regulators try to preserve robustness and resilience of financial markets by imposing some constraints to financial institutions, such as the Value-at-Risk constraint. All the regulators' constraints require an estimation of the riskiness of the investments as well as of the dependencies between extreme events of financial returns. Therefore both the capital requirement constraint and

the risk/dependency expectations play a crucial role in determining the systemic stability of financial markets.

On one hand, as documented by many papers (see for example [Adrian and Shin, 2013]), VaR capital requirements, as other risk constraints, can induce a perverse demand function: in order to target the leverage, a financial institution will sell more assets if their price drops and vice versa when their price rises. Thus, a marked-to-market and VaR constrained financial institution will have a positive feedback effect on the prices of the assets in its portfolio.

On the other hand, the implementation of any capital requirement depends on the expectations financial institutions have on the risk of the assets in the portfolio and on their statistical dependence. Since trading decisions, which drive endogenously the market in the presence of illiquid assets, are based on institutions' expectations and expectations of risk are usually formed by looking at the past history of prices, this creates a second feedback effect in addition to the one due to target leveraging.

We present an analytical model of the financial system where both feedbacks mechanisms are present. Building on [Corsi et al., 2016], we model a set of financial institutions having Value-at-Risk capital requirements and investing in a portfolio of risky and illiquid assets, whose prices evolve stochastically in time and are endogenously driven by the trading decisions of financial institutions. The estimations of risk of the investment assets, and as a consequence the leverage, are periodically updated and banks use a backward-looking expectation scheme which considers price returns in a past time window to build estimates. The two feedback mechanisms are coupled by the price dynamics, which on one side is used to mark-to-market the portfolio and to estimate risk and correlations, and on the other one is endogenously affected by the trading activity of financial institutions.

Interestingly, the two feedback mechanisms act on different time scales. In our model the time scale of leverage targeting is shorter than the time scale over which financial institutions update their risk expectations. This separation of time scales is crucial in our modeling. Since the slow variables, associated with updates of risk expectations, evolve in time as a function of averages over the fast variables, associated with leverage targeting, our model can be casted as a discrete time slow-fast dynamical system. The ratio between the two time scales is the key parameter determining the type of mathematical modeling. We show that when this ratio tends to infinity, *i.e.* financial institutions are continuously marked-to-market, the dynamics is described by a deterministic map. The window used to form expectations of risk plays a central role in determining systemic stability and leverage cycles appear when investors become more myopic relative to past history of asset prices, *i.e.* the memory becomes smaller than a given threshold. Our model predicts that the deterministic dynamics of the financial system becomes chaotic when the memory decreases further and goes below a second smaller threshold (see the bifurcation diagram associated with the dynamics of the financial leverage in Figure B). When the ratio between the two time scales is finite a random slow-fast dynamical system describes the system. Even

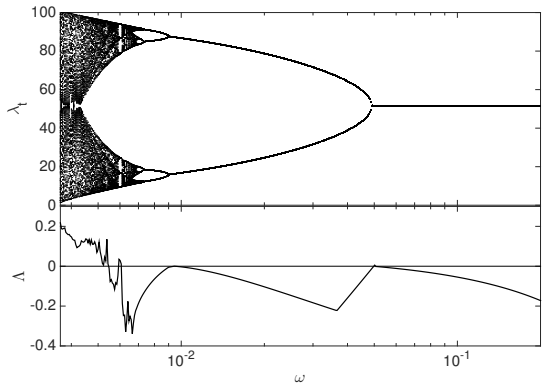


Fig. B Bifurcation diagram associated with the dynamics of the financial leverage λ_t as a function of the memory parameter ω of risk expectations (top panel) (the larger is ω, the larger is the memory) and the Lyapunov exponent associated with the dynamics (bottom panel).

if mathematically this is harder to study, because of the joint chaotic and stochastic dynamics, we show by analytical arguments and numerical simulations that the main dynamical characteristics remain unchanged.

We are therefore able to characterize the possible dynamical outcomes for the considered financial system as a function of the memory window used to form expectations, the tail parameter of the Value-at-Risk, the number of asset classes available for investments, the ratio between the two time scales (related to the presence of market frictions), and a parameter determining the level of financial innovation. We show how the breaking of the fixed point equilibrium for the financial system occurs via a period-doubling bifurcation when any of these parameters is varied and how the dynamics of the financial system may be intrinsically chaotic in certain parameter regions. Each of these parameters can at least in part be controlled by regulators, thus our model is able to provide some policy recommendations, which can be summarized (briefly) as follows:

a. increasing the memory of risk expectations has always a stabilizing effect because of increased information about the dynamics of the financial system;
b. less stringent capital constraint makes the financial system more unstable because of stronger feedback effects;
c. introducing financial innovations, which, for example, make easier to diversify the portfolio, and/or introducing new financial instruments, can break the fixed-point equilibrium of the system because of the combined effects of overlapping portfolios and diminished risk perception;
d. other things being equal, removing market frictions has a stabilizing effect within our model. In fact, the strategy of being marked-to-market in the capital structure represents the control strategy of the balance sheet which has the consequence of reducing the amplitude of the cycles of leverage.

Part I
Latent dynamics and persistence patterns in networks

Persistence is the art of being attached to the past. Change is hope in the future. Never forget, but don't stop.

Chapter 1
Statistical models of static and dynamic networks

Abstract In this chapter we review some statistical models of networks, with a particular attention to those models which have been generalized to dynamic networks. For the static case, we focus in particular on the fitness model, which is based on the idea of fitness, a latent node-specific characteristic determining how many links the node creates with neighbors, and to the stochastic block model, which describes the community structure of a network. We study further the problem of community detection from the observation of a network snapshot, with a focus on the conditions of detectability from a theoretical point of view. Then, we review dynamic network models, which can be grouped in two categories: static and time-varying parameter models. The first class aims to describe some Markovian behaviors of links, for instance link persistence captured by the mechanism of copying a link from the past. On the contrary for the second class of models, the network evolution is captured by a latent dynamics describing how some network topologies change over time, for instance nodes changing communities as time goes on, thus resulting in changing connections with other nodes too. Finally, we study the problem of community detection for dynamic networks.
This represents an introductory chapter to the research presented in Chapters 2 and 4 where we merge the two approaches based on static and time-varying parameters to disentangle the two linkage behaviors in dynamic networks. Finally, in Chapter 3 we study the problem of community detection where link direction is considered, thus showing an entire new class of macroscopic structures which are of interest for the problem.

Introduction

In recent years, there has been a growing interest in the study of complex networks [Newman, 2010, Holme and Saramäki, 2013]. One of the reasons is that many natural and artificial systems are characterized by the presence of a sparse structure of interactions, i.e. only a small fraction of the possible pairs of elements mutually in-

teract (at least at each time). Thus the topology of the network of interactions plays an important role in understanding the aggregate behavior of many complex systems. Moreover most of the investigated systems evolve over time and the structure of the network is generically not constant but new links are formed and old ones are destroyed at each time.

Statistical modeling of network data has been posited as a major topic of interest in different areas of study and, more recently, a great effort has been devoted to the statistical modeling of dynamic networks - the study of networks which evolve over time. Models of networks, both static and dynamic, are countless. Then, a comprehensive survey of all statistical models is out of the scope of this thesis. The goal is far more modest. Here, we aim to discuss two lines of research about statistical modeling of dynamic networks, namely static and time-varying parameter models. Then, our original research shows how the two approaches can be merged in order to disentangle different linkage behaviors in network dynamics.

Regarding static networks, statistical models aim to describe network ensembles which reproduce in average some specific real-world characteristics, ranging from density, degree distribution, motifs, community structure and so on. We focus in particular on those models which are of interest for this thesis.

Then, we present some generalizations to the dynamic case which can be grouped into two different classes: static parameter models and time-varying parameter models. We highlight how the two different approaches differ from each other, specifically what kind of temporal patterns they are able to capture. In particular, the first class of models describes a Markovian behavior for some observed network metrics, *e.g.* the mechanism of copying a link from the past. The second class of models describes, instead, some unobserved dynamics for node-specific characteristics, *e.g.* membership of a community, which determines how the network evolves in time.

The remainder of this chapter is organized as follows. In Section 1.1 we review some basic concepts in network theory and we fix the notation. Then, in Section 1.2 we present a review of statistical models of static networks, while in Section 1.3 we consider the case of dynamic networks.

1.1 Introduction to networks

A network is, in its simplest form, a set of points, *i.e. nodes*, connected by lines, *i.e. links*, see Figure 1.1. Many systems in many fields can be described as a network and thinking of them in this way has the advantage of capturing the complex emergent behavior arising from the interactions among the subparts. In Finance, *e.g.* , there exist several models describing the risk of a financial institution, that is a crucial aspect of asset pricing and portfolio management, but having largely the focus on the business of that specific institution. However, in some periods of financial distress, it is well known that systemic aspects involving more than one institution become significant in the propagation of risk, *e.g.* contagion processes or feedback systems which amplify the distress [Gai and Kapadia, 2010, Danielsson et al., 2012]. This is

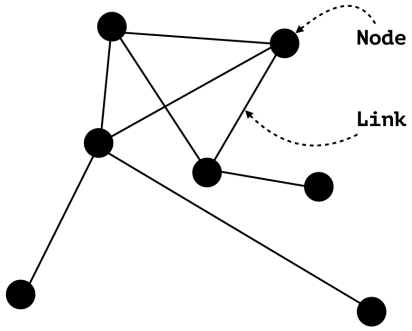


Fig. 1.1 A graphical example of a network composed by seven nodes and eight links.

the intuition behind the so called *systemic risk*. The description of a financial system as a network helps in describing the aspects related to systemic risk (to name just one of the many research fields).

Many real-world networks display substantial non-trivial topological features, with patterns of connections which are neither purely regular nor purely random. Such features include a heavy tail in the degree distribution (scale-free networks [Barabási et al., 2000]), a high clustering coefficient [Newman, 2001], the small-world behavior [Watts and Strogatz, 1998], assortativity among nodes [Newman, 2002], community structure [Girvan and Newman, 2002], and so on. In the presence of these nontrivial characteristics, we refer to a network as *complex network*. For instance, there are evidences in favor of scale-free degree distribution for the interbank market [Boss et al., 2004], as well as a nontrivial organization of banks in terms of both core-periphery and bipartite structures [Fricke and Lux, 2015, Barucca and Lillo, 2016]. In the following, we describe some statistical network models which are able to capture some of these characteristics for both the static and the dynamic cases.

In mathematical terms, a network, or a *graph*, is an ordered pair $G \equiv (V,E)$ of a set V of nodes, or vertices, and a set E of links, or edges, with $E \subseteq \{(i,j)|(i,j) \in V^2 \text{ and } i \neq j\}$, where we exclude the possibility of self-loops. When we consider link direction, any pair of nodes can be connected by two links, one for each direction, *i.e.* $i \to j$ and $j \to i$.

We can also open to the possibility of nodes connected by more than one edge. In that case, we refer to the link as a multi-edge and to the network as *weighted*, *i.e.* each link is associated with an integer number representing the weight. If we exclude the possibility of multi-edges, the graph is called *simple*.

A graph can be organized in two disjoint sets of nodes, such that an edge connects one node of a set to another node of the other set. In this case, the network is referred as *bipartite*. In the following, unless specified, we consider the case of simple undirected graphs.

When we study statistical models of networks, it is necessary to consider not only one single instance but a large number of networks and weighting them with some probability distribution. Hence, a network ensemble is an idealization consisting of a large number of virtual copies (sometimes infinitely many) of an observed network, considered all at once, each one representing a possible state that the real system might be in. Thus, the network ensemble is defined as the couple formed by the set of networks and the probability distribution associated with this set.

There are a number of different ways to represent a graph for an operative point of view. Here, we adopt the description in terms of the *adjacency matrix*. Given $N = |V|$ the number of nodes, the adjacency matrix is a $N \times N$ matrix $\boldsymbol{A}$ having entry A_{ij} equal to one if node i is linked to node j, zero otherwise. In the case of simple undirected graphs without self-loops, $\boldsymbol{A}$ is a symmetric matrix with diagonal elements equal to zero. For instance, in terms of the adjacency matrix, the degree sequence of a graph is given by $\{c_i\}_{i=1,\dots,N}$ with $c_i = \sum_j A_{ij}$.

When observed on a suitable timescale, any system is inherently dynamic, and a network with links evolving in time is called *temporal* or *dynamic network*. There exist different representations of a temporal network, ranging from the description in terms of contact sequence, to the one with the adjacency index, see [Holme and Saramäki, 2012, Casteigts et al., 2012] for further details. However, the most common way to represent a dynamic network in discrete time is describing it as a sequence of graphs, thus recovering the description in terms of adjacency matrices. That is, a dynamic network is seen as a time series of adjacency matrices $\boldsymbol{\mathcal{A}} \equiv \{\boldsymbol{A}^1, \dots, \boldsymbol{A}^T\}$. In this case, the unit value from t to $t+1$ represents the timescale at which we observe the network evolution. For instance, an overnight interbank market can be described by a time series of network snapshots, where each snapshot represents a day of transactions. In other systems, *e.g.* , with overlapping links, each one having different duration, the definition of the timescale could be more challenging and spurious effects may be arising. In this case, other descriptions must be adopted.

Below, we review static and dynamic network models which describe (from a statistical point of view) some topological features of networks and how they evolve in time.

1.2 Statistical models of static networks

The typical scenario addressed in the creation of a statistical network model is the following: we observe some number of network properties of a real-world network, such as number of links, vertex degrees, clustering coefficients, correlation functions, and so forth, and we aim to construct an ensemble of networks reproducing on average such metrics.

In mathematical terms, given a graph $G \in \mathscr{G}$ where $\mathscr{G}$ represents the ensemble of graphs, we associate to each graph G a probability $\mathbb{P}(G|\boldsymbol{\pi})$ depending on model parameters $\boldsymbol{\pi} \equiv \{\pi_i\}_{i=1,\dots,n}$ and, for a given set of metrics $\boldsymbol{x} \equiv \{x_i\}_{i=1,\dots,M}$, we have

$$\sum_G \mathbb{P}(G|\boldsymbol{\pi})\boldsymbol{x}(G) = \mathbb{E}(\boldsymbol{x}|\boldsymbol{\pi}) \tag{1.1}$$

where $\mathbb{E}(\boldsymbol{x}|\boldsymbol{\pi})$ is the expected value of the metrics over the ensemble of graphs specified by parameters $\boldsymbol{\pi}$.

Then, assume to observe a real-world network G^* with metrics $\boldsymbol{x}^*$, thus asking what are the parameters $\boldsymbol{\pi}$ which describe an ensemble of graphs whereby G^* is a specific realization of the ensemble, and, in particular, the values $\hat{\boldsymbol{\pi}}$ such that $\mathbb{E}(\boldsymbol{x}|\hat{\boldsymbol{\pi}}) = \boldsymbol{x}^*$.

This problem is known as *statistical inference*. The conclusion of the statical inference of a model is a *point estimate*, *i.e.* the values that best approximate some parameters of interest given the data, or an *interval estimate*, *i.e.* the confidence interval constructed around the point estimate [Casella and Berger, 2002]. Model inference is based on some assumptions, such as the correct specification of the data-generating process together with the consistency of estimates, as well as the condition of model identifiability. In fact, a model is identifiable if it is theoretically possible to learn the true values of parameters after obtaining an infinite number of observations from it. Mathematically, it is equivalent to say that different values of the parameters give rise to different probability distributions of the observable variables. When a model is identifiable, the inference problem is well specified and there exist different methods to estimate the parameters of the model, ranging from maximum likelihood types of estimators which involve the maximization of the likelihood $\mathbb{P}(G^*|\boldsymbol{\pi})$, Bayesian inference which is based on the update of the posterior distribution of parameters $\mathbb{P}(\boldsymbol{\pi}|G^*)$, up to the method of moments. For further details about estimation methods, see Section 1.2.1.

In the Subsections from 1.2.2 to 1.2.6 we review some statistical models of static networks which are of interest for this thesis. Finally, applications to network data encompass as many as possible research areas, to face different problems, such as clustering of data [Karrer and Newman, 2011, Peixoto, 2014b], network reconstruction in the presence of missing information [Mistrulli, 2011, Caldarelli et al., 2013, Mazzarisi and Lillo, 2017], link prediction [Liben-Nowell and Kleinberg, 2007, Richard et al., 2014], statistical validation of network interactions [Tumminello et al., 2011], to name but a few.

1.2.1 Inference of network models: methods

There exist different approaches to face the problem of obtaining both point and interval estimates of network model parameters, once the inference problem of the model under investigation is well posed, *i.e.* the conditions of both identifiability and consistency of estimates hold.

Without any intention of being exhaustive in the description of the whole statistical inference problem, we sketch the three main methods adopted for network model inference.

First, maximum likelihood estimation is a method of obtaining point estimate of the model parameters $\boldsymbol{\pi}$ given the observations G^*, by maximizing the likelihood function $L(\boldsymbol{\pi}) \equiv \mathbb{P}(G^*|\boldsymbol{\pi})$ or, equivalently, the log-likelihood function $\log L(\boldsymbol{\pi}) \equiv \log \mathbb{P}(G^*|\boldsymbol{\pi})$, *i.e.* by solving

$$\arg\max_{\boldsymbol{\pi} \in Dom(\boldsymbol{\pi})} \log L(\boldsymbol{\pi}), \tag{1.2}$$

if the maximum exists. If yes, $\hat{\boldsymbol{\pi}}$ solving (1.2) is the Maximum Likelihood Estimator (MLE) of the model defined by the probability distribution $\mathbb{P}(G^*|\boldsymbol{\pi})$. In the case of no analytical solution, we have different numerical methods, ranging from the Newton-Raphson method [Friedman et al., 2001] to more sophisticated ones, to solve (1.2). Finally, we aim to stress the fact that, if the likelihood function is differentiable (in $\boldsymbol{\pi}$), a necessary condition for the existence of MLE is

$$\frac{\partial}{\partial \pi_i} \log L(\boldsymbol{\pi}) = 0, \ \forall i = 1, ..., n. \tag{1.3}$$

Hence, solutions of (1.3) represent possible candidates for MLE. However, if $f(\boldsymbol{\pi}) \equiv -\log L(\boldsymbol{\pi})$ is a convex function in the n-dimensional domain of parameters $\boldsymbol{\pi}$ and $\hat{\boldsymbol{\pi}}$ and the solution of (1.3) is interior to the domain, then this solution is the unique maximum $\hat{\boldsymbol{\pi}}$. See [Casella and Berger, 2002] for further details.

Second, Bayesian inference is a method to obtain the interval estimate of the model parameters by using the Bayes' theorem as the update rule when more information becomes available (in fact, Bayesian updating is particularly important in the case of time sequence of data). From a conceptual point of view, Bayesian inference differs from MLE because parameters are considered as quantities whose variations are described by a probability distribution, *i.e.* the *prior distribution* $\mathbb{P}(\boldsymbol{\pi})$, and not fixed. The observations are then taken from the conditional probability of G^* given the parameters $\boldsymbol{\pi}$, *i.e.* the likelihood $\mathbb{P}(G^*|\boldsymbol{\pi})$. Finally, the prior updated with observations is called the *posterior distribution*, *i.e.* $\mathbb{P}(\boldsymbol{\pi}|G^*)$. Hence, the posterior is obtained according to the Bayes' rule

$$\mathbb{P}(\boldsymbol{\pi}|G^*) = \frac{\mathbb{P}(G^*|\boldsymbol{\pi})\mathbb{P}(\boldsymbol{\pi})}{\mathbb{P}(G^*)} \tag{1.4}$$

where $\mathbb{P}(G^*)$ is sometimes termed the marginal likelihood or "model evidence". This factor is the same for all possible $\boldsymbol{\pi}$, then it is not important in determining the interval estimate of parameters. Usually, Monte Carlo Markov Chain (MCMC) methods are adopted to determine the posterior. In particular, by constructing a Markov chain that has the desired distribution as its equilibrium distribution, one can obtain a sample of the desired distribution by observing the chain after a number of steps. Once the posterior distribution is obtained, the interval estimate of $\boldsymbol{\pi}$ is obtained accordingly, for further details see [Casella and Berger, 2002].

In Bayesian statistics, there exists also a hybrid method between Bayesian inference and MLE, namely Maximum A Posteriori (MAP) estimation, that is a point estimate but based on the maximization of the posterior distribution of parameters

[Gauvain and Lee, 1994, Bassett and Deride, 2019]. It is particularly useful in the case of models with latent variables, *e.g.* networks generated with unobserved community structure where each node is characterized by its (latent) membership. Let $\boldsymbol{\theta}$ be the set of latent variables and $\boldsymbol{\pi}$ being the other model parameters, the joint posterior can be expressed as

$$\mathbb{P}(\boldsymbol{\theta}, \boldsymbol{\pi}|G^*) = \mathbb{P}(\boldsymbol{\theta}|G^*, \boldsymbol{\pi})\mathbb{P}(\boldsymbol{\pi}|G^*). \tag{1.5}$$

The MAP approach consists in alternating the maximization of the two terms on the right hand side of (1.5). In particular, according to the Bayes' rule, the first term in the right hand side of (1.5) is

$$\mathbb{P}(\boldsymbol{\theta}|G^*, \boldsymbol{\pi}) \propto \mathbb{P}(G^*|\boldsymbol{\theta}, \boldsymbol{\pi})\mathbb{P}(\boldsymbol{\theta}). \tag{1.6}$$

The maximization over $\boldsymbol{\theta}$ of (1.6) leads to a point estimate similar to MLE but corrected with the prior information on latent variables, if any. Then, the second term of (1.5) is more challenging because it usually involves the integral over the posterior distribution of latent variables, *i.e.*

$$\mathbb{P}(\boldsymbol{\pi}|G^*) = \sum_{\boldsymbol{\theta}} \mathbb{P}(G^*, \boldsymbol{\theta}|\boldsymbol{\pi})\mathbb{P}(\boldsymbol{\pi}) = \sum_{\boldsymbol{\theta}} \mathbb{P}(G^*|\boldsymbol{\theta}, \boldsymbol{\pi})\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{\pi})\mathbb{P}(\boldsymbol{\pi}), \tag{1.7}$$

and, sometimes, some approximation could be required. However, as first recognized by [Neyman et al., 1948, Little and Rubin, 1983], MAP approach should be preferred in any case to a point estimate of both $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ obtained by maximizing jointly the posterior distribution on the left hand side of (1.5). In fact, the overall point estimation of both parameters and latent variables is not necessarily consistent as sample size increases.

Finally, the method of moments is, perhaps, the oldest method of finding point estimators of the model parameters (without considering the presence of latent variables). It has the advantage of being quite simple to use. However, in many cases it is not the optimal method, that is it yields estimators that may be improved upon. The method of moments is found by equating n sample statistics to the corresponding n observed (population) moments (given the likelihood of data). In the case of statistical network models, it means to find the solution of the system of equations

$$\boldsymbol{x}^* = \mathbb{E}(\boldsymbol{x}|\boldsymbol{\pi}), \tag{1.8}$$

where $\boldsymbol{x}^*$ represents a set of n complete and independent[1] measurements (moments) of G^* which we preserve by averaging over the ensemble ($\mathbb{E}(\boldsymbol{x}|\boldsymbol{\pi})$).

In the following, when required, we describe more specifically the adopted estimation method for the inference of specific models.

[1] In the sense that we obtain a system of equations (1.8) admitting a unique solution.

1.2.2 Exponential random graphs

When we aim to describe a network ensemble, the question number one is how choosing the probability distribution over the ensemble. Even if there may exist different approaches to answer the question, a possible choice is to describe the probability distribution by requiring the expected characteristics of the graph ensemble to match a given set of measurements of the real-world network, while maximizing the entropy of the ensemble [Park and Newman, 2004]. A probability distribution over a network ensemble of this type has the same role of the Boltzmann distribution in classical statistical mechanics: for each graph in the ensemble (*microstate*), the characteristics are the ones we aim to preserve, either exactly (*microcanonical* ensemble) or on average (*canonical* ensemble), and the probability distribution over the ensemble is the one which maximizes the Shannon entropy [Shannon, 1948], meaning that no other characteristics besides the ones we have chosen are considered.

Let $G \in \mathscr{G}$ be a graph in the ensemble with N nodes and let $\mathbb{P}(G)$ be the probability associated with the same graph within the ensemble. We choose $\mathbb{P}(G)$ such that the expected values of n observable characteristics, $\mathbb{E}(x_i(G))\ \forall i = 1,...,n$ are equal to the measurements $x_i^*\ \forall i = 1,...,n$ observed in the real-world network. [2] The probability distribution is obtained by maximizing the Shannon entropy with the normalization constraint,

$$S = -\sum_{G\in\mathscr{G}} \mathbb{P}(G)\ln\mathbb{P}(G)\ \ s.t.\ \sum_{G\in\mathscr{G}} \mathbb{P}(G)x_i(G) = x_i^*,\ \sum_{G\in\mathscr{G}} \mathbb{P}(G) = 1. \tag{1.9}$$

As shown by Shannon [Shannon, 1948], the quantity S defined in (1.9) is the one that best represents the lack of information beyond the known observables. By solving this problem, we obtain a network probability distribution.

By introducing the Lagrange multipliers $\alpha,\boldsymbol{\theta} \equiv \{\theta_i\}_{i=1,...,M}$ and considering a generic probability mass function $p(G)$, the maximum entropy probability distribution is the one solving

$$\frac{\partial}{\partial p(G)}\left\{S + \alpha\left(1 - \sum_{G\in\mathscr{G}} p(G)\right) + \sum_{i=1}^{n}\theta_i\left(x_i^* - \sum_{G\in\mathscr{G}} p(G)x_i(G)\right)\right\} = 0. \tag{1.10}$$

The formal solution is

$$\mathbb{P}(G,\boldsymbol{\theta}) = \frac{e^{-H(G,\boldsymbol{\theta})}}{Z(\boldsymbol{\theta})},\ \text{with } H(G,\boldsymbol{\theta}) = \sum_{i=1}^{n}\theta_i x_i(G),\ Z(\boldsymbol{\theta}) = e^{\alpha+1} = \sum_{G\in\mathscr{G}} e^{-H(G,\boldsymbol{\theta})}, \tag{1.11}$$

[2] $x_i(G)$ is the value of the observable x_i for a given i associated with graph G of the ensemble. In the *micro-canonical* ensemble, we choose $x_i(G)$ equal to the observed quantity x_i^* for each graph G in the ensemble. In the *canonical* ensemble, this equality holds only on average.

where $e^{-H(G,\boldsymbol{\theta})}$ corresponds to the Boltzmann's factor with $H(G)$ being the Hamiltonian of the model and $Z(\boldsymbol{\theta})$ the partition function.

The solution (1.11) corresponds to consider the Exponential Random Graphs (ERG) ensemble, *i.e.* the graph ensemble having the family of exponential distributions as probability distributions over the graphs.

Given a specific exponential random graph model, the expected value of any graph property y within the model is the average over the probability distribution (1.11), *i.e.*

$$\mathbb{E}(y|\boldsymbol{\theta}) = \sum_{G\in\mathscr{G}} \mathbb{P}(G|\boldsymbol{\theta})y(G). \tag{1.12}$$

The exponential random graph, like all such maximum entropy ensembles, gives the expected value of an unknown quantity y, given a set of known measurements.

Several models of networks can be described rigorously within the ERG ensembles. In the following, we show some examples of interest for this thesis.

1.2.3 Erdös-Rényi model

The study of random graphs was introduced for the first time by [Erdös and Rényi, 1959, Gilbert, 1959] which describe the so-called Erdös-Rényi model. In this model, all graphs on a fixed vertex set with a fixed number of edges are equally likely and each link has a fixed probability of being present or absent, independently from other links. According to this definition, each graph $G \in \mathscr{G}$ with N nodes and M links has probability

$$\mathbb{P}(G) = p^M(1-p)^{\binom{N}{2}-M} \tag{1.13}$$

with p the link probability equal for any possible couple of nodes.

The model described by (1.13) is the simplest case of exponential random graphs with the total number of links as the characteristic we aim to preserve over the ensemble. In the case of undirected graphs, the Hamiltonian is

$$H(G,\theta) = -\theta M, \tag{1.14}$$

and the partition function can be computed as

$$Z(\theta) = \sum_{G\in\mathscr{G}} e^{-H(G,\theta)} = \sum_{\{A_{ij}\}} e^{\theta\sum_{i,j>i}A_{ij}} = \prod_{i,j>i}\sum_{A_{ij}=0}^{1} e^{\theta A_{ij}} = \prod_{i,j>i}(1+e^{\theta}), \tag{1.15}$$

according to the independence of links A_{ij} $\forall i,j>i$. Since we have $\binom{N}{2}$ possible couples of nodes, the probability of a graph is

$$\mathbb{P}(G,\theta) = \frac{e^{-H(G,\theta)}}{Z(\theta)} = \frac{e^{\theta M}}{(1+e^{\theta})^{\binom{N}{2}}} = p^{M}(1-p)^{\binom{N}{2}-M}, \tag{1.16}$$

where the last equality follows by defining p in (1.13) as $p \equiv (1+e^{-\theta})^{-1}$. Then, for instance, the expected total of links in the model is

$$\mathbb{E}(M|\theta) = \sum_{G\in\mathscr{G}} M(G)\mathbb{P}(G,\theta) = \binom{N}{2}\frac{1}{1+e^{-\theta}}. \tag{1.17}$$

Given the probability distribution in (1.16), we can characterize the network model. For example, the degree distribution within the model is described by a binomial distribution, *i.e.* $\mathbb{P}(c_i = c) = \binom{N-1}{c} p^{c}(1-p)^{N-1-c}$ where c is the generic degree of the node i. Hence, in the asymptotic limit $N \to \infty$ with constant Np the degree distribution is Poisson. Finally, the model can be estimated on data by using (1.17), *i.e.* considering the expected number of links as equal to the observed one and solving for θ.

This model is known also as Bernoulli random graph and is one of the most studied of graph models. However, many scientific works, *e.g.* see [Watts and Strogatz, 1998, Strogatz, 2001, Newman et al., 2002] to name but a few, have pointed out that the Erdös-Rényi model may be inappropriate to describe real-world network systems, first of all because of observed degree distributions which are best described by heavy-tailed distributions, a phenomenon known as scale-free networks.

The fitness model (described below) has been proposed as one possible mechanism to explain generic degree distributions in real-world networks.

1.2.4 Fitness model

The fitness model describes network with nodes characterized by a latent variable, *i.e.* the *fitness*, governing the link probability for all links incident to the node. When we study the statistical inference of the fitness model given a real-world network, the fitnesses are the ones which describe nodes having on average the degree equal to observations. Hence, the fitness models represents one possible mechanism to explain generic degree distributions in real-world networks.

The fitness model has had a remarkable history. In statistical physics literature, it has been introduced by [Caldarelli et al., 2002] to capture the power law distribution of real-world networks. Then, it has been renamed as β-model by [Chatterjee et al., 2011], which have studied some asymptotic properties about identifiability and consistency of the maximum likelihood estimator. Within the modeling framework of ERG, the fitness model is known also as generalized random graphs [Park and Newman, 2004]. Applications of this version of the fitness model are countless. For instance, it has been successfully applied to describe the World Trading Web between countries by using GDP (opportunely normalized) γ_i as the

fitness of the node-country i and assuming $f(\gamma_i,\gamma_j) \equiv \frac{\delta\gamma_i\gamma_j}{1+\delta\gamma_i\gamma_j}$ as linking function with parameter δ, see [Garlaschelli and Loffredo, 2004a].

Moreover, in the literature of growing networks, for fitness model some authors mean the goodness-get-richer mechanism describing the capability of a node in creating new links as long as new nodes enters the system, depending on the latent node fitness. There exist many models capturing this mechanism. The first one was introduced by [Bianconi and Barabási, 2001] to describe a microscopic behavior leading to scale-free networks, by combining the concept of growing networks with preferential attachment driven by the node fitness. Applications of this version of the model are countless. For instance, the fitness model has been used to describe the scale-free structure of the World Wide Web considered as a growing network whose evolutionary dynamics arises from the combined effects of preferential attachment driven by the fitness model, the current state of the network, and the high rates of birth and death of pages on the web, see [Kong et al., 2008].

The fitness model is an exponential random graph having the degree sequence as the characteristic we take fixed on average over the ensemble. As shown in [Park and Newman, 2004], suppose that we measure the degrees of all the nodes. Let us denote c_i the degree of node i, then the set of all degrees $\{c_i\}_{i=1,\dots,N}$ is called the degree sequence. By using it as the characteristic describing the ERG model, the network ensemble has Hamiltonian

$$H(G,\boldsymbol{\theta}) = -\sum_{i=1}^{N} \theta_i c_i, \tag{1.18}$$

with $\boldsymbol{\theta} \equiv \{\theta_i\}_{i=1,\dots,N}$ the set of parameters and θ_i defining the fitness of node i. It is trivial to prove, in a similar fashion of (1.15), that the partition function is

$$Z(\boldsymbol{\theta}) = \prod_{i,j>i} (1+e^{\theta_i+\theta_j}), \tag{1.19}$$

resulting in a probability distribution equal to

$$\mathbb{P}(\boldsymbol{A}|\boldsymbol{\theta}) = \prod_{i,j>i} \frac{e^{A_{ij}(\theta_i+\theta_j)}}{1+e^{\theta_i+\theta_j}}, \tag{1.20}$$

where the graph G is described by its adjacency matrix $\boldsymbol{A} \equiv \{A_{ij}\}_{i,j>i}$.

The fitness model having probability distribution (1.20) describes a random graph whose nodes are characterized by a latent quantity, *i.e.* the fitness, which determines their degree. The larger is the node fitness, the larger is the probability of observing a larger node degree. In the statistically-oriented literature, some points are still under investigation, in particular how to obtain some rigorous results about the inference of the model. In fact, by reverting the point of view, assume to observe a real-world network G^*, thus asking for the values of parameters $\{\theta_i^*\}_{i=1,\dots,N}$ which describe an ensemble of graphs with G^* representing a specific realization

over the ensemble. Below, we show when the inference problem is well specified and how to obtain a point estimate of the parameters.

Finally, in the case of directed networks, each node may have both incoming and outgoing links, resulting in two types of degree, namely *out-degree* and *in-degree*. The fitness model can be easily generalized by introducing two fitnesses, *i.e.* θ_i^{out} and θ_i^{in}, for each node i. In this case, the two node fitnesses describe the propensity of the node in creating outgoing links and incoming links, respectively. Hence, (1.20) becomes

$$\mathbb{P}(\boldsymbol{A}|\boldsymbol{\theta}) = \prod_{i,j\neq i} \frac{e^{A_{ij}(\theta_i^{out}+\theta_j^{in})}}{1+e^{\theta_i^{out}+\theta_j^{in}}}, \tag{1.21}$$

where $\boldsymbol{A} = \{A^{ij}\}_{i,j=1,\dots,N,\, i\neq j}$ describes the directed graph without the possibility of self-loops.

1.2.4.1 Inference of the fitness model

MLE (1.3) of the fitness model described by (1.20) leads to the following system of equations

$$\sum_{j\neq i} \frac{e^{(\theta_i+\theta_j)}}{1+e^{(\theta_i+\theta_j)}} = c_i^*, \;\; \forall i = 1,\dots,N. \tag{1.22}$$

where c_i^* is the observed degree of node i. For the specific case of the fitness model, MLE is equivalent to the method of moments (1.8) where the moments correspond to the node degrees $\{c_i\}_{i=1,\dots,N}$.

Since $-\log L(\boldsymbol{\theta})$ of the fitness model is a convex function in the domain of parameters $\boldsymbol{\theta} \equiv \{\theta_i\}_{i=1,\dots,N} \in \mathbb{R}^N$, see [Rinaldo et al., 2013], the solution $\hat{\boldsymbol{\theta}}$ of (1.22) is the maximum in the domain, thus $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of the fitness model. [Chatterjee et al., 2011] have proved the consistency of MLE in the asymptotic limit of diverging number of nodes but fixed values for the parameters, *i.e.* dense graphs with arbitrarily large number of nodes. In this case, MLE is a consistent point estimate and the fitness model results as identifiable, since consistency implies identifiability [Gabrielsen, 1978]. Furthermore, [Csiszár et al., 2012] have proved that the degree sequence is a sufficient statistic in the dense regime. However, the problem of MLE consistency and identifiability of the fitness model in the limit of sparse graphs is still an open question.

Finally, let us describe a fast algorithm for computing the MLE, if it exists, by solving (1.22). Let us define $y_i \equiv e^{\theta_i} \;\forall i = 1,\dots,N$, $\boldsymbol{y} \equiv \{y_i\}_{i=1,\dots,N}$, $\varphi_i(\boldsymbol{y}) \equiv \sum_{j\neq i} \frac{y_j}{1+y_i y_j}$, and let $\boldsymbol{\varphi} : \mathbb{R}^N \to \mathbb{R}^N$ be the function whose i-th component is φ_i. Then,

1. start from any $\boldsymbol{y}^{(0)} \in \mathbb{R}^N$,
2. use the update rule $y_i^{(k)} = \frac{c_i^*}{\varphi_i(\boldsymbol{y}^{(k-1)})} \;\forall i = 1,\dots,N$,

3. continue until convergence.

[Chatterjee et al., 2011] have proved that this algorithm converges in the dense regime and the MLE point estimate of parameters is $\hat{\theta}_i = \log \hat{y}_i \ \forall i = 1,...N$ where $\hat{\boldsymbol{y}}$ is the convergence point.

For directed networks, the link probability conditional to the value of the fitnesses is

$$\mathbb{P}(A^t_{ij}|\theta^{out}_i, \theta^{in}_j) = \frac{e^{A_{ij}(\theta^{out}_i + \theta^{in}_j)}}{1 + e^{\theta^{out}_i + \theta^{in}_j}}, \tag{1.23}$$

where $\theta^{out}_i, \theta^{in}_j$ are the fitness of nodes i and j, outgoing and incoming respectively. In this case, each generic link probability is invariant under the following linear transformation (shift) of the fitnesses, $\theta^{out}_i \mapsto \theta^{out}_i + \tilde{c} \ \ \forall i = 1,...,N$, $\theta^{in}_j \mapsto \theta^{in}_j - \tilde{c} \ \ \forall j = 1,...,N$, with $\tilde{c} \in \mathbb{R}$. This symmetry arises because the total number of outgoing links has to be equal to the total number of incoming links at each time t and makes the model as non-identifiable. However, let us notice that, for any choice of $\tilde{c} \in \mathbb{R} \ \forall t$, the expectation of any network metric remains the same as long as the sum of the fitnesses in (1.23) remains the same. Hence, in the case of directed networks, we need to specify a given parameterization of the fitness model. The most common one is to take one (outgoing) fitness as equal to zero. Notice that the identifiability problem for the exponential family of probability distributions for directed static graphs is well known since the milestone work of [Holland and Leinhardt, 1981]. Furthermore, in successive investigations, several authors suggested the same parameterization with one zero fitness and several studies on the inference problem of the model have been accomplished, *e.g.* see [Rinaldo et al., 2013, Luo et al., 2017].

1.2.5 Stochastic block models

Sometimes the nodes of a network are partitioned into communities and the observed network structure crucially depends on the nodes' membership. Stochastic Block Model(s) (SBM), introduced by [Holland et al., 1983, Snijders and Nowicki, 1997, Nowicki and Snijders, 2001], generalizes the Erdös-Rényi random graph by giving each pair of nodes a connection probability depending on the communities they belong to, thus describing a large variety of macroscopic structures observed in real-world networks [Porter et al., 2009].

The Stochastic block model is defined as follows. Each node i is associated with a latent variable, *i.e.* the node membership or label $g_i \in \{1,2,...,k\}$ with k the number of communities, indicating which community the node belongs to. The labels are assumed to be chosen independently, where for each node i the probability that $g_i = a$ is q_a and $\boldsymbol{q} \equiv \{q_a\}_{a=1,...,k}$ the expected fractions N_a/N of nodes N_a in each group a.

Then, the $k \times k$ *affinity matrix* $\boldsymbol{p} = \{p_{ab}\}_{a,b=1,...,k}$ describes the probability of a link between group a and group b. Finally, a link A_{ij} between i and j is drawn with probability $p_{g_i g_j}$ where g_i and g_j are the labels of node i and node j, respectively.

In mathematical terms, the prior distribution of labels $\boldsymbol{g} \equiv \{g_i\}_{i=1,...,N}$ is

$$\mathbb{P}(\boldsymbol{g}|\boldsymbol{q}) = \prod_{i=1}^{N} q_{g_i} \tag{1.24}$$

and the likelihood of SBM, *i.e.* the probability of the random graph described by its adjacency matrix $\boldsymbol{A}$ conditional on the model parameters $\boldsymbol{g}$ and $\boldsymbol{p}$, is

$$\mathbb{P}(\boldsymbol{A}|\boldsymbol{g},\boldsymbol{p}) = \prod_{i,j>i} p_{g_i g_j}^{A_{ij}} (1 - p_{g_i g_j})^{1-A_{ij}}, \tag{1.25}$$

with $p_{ab} \equiv \mathbb{P}(A_{ij} = 1 | g_i = a, g_j = b)$.

According to SBM, the average number of links from group a to group b is $M_{ab} = p_{ab} N_a N_b$ or $M_{aa} = p_{aa} N_a (N_a - 1)$ if $a = b$ (because we exclude the possibility of self-loops). Then, the mean degree of a node is $\bar{c} = N/k^2 \sum_{ab} p_{ab} \equiv N\bar{p}$. We are interested in the sparse regime $\bar{p} = \bar{c}/N$, that is the most challenging from the inference perspective, since most of real networks of interest are sparse and because sparsity allows to carry out asymptotically optimal analysis. In this case, we can define a rescaled affinity matrix $c_{ab} \equiv N p_{ab}$ that is of order one by construction when $p_{ab} = O(1/N)\ \forall a, b = 1, ..., k$.

The stochastic block model is in effect an exponential random graph ERG model. It is enough to notice that the likelihood (1.25) is equivalent to the exponential probability distribution

$$\mathbb{P}(\boldsymbol{A}|\boldsymbol{g},\boldsymbol{\gamma}) = \prod_{i,j>i} \frac{e^{\gamma_{g_i g_j} A_{ij}}}{\left(1 + e^{\gamma_{g_i g_j}}\right)}, \tag{1.26}$$

where we have defined $\gamma_{g_i g_j} \equiv \log \frac{p_{g_i g_j}}{1 - p_{g_i g_j}}$.

There exist several special cases of SBM with different parameterization of the affinity matrix in order to capture different community structures, ranging from assortative mixing [Newman, 2002], core-periphery [Zhang et al., 2015], bipartite [Barucca and Lillo, 2016] up to hierarchical structures [Peixoto, 2014b], to name but a few. Furthermore, different extensions exist to capture, *e.g.* , the heterogeneity of node degree observed in real-world networks [Karrer and Newman, 2011], or to allow mixed-membership in order to describe the multiple roles of nodes in some networked systems [Airoldi et al., 2008]. See [Kim et al., 2018] for a review about stochastic block models. Finally, notice that both Erdös-Rényi and fitness models can be interpreted as limiting cases of SBM: (i) when $p_{ab} = p$, $\forall a, b = 1, ..., k$, SBM reduces to Erdös-Rényi random graphs described by (1.13); (ii) when we have one community for each node and the affinity matrix is parameterized as $p_{ij} = \text{logit}^{-1}(\theta_i + \theta_j)$, $\forall i = 1, ..., N, j > i$ where θ_i is the fitness of node i, then SBM reduces to the fitness model described by (1.20).

From the point of view of statistical inference, the information about communities is often not known a priori and we only observe the links connecting the nodes of the network. Hence, a fundamental problem is to detect the communities and understand what is the role they play in the network's structure and dynamics.

There exists several algorithms for community detection, *e.g.* see [Danon et al., 2005, Fortunato, 2010], and most of them are built by trying to 'optimize' some network metrics over the possible partitions of the nodes, *e.g.* modularity [Newman, 2006]. However, this approach suffers the problem of not providing a measure of the significance of the division into communities, and they falsely detect communities even in purely random graphs.

The problem of community detectability, *i.e.* the possibility of inferring the correct communities, at least in part, is formalized mathematically as follows. Let us consider $\boldsymbol{g} = \{g_i\}_{i=1,\dots,N}$ as the original assignments and $\hat{\boldsymbol{g}} = \{\hat{g}_i\}_{i=1,\dots,N}$ as its estimate according to some algorithm. Hence, we define the overlap between original and inferred assignement as

$$Q(\boldsymbol{g},\hat{\boldsymbol{g}}) = \max_{\{\sigma\}} \frac{\frac{1}{N}\sum_i \delta_{g_i,\sigma(\hat{g}_i)} - \max_a q_a}{1-\max_a q_a}, \tag{1.27}$$

where $\{\sigma\}$ ranges over the permutations σ on k elements and $\delta_{a,b}$ is the Kronecker delta taking value equal to one if $a=b$, zero otherwise. The overlap is defined such that if $\hat{g}_i = g_i \ \forall i = 1,\dots,N$, *i.e.* , if we find the exact labeling, then $Q=1$. On the contrary, it is $Q=0$ when the detection algorithm performs equivalently to a randon guess. When a labeling $\hat{\boldsymbol{g}}$ is correlated with the original one $\boldsymbol{g}$ the overlap is strictly positive on average. Communities are considered as *detectable* if there exists at least one algorithm for which the overlap is strictly positive on average.

Recently, there is a growing interest in understanding the limitations of community detection [Reichardt and Leone, 2008, Decelle et al., 2011b, Decelle et al., 2011a, Nadakuditi and Newman, 2012] and interpreting results obtained for empirical networks [Peel et al., 2017]. To this end, the problem of community detectability can be solved by studying the conditions for the inference of stochastic block models, in two steps: (i) first, determining what are the 'difficult' cases for community detectability, *e.g.* when all nodes of the network have the same degree in average, otherwise classifying nodes according to their degree leads always to a positive overlap; (ii) then, assessing the significance of the community structure inferred from data generated according to the stochastic block model describing such difficult case, thus determining the condition of detectability, or the so-called *detectability threshold*.

The notion of detectability threshold was first introduced for sparse graphs with block structure of connections in [Reichardt and Leone, 2008] and the results were later found using Bayesian inference [Decelle et al., 2011b, Decelle et al., 2011a], information theory [Abbe et al., 2016], and Random Matrix Theory [Nadakuditi and Newman, 2012]. The work was also extended to the dynamic case [Ghasemian et al., 2016] and for bipartite networks [Larremore et al., 2014]. Detectability threshold refers to a sharp transition in parameters space separating two distinct regimes: (i) when it is possible,

at least in part, to infer communities of nodes, and (ii) when inference cannot work better than random assignments. This phase transition is of practical importance and sheds new light on the problem of community detection algorithms optimality. A rigorous proof of the advocated phenomenon in static and symmetric case was given by [Mossel et al., 2015a, Bordenave et al., 2015]. A recent summary on the subject can be found in [Abbe, 2017]. Finally, one should also remember that even though the results were shown for block models, the conclusions are more universal [Young et al., 2018].

To explain how the detectability threshold can be found, let us consider the assortative planted partition model, which is a special case of SBM. It is the widely used benchmark in the mathematics and computer science community detection literature [Krzakala et al., 2013, Decelle et al., 2011a, Dyer and Frieze, 1989, Condon and Karp, 2001] and it is specified as the SBM with uniform prior, *i.e.* $q_r = 1/k \forall r = 1,...,k$, and affinity matrix with a constant p_{in} on the diagonal and another constant $p_{out} \leqslant p_{in}$ off diagonal.[3] In fact, for the assortative planted partition model, every group has the same average degree. In this case, the degree sequence does not contain any information about which node should be in which group, and phases exist in parameter space where the original group assignment cannot be inferred from the structure of the network.

A classical result [Dyer and Frieze, 1989] shows that for $p_{in} - p_{out} > O(\log N/N)$ the planted partition is with high probability equal to the best possible partition, in terms of minimizing the number of links between groups. Another classical result [Airoldi et al., 2008] shows that the planted partition can be easily found as long as $p_{in} - p_{out} > O(N^{-1/2+\xi})$ for arbitrarily small ξ. Furthermore, several studies [Snijders and Nowicki, 1997, Peixoto, 2014a, Peixoto, 2017, Kao et al., 2018] show great efficiency of MCMC methods in the inference of SBM on real-world data.

Finally, the problem of community detection for static graph generated according to SBM has been theoretically solved by [Decelle et al., 2011b, Decelle et al., 2011a]. In the specific case of the planted partition model, [Decelle et al., 2011a] show that polynomial-time algorithms can find a partition that is correlated with the planted partition when

$$p_{in} - p_{out} > \frac{\sqrt{k p_{in} + k(k-1) p_{out}}}{\sqrt{N}}. \tag{1.28}$$

In particular, Decelle et al. provide a Bayesian estimation method, called *Belief Propagation* (BP) algorithm, for the inference of SBM in the regime of sparse networks. This becomes exact in the asymptotic limit of diverging number of nodes but fixed mean degree. They further argue that the inference with BP is optimal, *i.e.* no better algorithm exists for learning the model parameters of the generative SBM. Finally, they provide a rather complete and asymptotically exact analysis of the phase diagram associated with the problem of community detection of the as-

[3] Equivalently for the rescaled affinity matrix described by c_{in} and c_{out}.

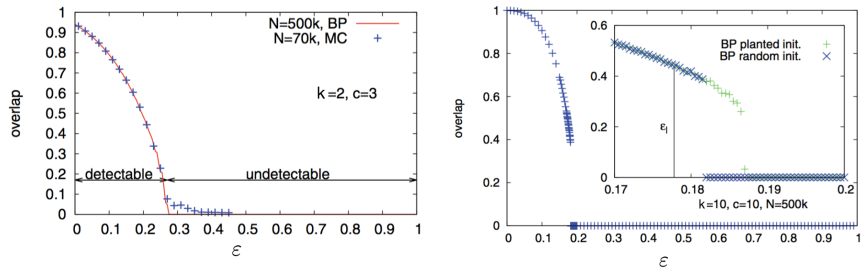


Fig. 1.2 Left: second order phase transition. Overlap (1.27) between the original assignment and its best estimate given the structure of the graph generated according to the planted partition model, computed by both BP algorithm (see Subsection 1.2.5.1) and MCMC method, in both cases with a random initialization of the algorithm. Right: first order phase transition. Overlap (1.27) obtained with BP algorithm. The inset zooms in on the critical region. The overlap with both random initialization (BP random init.) and initialization with the original assignment (BP planted init.) for the BP algorithm is shown. In both plots, graphs are generated using N nodes, a number k of groups of the same size, average degree c, and different ratios $\varepsilon = c_{out}/c_{in}$. The Figures are taken from [Decelle et al., 2011a].

sortative planted partition model, thus showing the approach to solve the problem also for different specification of SBM.

In Figure 1.2 we show two examples of a phase diagram associated with the community detection of the planted partition model in the case of k groups and average degree equal to c. In the left plot of Figure 1.2, the continuous red line represents the overlap resulting from BP inference while the crosses represent the results obtained from Gibbs sampling, *i.e.* , by using MCMC methods [Snijders and Nowicki, 1997]. In both cases, the algorithms start with a random initial group assignment. We can notice how the problem of community detection associated with the inference of SBM is characterized by a phase transition which divides the parameter space into two regions: (i) a detectable region where the observed graph contains significant information about the original group assignment, and using BP or MCMC yields an assignment that is strongly correlated with the original one; (ii) an undetectable region where the observed graph does not contain any significant information about the original group assignment, and community detection is impossible. This behavior has been analytically proved at least for $k = 2$ groups [Mossel et al., 2018]. For the planted partition model, the detectability transition is governed by the ratio $\varepsilon \equiv \frac{c_{out}}{c_{in}}$ which determines the degree of assortativity of the network, *i.e.* the link density within the same community with respect to the one between two different communities. Thus we can observe two limiting cases: (i) $\varepsilon = 1$ gives an Erdös-Rényi random graph, (ii) and $\varepsilon = 0$ gives completely disconnected groups. According to (1.28), a threshold value ε_c for ε can be found, in Figure 1.2 $\varepsilon_c \approx 0.27$.

Furthermore, it can be numerically shown that, when $k < 4$, the problem of community detectability is characterized by a second order phase transition, as in the left plot of Figure 1.2: in the regime of detectability, the BP solution displays a positive overlap with the original assignment for any random initial point used by the BP al-

gorithm, thus this case is termed *easy phase transition* for the community detection problem. On the contrary, when $k \geqslant 4$, the problem of community detectability is characterized by a first order phase transition. Moreover, there exists a region in the parameter space where inference is possible but algorithmically hard (with diverging convergence time for BP). In fact, the overlap of labels inferred by the randomly initialized algorithm becomes positive below the threshold for ε, see the right plot of Figure 1.2. On the contrary, when BP is initialized with the original assignment ('BP planted init.' in the right plot of Figure 1.2), we recover a positive overlap up to the detectability threshold. In this region, inference is possible but requires an extensive investigation in the parameter space. Thus, it is termed *hard phase transition* for the community detection problem. See [Decelle et al., 2011a] for further details.

In Subsection 1.2.5.1 we better explain how to face the problem of community detection for SBM and some technical aspects about the inference of both labels and parameters are presented. These results represent some building blocks of the research in Chapters 3 and 4.

Finally, let us notice that introducing link directions has a non trivial impact on detectability of communities in SBM. In particular, non-zero correlation with the correct assignments can be achieved not only by increasing assortativity but also by increasing asymmetry, *i.e.* given the total number of links between two groups, there are more links in one direction than the other one. The detectability condition can be analytically assessed as a function of the parameters governing the assortativity and the asymmetry of the network. Moreover, the type of detectability transition depends non trivially on the number of groups. This research is presented in Chapter 3.

1.2.5.1 Inference of SBM

Let us assume to observe a graph G generated according to a stochastic block model and we aim to infer the most likely values of model parameters as well as what is the most likely assignment of a label to a given node. This inference problem is solved by [Decelle et al., 2011b, Decelle et al., 2011a] which adopt a Bayesian approach. Throughout this Subsection we review the estimation method for the inference of communities which is named as *Belief Propagation* algorithm and it is based on the so called *replica symmetric cavity method* [Mézard and Parisi, 2001]. We show also how to learn the other parameters of SBM, by using the BP algorithm.

Let us consider the set of SBM parameters $\boldsymbol{\pi} \equiv \{\{q_a\}_{a=1,\dots,k}, \{p_{ab}\}_{a,b=1,\dots,k}\}$ for a given number k of groups, along with a given group assignment $\boldsymbol{g} = \{g_i\}_{i=1,\dots,N}$ for the N nodes of the network. Finally, assume to observe a graph G generated by SBM, having $\boldsymbol{A}$ as adjacency matrix.

The probability that the stochastic block model generates G with group labels $\boldsymbol{g}$, *i.e.* the SBM likelihood, is

$$\mathbb{P}(\boldsymbol{A}, \boldsymbol{g} | \boldsymbol{\pi}) = \prod_{i,j>i} p_{g_i g_j}^{A_{ij}} (1 - p_{g_i g_j})^{1-A_{ij}} \prod_i q_{g_i}. \tag{1.29}$$

The node labels $\boldsymbol{g}$ are, however, latent variables which we do not observe, thus standard maximum likelihood methods cannot be applied for learning $\boldsymbol{\pi}$.

A Bayesian approach is adopted. In particular, the two sets of parameters, *i.e.* $\boldsymbol{g}$ and $\boldsymbol{\pi}$, are not inferred overall but the two different posterior distributions for both $\boldsymbol{g}$ and $\boldsymbol{\pi}$ are 'optimized' iteratively, thus giving rise to an Expectation- (inference of the group assignment $\boldsymbol{g}$) Maximization (learning of parameters $\boldsymbol{\pi}$) algorithm. It is in effect a Maximum A Posteriori (MAP) estimation.

Given $\boldsymbol{\pi}$, then the (*posterior*) probability of a given group assignment $\boldsymbol{g}$ given $\boldsymbol{\pi}$ is (using Bayes' rule)

$$\mathbb{P}(\boldsymbol{g}|\boldsymbol{A},\boldsymbol{\pi}) = \frac{\mathbb{P}(\boldsymbol{A},\boldsymbol{g}|\boldsymbol{\pi})}{\sum_{\{t_i\}}\mathbb{P}(\boldsymbol{A},\{t_i\}|\boldsymbol{\pi})}, \tag{1.30}$$

where the sum runs over all possible assignments $\{t_i\}\ \forall i$, or, equivalently,

$$\mathbb{P}(\boldsymbol{g}|\boldsymbol{A},\boldsymbol{\pi}) = \frac{e^{-H(\boldsymbol{g}|\boldsymbol{A},\boldsymbol{\pi})}}{Z(\boldsymbol{A},\boldsymbol{\pi})}, \tag{1.31}$$

where

$$H(\boldsymbol{g}|\boldsymbol{A},\boldsymbol{\pi}) = -\sum_i \log q_{g_i} - \sum_{i,j>i}\left(A_{ij}\log c_{g_i g_j} + (1-A_{ij})\log\left(1-\frac{c_{g_i g_j}}{N}\right)\right) \tag{1.32}$$

is the Hamiltonian of the stochastic block model and the partition function is $Z(\boldsymbol{A},\boldsymbol{\pi}) \equiv \sum_{t_i} e^{-H(\{t_i\}|\boldsymbol{A},\boldsymbol{\pi})}$.

In order to infer what is the most likely assignment of the label of node i, we need to compute the (*posterior*) *marginal*

$$\nu_i(g_i) = \sum_{\{g_j\}_{j\neq i}} \mathbb{P}(\boldsymbol{g}|\boldsymbol{A},\boldsymbol{\pi}), \tag{1.33}$$

that is the probability $\nu_i(g_i)$ that node i belong to group g_i, with g_i ranging from 1 to k. It is in effect the posterior distribution of the label of node i (whereas q_i determines the prior distribution).

Thus, the best estimate $\hat{g}_i$ for the label of node i is

$$\hat{g}_i = \arg\max_{g_i} \nu_i(g_i). \tag{1.34}$$

Now, assume to know the posterior marginal distribution of the group assignments $\boldsymbol{g}$. The posterior distribution of parameter $\boldsymbol{\pi}$ can be obtained by using the Bayes' rule as

$$\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{A}) = \frac{\mathbb{P}(\boldsymbol{\pi})}{\mathbb{P}(\boldsymbol{A})}\mathbb{P}(\boldsymbol{A}|\boldsymbol{\pi}) = \frac{\mathbb{P}(\boldsymbol{\pi})}{\mathbb{P}(\boldsymbol{A})}\sum_{\{g_i\}}\mathbb{P}(\boldsymbol{A},\{g_i\}|\boldsymbol{\pi}), \tag{1.35}$$

where $\mathbb{P}(\boldsymbol{\pi})$ is the prior distribution which can include any graph-independent information we might have about the values of the parameters. In absence of any prior

information, we can assume a uniform prior, *i.e.* $\mathbb{P}(\boldsymbol{\pi}) = 1$ up to normalization. Notice, however, that since the number of terms in the sum in (1.35) typically grows exponentially with N, we can assume any smooth prior as long as it is independent of N. In fact, the prior is less and less important when N grows, leaving us with the same posterior distribution we would have if the prior were uniform. By considering $f(\boldsymbol{\pi},\boldsymbol{A}) = -\frac{\log Z(\boldsymbol{\pi},\boldsymbol{A})}{N}$ (referred as 'free-energy' in Statistical Physics) and maximizing it over $\boldsymbol{\pi}$, we obtain the equations, also known as *Nishimori conditions*, for learning the model parameters. We obtain for q_a, subject to the constraint $\sum_a q_a = 1$,

$$q_a = \frac{1}{N}\sum_i \mathbb{E}(\delta_{g_i a}) = \frac{\mathbb{E}(N_a)}{N} \;\; \forall a = 1,...,k, \tag{1.36}$$

where the expected value is obtained by averaing over the marginal distribution (1.30).

Analogously for the generic element of the rescaled affinity matrix c_{ab}, we obtain

$$c_{ab} = \frac{1}{N q_a q_b}\sum_{(i,j)\in E} \mathbb{E}(\delta_{g_i a}\delta_{g_j b}) = \frac{\mathbb{E}(M_{ab})}{N q_a q_b} \;\; \forall a,b = 1,...,k,\, a \neq b, \tag{1.37}$$

where the sum is over the set of edges. That is, the most likely value of c_{ab} is proportional to the average number of edges from group a to group b. When $a = b$, it is

$$c_{aa} = \frac{1}{N\frac{q_a^2}{2}}\sum_{(i,j)\in E} \mathbb{E}(\delta_{g_i a}\delta_{g_j a}) = \frac{\mathbb{E}(M_{aa})}{N\frac{q_a^2}{2}} \;\; \forall a = 1,...,k. \tag{1.38}$$

Hence, the inference of SBM can be accomplished by optimizing iteratively the posterior marginals for $\boldsymbol{g}$ (1.33) (E-step) and the posterior distribution of parameters $\boldsymbol{\pi}$ (1.35) (M-step). The method which allows to maximize the marginal (1.33) is the Belief Propagation algorithm described below.

1.2.5.2 Belief Propagation (BP) algorithm

The belief propagation equations are derived from a recursive computation of the partition function with the assumption that the network is a tree. In fact, in the sparse regime $c_{ab} = O(1)$ graphs generated according to the SBM are locally treelike, *i.e.* almost all nodes have a neighborhood which is a tree up to distance $O(\log N)$. Then, in the limit $N \to \infty$, BP equations become asymptotically exact.

Let us define the conditional marginals, or *messages*, $\psi_{g_i}^{i\to j}$ as the marginal probability that node i belongs to group g_i if the node j were absent. BP assumes that the dependence between neighbors of i is short-range, *i.e.* it is mediated only by i. Thus, if i were missing, or if its label were fixed, the distribution of the states of the neighbors would be a product distribution, because of independence. In practice, let

us consider the marginal probability that i belongs to g_i starting from (1.30), *i.e.*

$$\mathbb{P}(g_i|\boldsymbol{A},\boldsymbol{\pi}) \propto \sum_{\{g_k\}_{k\neq i}} \mathbb{P}(g_i|\{g_k\}_{k\neq i},\boldsymbol{A},\boldsymbol{\pi})\mathbb{P}(\{g_k\}_{k\neq i}|\boldsymbol{A},\boldsymbol{\pi}). \tag{1.39}$$

Then, by assuming treelike approximation, the posterior $\mathbb{P}(\{g_k\}_{k\neq i}|\boldsymbol{A},\boldsymbol{\pi})$ can be factorized as a product distribution of the messages that neighbors k send to i, thus obtaining the following equation for the message that i sends to a generic node j,

$$\psi^{i\to j}_{g_i} = \frac{1}{Z^{i\to j}} q_{g_i} \prod_{k\neq i,j} \left(\sum_{g_k} c^{A_{ik}}_{g_i g_k} \left(1 - \frac{c_{g_i g_k}}{N}\right)^{1-A_{ik}} \psi^{k\to i}_{g_k} \right), \tag{1.40}$$

where $Z^{i\to j}$ is a normalization constant ensuring $\sum_{g_i} \psi^{i\to j}_{g_i} = 1$.

Equation (1.40) represents the most general BP equations to be applied iteratively up to convergence to the fixed point $\{\bar{\psi}^{i\to j}_{g_i}\}$. Then, the marginal probability (1.33) is estimated to be $\nu_i(g_i) = \psi^i_{g_i}$ where

$$\psi^i_{g_i} = \frac{1}{Z^i} q_{g_i} \prod_{k\neq i} \left(\sum_{g_k} c^{A_{ij}}_{g_i g_k} \left(1 - \frac{c_{g_i g_k}}{N}\right)^{1-A_{ij}} \bar{\psi}^{k\to i}_{g_k} \right), \tag{1.41}$$

where Z^i is the normalization term.

However, the product over messages in (1.40) is over all possible 'neighbors' of i except j, but for both links $A_{ik} = 1$ (weighted with $c_{g_i g_k}$) and no-links $A_{ik} = 0$ (weighted with $1 - \frac{c_{g_i g_k}}{N}$). Hence, in general, we have $N(N-1)$ messages to update at each iteration of the algorithm, thus requiring $O(N^2)$ time. This is suitable only for networks with a low number of nodes. Notice, however, that in the sparse regime where $c_{ab} = O(1)$ and for large N, the interactions mediated by no-links are of subleading order in N, then we can neglect them and considering only $2M$ messages where M is the total number of links. That is,

$$\begin{aligned} \psi^{i\to j}_{g_i} &= \frac{1}{Z^{i\to j}} q_{g_i} \left[\prod_{k\in\partial i\setminus j} \left(\sum_{g_k} c_{g_i g_k} \psi^{k\to i}_{g_k} \right) \right] \left[\prod_{k\notin\partial i} \left(1 - \frac{1}{N} \sum_{g_k} c_{g_i g_k} \psi^{k\to i}_{g_k} \right) \right] \approx \\ &\approx \frac{1}{Z^{i\to j}} q_{g_i} e^{-h_i} \prod_{k\in\partial i\setminus j} \left(\sum_{g_k} c_{g_i g_k} \psi^{k\to i}_{g_k} \right), \end{aligned} \tag{1.42}$$

where ∂i represents the set of neighbors of i, and we neglected terms that contribute $O(1/N)$ to $\psi^{i\to j}_{g_i}$, by replacing them with an external field

$$h_i = \frac{1}{N} \sum_k \sum_{g_k} c_{g_i g_k} \psi^k_{g_k}. \tag{1.43}$$

Finally, (1.41) becomes

$$\psi^i_{g_i} = \frac{1}{Z^i} q_{g_i} e^{-h_i} \prod_{k \in \partial i} \left(\sum_{g_k} c_{g_i g_k} \bar{\psi}^{k \to i}_{g_k} \right), \tag{1.44}$$

where $\{\bar{\psi}^{k\to i}_{g_k}\}$ is the fixed point of the approximated BP equations (1.42) applied iteratively up to convergence. The normalization term Z^i can be expressed in terms of the fixed point as $Z^i = \sum_{g_i} q_{g_i} e^{-h_i} \prod_{k\in\partial i} \left(\sum_{g_k} c_{g_i g_k} \bar{\psi}^{k\to i}_{g_k}\right)$.

Then, the equations (1.36-1.38) to be solved for learning the model parameters can be expressed in terms of the messages as

$$q_a = \frac{1}{N} \sum_i \psi^i_{g_a} \;\; \forall a = 1, ..., k, \tag{1.45}$$

$$c_{ab} = \frac{1}{N q_a q_b} \sum_{(i,j)\in E} \frac{c_{ab}(\psi^{i\to j}_a \psi^{j\to i}_b + \psi^{i\to j}_b \psi^{j\to i}_a)}{Z^{ij}} \;\; \forall a, b = 1, ..., k, \, a \neq b, \tag{1.46}$$

$$c_{aa} = \frac{1}{N q_a^2/2} \sum_{(i,j)\in E} \frac{c_{aa}(\psi^{i\to j}_a \psi^{j\to i}_a)}{Z^{ij}} \;\; \forall a = 1, ..., k, \tag{1.47}$$

where $Z^{ij} = \sum_{a<b} c_{ab}(\psi^{i\to j}_a \psi^{j\to i}_b + \psi^{i\to j}_b \psi^{j\to i}_a) + \sum_a c_{aa}(\psi^{i\to j}_a \psi^{j\to i}_a)$ for $(i,j) \in E$.

Hence, the BP algorithm for the inference of SBM works as it follows:

1. initialize randomly the k-component of the vector $\{\psi^{i\to j}_{g_i}\}$ for each $(i,j) \in E$;
2. initialize randomly both $q_a \; \forall a = 1, .., k$ and $c_{ab} \; \forall a \geqslant b$;
3. initialize the marginals (1.44);
4. update the messages up to convergence by applying iteratively (1.42);
5. update the marginals (1.44);
6. learn the new parameters (1.45-1.47);
7. continue up to convergence.

For further technical details see [Decelle et al., 2011a].

Finally, notice that the BP algorithm needs the number k of groups as initial input. In order to infer what is the best number of groups describing a given observed network, different methods can be applied, *e.g.* Bayesian Information Criterium [Friedman et al., 2001], once we have estimated the SBM model for different k and the marginal likelihood of data has been obtained, see [Decelle et al., 2011a].

1.2.5.3 Phase transition in community detection

When we study the inference of the group assignment of the nodes in a network generated according to the stochastic block model, the most difficult case is represented by nodes having the same average degree c. In fact, any other case includes

information on degree heterogeneity which can be exploited together with BP inference in order to obtain a positive correlation between inferred group assignment and the real one, *i.e.* what we define as *detectablility*. Hence, when aiming to define the condition for detectability, we restrict the analysis to the difficult case.

In Figure 1.2, we show the transition from detectability to undetectability of communities for the planted partition model governed by the parameter $\varepsilon = c_{out}/c_{in}$ describing the level of assortativity of the network. In particular, in the asymptotic limit $N \to \infty$ there exists a threshold value ε_c, above which communities are not detectable, *i.e.* BP works equivalently to a random guess.

In order to see this, let us notice that $\psi_a^{i\to j} = q_a$ is always a fixed point of (1.42). This solution corresponds to the random assignment of labels according to the expected number of nodes in each group $a = 1,...,k$. In fact, it does not depend on indexes i and j, thus it does not account for the information on the observed network. Then, this fixed point does not provide any information about the original assignment, *i.e.* it is no better than a random guess.

Hence, let us consider a perturbation of this solution with the aim of studying when this fixed point becomes unstable, thus converging to another solution of (1.42). To this end, let us assume that the network is a tree with d levels and we aim to study how a perturbation on one leaf propagates to the root. A perturbation

$$\psi_a^j = q_a + s_a^j \tag{1.48}$$

propagating to a neighbor l can be described by the transfer matrix

$$T_{ab} \equiv \left.\frac{\partial \psi_a^l}{\partial \psi_b^j}\right|_{\boldsymbol{\psi}=\boldsymbol{q}} = \left.\left(\frac{c_{ab}\psi_a^l}{\sum_b c_{ab}\psi_b^j} - \psi_a^l \frac{\sum_a c_{ab}\psi_a^l}{\sum_b c_{ab}\psi_b^j}\right)\right|_{\boldsymbol{\psi}=\boldsymbol{q}} = q_a\left(\frac{c_{ab}}{c} - 1\right), \tag{1.49}$$

where we have used (1.42) to leading order in N in order to derive this expression. Then, a perturbation on the leaf at level d, *i.e.* $s_{g_d}^{l_d}$, propagates to the root (level 0), *i.e.* $s_{g_0}^{l_0}$, as

$$s_{g_0}^{l_0} = \sum_{\{g_i\}_{i=1,\dots,d}} \left(\prod_{i=0}^{d-1} T_{g_i g_{i+1}}\right) s_{g_d}^{l_d}. \tag{1.50}$$

Observe that the transfer matrix does not depend on the node index, thus (1.50) becomes $s^{l_0} = (T)^d s^{l_d}$. When $d \to \infty$, $(T)^d$ is dominated by the largest eigenvalue λ, *i.e.* $s^{l_0} \approx (\lambda)^d s^{l_d}$. When we consider the perturbations from all $(c)^d$ of the leaves, the influence on the root is zero because each perturbation has zero mean. However, for the variance we have $\mathbb{E}\left((s^{l_0})^2\right) \approx \mathbb{E}\left(\left(\sum_l^{(c)^d} (\lambda)^d s^{l_d}\right)^2\right) \approx (c)^d(\lambda)^{2d}\mathbb{E}((s^{l_d})^2)$, thus obtaining the following criterion for stability,

$$c\lambda^2 = 1. \tag{1.51}$$

Hence, we have

1. for $c\lambda^2 < 1$, the perturbations vanish as we move throughout the tree and the trivial fixed point is stable. In this case, BP inference is equivalent to a random assignment;
2. for $c\lambda^2 > 1$, the perturbation is amplified exponentially, the trivial fixed point is unstable, and communities are easily detectable because of convergence to a new fixed point accounting for network information.

In the case of the assortative planted partition model with k groups, the largest eigenvalue of the transfer matrix can be computed and it is $\lambda = \frac{c_{in}-c_{out}}{kc}$. Then, we obtain the following criterion for detectability,

$$|c_{in} - c_{out}| > k\sqrt{c}, \tag{1.52}$$

or, equivalently, Equation (1.28).

1.2.6 Other network models

In the previous Subsections we reviewed the statistical network models which are of particular interest for this thesis. In particular, we introduced the general framework of exponential random graphs which allows, in principle, to describe network ensembles with specified characteristics. Two possible applications are represented by the Erdös-Rényi random graphs and the fitness model, *i.e.* when the characteristics we aim to describe are the total number of links and the degree sequence of nodes, respectively. Then, we introduced the stochastic block model which describes the community structure of a network.

Describing different elements of a network leads to different models. Here, we mention the most notable examples and refer to [Newman, 2010, Goldenberg et al., 2010] for a complete review. For example, in the case of directed graphs, when we aim to describe the node degree as well as link reciprocity, *i.e.* how much it is likely that two nodes are mutually linked, we can focus on the probability of the dyadic pairing between nodes, thus considering distinctly the four cases: whether node i links to j, j to i, neither, or both. By extending the fitness model (1.21), we can introduce a new parameter for each pairing of nodes describing locally the effect of reciprocity. This model is known as p_1 model and has been studied for the first time by [Holland and Leinhardt, 1981]. The drawback refers to identifiability of the model in the most general setting, thus only few special cases are identifiable and of interest. Finally, a generalization can be obtained by considering the p_1 model parameters as a sample drawn from some underlying distribution, and then estimate the parameters of that distribution. This Bayesian extension is known as p_2 model [Van Duijn et al., 2004, Zijlstra et al., 2006].

Another rich line of research refers to a class of models that describes the probabilistic dependence between the latent state variables and the observed network measurements, namely latent space models of networks, whose milestone work is represented by [Hoff et al., 2002]. The intuition at the core of latent space models is that

each node i can be represented as a point, *i.e.* the latent state, in an Euclidean space and the link probability between two nodes is determined by the distance in the latent space. The work by [Hoff et al., 2002] has been recently extended in a number of directions to include treatment of transitivity, homophily on node-specific attributes, clustering, and heterogeneity of nodes [Hoff, 2005, Krivitsky et al., 2009]. Moreover, the study of how nodes cluster in the latent space opens naturally to the problem of community detection. In fact, [Handcock et al., 2007] introduce an explicit clustering model in the latent space in order to infer some kind of community structure of networks. Finally, a natural extension to the dynamic case can be obtained by allowing for the latent positions to change over time. This has been firstly introduced by [Sarkar and Moore, 2006] and a number of extensions exist.

In the next Section, we present some generalizations of the reviewed models to the dynamic case, by highlighting for each of them the temporal patterns they are able to capture.

1.3 Statistical models of dynamic networks

Any system is dynamic in nature. When we observe that system at some time scale which allows appreciating how its structure changes over time, then we need to adopt dynamic models to describe its evolution. In the case of a networked system, a common approach is to assume knowing the time scale at which it changes over time, thus studying a sequence of graphs, *i.e.* a time series of network snapshots described by a set of ordered adjacency matrices $\boldsymbol{\mathcal{A}} \equiv \{\boldsymbol{A}^t\}_{t=1,\dots,T}$.

Hence, there exist two main approaches in the statistical modeling of dynamic networks: (i) considering models which explicitly describe how network metrics evolve in time, or (ii) describing a latent process for some node-specific characteristics, thus resulting in the conditional independence of each network snapshot given the current information on the latent variables.

The milestone work for the first class of model is represented by [Hanneke et al., 2010] which extends exponential random graphs to the dynamic case, namely *Temporal Exponential Random Graphs* (TERG). In particular, Hanneke et al. propose a general framework to describe ensembles of dynamic networks by making a Markov assumption on the evolution of the network from one snapshot to the next one, *i.e.*

$$\mathbb{P}(\boldsymbol{\mathcal{A}}|\boldsymbol{A}^0,\boldsymbol{\pi}) = \prod_{t=1}^{T} \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\boldsymbol{\pi}), \tag{1.53}$$

where $\boldsymbol{\pi}$ is some set of static parameters. We refer to this approach also as *static parameter models of dynamic networks* and review it and other possible generalizations in Subsection 1.3.1.

Regarding the second class of models, the key idea behind introducing latent variables into network analysis is to capture various forms of dependence between links, but preserving the conditional independence snapshot by snapshot. That is,

given the latent variables $\boldsymbol{\Theta} \equiv \{\boldsymbol{\theta}^t\}_{t=0,1,\dots,T}$, we have

$$\mathbb{P}(\boldsymbol{\mathcal{A}}|\boldsymbol{\Theta}) = \prod_{t=1}^{T} \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{\theta}^t). \tag{1.54}$$

This approach, namely *time-varying parameter models of dynamic networks*, encompasses many network models with latent variables, such as latent space models or stochastic block models. For these models, each node is associated with a latent variable, *e.g.* fitness, position in the latent space, or community membership, and the link probability depends on the unobserved characteristics of nodes. Depending on the specific latent process and how the link probability is described, we are able to capture specific network structures and how they evolve over time. We review these models in Section 1.3.2. Finally, in Section 1.3.3 we face the problem of community detection in the case of SBM with dynamic community structure, by showing how the detectability condition is modified in the dynamic case.

1.3.1 Dynamic graph ensembles

In a similar fashion of the static case, TERG are defined as follows. Let us assume to observe a sequence of graphs at $T+1$ time snapshots, each network snapshot at time t described by its adjacency matrix $\boldsymbol{A}^t$. The dynamic network ensemble is described by the probability distribution $\mathbb{P}(\boldsymbol{A}^1,\dots,\boldsymbol{A}^T|\boldsymbol{A}^0,\boldsymbol{\pi})$ which factorizes as in (1.53) according to the Markov assumption. Then, inspired by ERG, [Hanneke et al., 2010] define the conditional transition probability as

$$\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\boldsymbol{\pi}) = \frac{1}{Z(\boldsymbol{A}^{t-1},\boldsymbol{\pi})} e^{\boldsymbol{\pi}'\boldsymbol{\Psi}(\boldsymbol{A}^t,\boldsymbol{A}^{t-1})}, \tag{1.55}$$

where $\boldsymbol{\Psi} : \mathbb{R}^{N\times N} \times \mathbb{R}^{N\times N} \to \mathbb{R}^n$ maps the pair of consecutive network snapshots to n network metrics, N is the number of nodes, $Z(\boldsymbol{A}^{t-1},\boldsymbol{\pi})$ is the partition function at generic time t, and $\boldsymbol{\pi}$ is the vector of n static parameters, one for each considered network metric. Equation (1.55) describes the general framework of temporal exponential random graph (TERG) models.

As explicit examples, let us consider two network measurements, *i.e.* the total number of links and the overall link persistence (or link stability),

$$\Psi_{density}(\boldsymbol{A}^t,\boldsymbol{A}^{t-1}) = \sum_{i,j>i} A^t_{ij}, \tag{1.56}$$

$$\Psi_{persistence}(\boldsymbol{A}^t,\boldsymbol{A}^{t-1}) = \sum_{i,j>i} \left[A^t_{ij}A^{t-1}_{ij} + (1-A^t_{ij})(1-A^{t-1}_{ij})\right]. \tag{1.57}$$

By considering $\boldsymbol{\Psi}(\boldsymbol{A}^t,\boldsymbol{A}^{t-1}) \equiv (\Psi_{density}(\boldsymbol{A}^t,\boldsymbol{A}^{t-1}),\Psi_{persistence}(\boldsymbol{A}^t,\boldsymbol{A}^{t-1}))'$ and $\boldsymbol{\pi} \equiv (\theta,\mu)' \in \mathbb{R}^2$, the TERG model is specified by the following probability distribution

$$\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\theta,\mu) = \prod_{i,j>i} \frac{e^{A_{ij}^t(\theta+\mu(2A_{ij}^{t-1}-1))}}{1+e^{\theta+\mu(2A_{ij}^{t-1}-1)}}. \tag{1.58}$$

Equation (1.58) describes a temporal generalization of the Erdös-Rényi model where θ determines the link probability (see (1.16) for a comparison with the static case) and μ measure the overall link (or no-link) persistence. In this simple case, the TERG model can be estimated by MLE methods.

We discuss this simple case because of interest for the research presented in the following. However, one can extend this framework to capture the temporal patterns of other metrics, such as reciprocity [Garlaschelli and Loffredo, 2004b], transitivity [Holland and Leinhardt, 1971], and longer-range Markov dependencies, to name but a few. For example, persistence patterns of reciprocated relations together with all possible triadic configurations in social networks have been investigated in TERG by [Robins and Pattison, 2001]. Furthermore, [Krivitsky and Handcock, 2014] have studied longer-range Markov dependencies in TERG as well as possible asymmetries in the persistence of links and no-links in social networks, according to the intuition that social processes resulting in the creation of a 'link' could be not the same as those resulting in the disruption of a relation (no-link). In these cases, MLE estimation methods can be adopted as well as MCMC sampling techniques, see [Snijders, 2002, Hanneke et al., 2010, Caimo and Friel, 2011] for further details. Recently, longer-range Markov dependencies for links have been studied also by [Peixoto and Rosvall, 2017] with the aim of selecting the most appropriate Markov order which better capture the evolution of the network. [Zhang et al., 2017] have proposed generalizations of a number of standard network models, including the classic random graph and the configuration model, to the case of dynamic networks in a similar fashion to [Peixoto and Rosvall, 2017].

When we are interested in studying the persistence of each link, we can 'neglect' the network, thus focusing on the time series analysis of single binary sequences of links. To this end, the DAR(p) process [Jacobs and Lewis, 1978b] represents a possible scheme for obtaining a stationary sequence of binary random variables which has $p-th$ order Markov dependence and a specified marginal distribution.

1.3.1.1 Discrete AutoRegressive (DAR) network models

The discrete autoregressive model DAR(p) has been introduced for the first time by [Jacobs and Lewis, 1978b]. It describes, in the more general setting, a categorical variable which has p-th order Markov dependence and a multinomial marginal distribution. In the case of a binary random variable X, the marginal distribution is Bernoulli. If X^t is the realization of the random variable at time t, we have

$$X^t = V^t X^{t-l_t} + (1-V^t)Y^t \tag{1.59}$$

where $X^t \in \{0,1\}\,\forall t$, $V^t \sim \mathscr{B}(\alpha)$ is a Bernoulli random variable with $\alpha \in [0,1]$, $l_t \sim \mathscr{M}(\beta_1,...,\beta_p)$ is a multinomial random variable with $\sum_{i=1}^{p}\beta_i = 1$ and Y^t is a binary random variable sampled according to the Bernoulli marginal distribution $\mathscr{B}(\chi)$ with $\chi \in [0,1]$. In other words, at each time V^t determines if copying from the past or sampling according to the marginal. When we copy from the past, then the multinomial random variable l_t selects the time lag and, accordingly, which past realization of X we copy.

The process defined by (1.59) has the property that the correlation at any lag is larger than or equal to zero. This is by construction, since $\alpha, \beta_1, ..., \beta_p$ are non-negative definite. Thus, the DAR(p) model describes only patterns of persistence for the binary random variable. Furthermore, the correlation structure of (1.59) is equivalent to the one of the standard autoregressive AR(p) process for real-valued random variable, see [Jacobs and Lewis, 1978b], and Yule-Walker equations associated with (1.59) are formally the same of AR(p). Hence, the method of moments can be used for model estimation [Tsay, 2005] as well as MLE. The method of moments is equivalent to solving the following Yule-Walker equations,

$$\rho_m = \alpha \sum_{k=1}^{p} \beta_k \rho_{m-k}$$

where $m = 1,...,p$ and ρ_m the autocovariance function of X^t at lag m, yielding p equations ($\beta_1,...,\beta_p$ are not independent because of the condition $\sum_{m=1}^{p}\beta_m = 1$).

As explanatory example, let us consider the case of the DAR(1) model applied to a networked system. We can model link stability with the following discrete autoregressive process,

$$A_{ij}^t = V_{ij}^t A_{ij}^{t-1} + (1-V_{ij}^t)Y_{ij}^t \;\; \forall i,j = 1,...,N \text{ and } j > i \tag{1.60}$$

where $V_{ij}^t \sim \mathscr{B}(\alpha_{ij})$ with $\alpha_{ij} \in [0,1]$, $Y_{ij}^t \sim \mathscr{B}(\chi_{ij})$ with $\chi_{ij} \in [0,1]$. In (1.60), the value of A_{ij}^t is copied from the past value with probability α_{ij} or obtained by tossing a coin according to the marginal distribution $\mathscr{B}(\chi_{ij})$ with probability $1-\alpha_{ij}$. Highly persistent links (or no-links) are described by high values of α_{ij}. As a consequence, networks characterized by high values of $\boldsymbol{\alpha} \equiv \{\alpha_{ij}\}_{i,j=1,...,N}$ tend to preserve significantly the past structure through time. The persistence pattern of this model can be quantified by the autocorrelation functions (ACF) of the links. It is the one of a standard autoregressive process AR(1) but with non negative autoregressive coefficient α_{ij}, i.e. the DAR(1) model is able to describe only non negative ACF. The generalization of this model to directed networks is simply obtained by considering not symmetric adjacency matrices. This model, termed DAR(1) network model, has been proposed for the first time in the context of dynamic networks in [Mazzarisi et al., 2019a].

The probability of observing a given sequence of graphs $\boldsymbol{\mathcal{A}} = \{\boldsymbol{A}^1,...,\boldsymbol{A}^T\}$ conditional to observation $\boldsymbol{A}^0$ according to the DAR(1) model is

$$\begin{aligned}\mathbb{P}(\boldsymbol{\mathcal{A}}|\boldsymbol{A}^0,\boldsymbol{\alpha},\boldsymbol{\chi}) &= \prod_{t=1}^{T}\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\boldsymbol{\alpha},\boldsymbol{\chi}) = \prod_{t=1}^{T}\prod_{i,j>i}\mathbb{P}(A_{ij}^t|A_{ij}^{t-1},\alpha_{ij},\chi_{ij}) = \\ &= \prod_{t=1}^{T}\prod_{i,j>i}\left(\alpha_{ij}\delta_{A_{ij}^t,A_{ij}^{t-1}} + (1-\alpha_{ij})\chi_{ij}^{A_{ij}^t}(1-\chi_{ij})^{1-A_{ij}^t}\right),\end{aligned} \tag{1.61}$$

where $\boldsymbol{\chi} = \{\chi_{ij}\}_{i,j=1,\dots,N}$ and $\delta_{A_{ij}^t,A_{ij}^{t-1}}$ is the indicator function (or Kronecker delta) taking value equal to 1 if $A_{ij}^t = A_{ij}^{t-1}$ and zero otherwise. Equation (1.61) describes $\binom{N}{2}$ independent Markov chains of order one for each link. This model of dynamic networks is fully determined by the $N(N-1)$ parameters $\{\boldsymbol{\alpha},\boldsymbol{\chi}\} \equiv \{\alpha_{ij},\chi_{ij}\}_{i=1,\dots,N;j>i}$ and we can estimate them by maximum likelihood method. [Jacobs and Lewis, 1978a] describe, in a second work, the asymptotic properties of generic DAR(p) processes, thus proving the consistency of maximum likelihood estimator. For completeness, the maximum likelihood estimation of DAR(1) can be obtained as follows.

Let us focus on a sequence of links $\{A_{ij}^t\}^{t=0,1,\dots,T}$. We aim to find the maximum likelihood estimators of α_{ij} and χ_{ij}. The probability (likelihood) of the observations is

$$\begin{aligned}\mathbb{P}(\{A_{ij}^t\}^{t=0,1,\dots,T}|\alpha_{ij},\chi_{ij}) &= \prod_{t=1}^{T}\mathbb{P}(A_{ij}^t|A_{ij}^{t-1},\alpha_{ij},\chi_{ij}) = \\ &= \prod_{t=1}^{T}\left(\alpha\delta_{A_{ij}^t,A_{ij}^{t-1}} + (1-\alpha)\chi^{A_{ij}^t}(1-\chi)^{1-A_{ij}^t}\right).\end{aligned} \tag{1.62}$$

The maximum likelihood estimators of α_{ij} and χ_{ij} are the values which maximize the posterior $\mathbb{P}(\alpha_{ij},\chi_{ij}|\{A_{ij}^t\}^{t=0,1,\dots,T})$ or, equivalently, $\log\mathbb{P}(\alpha_{ij},\chi_{ij}|\{A_{ij}^t\}^{t=0,1,\dots,T})$. Hence, possible candidates are the solution of the following system of equations

$$\partial_{\alpha_{ij},\chi_{ij}}\sum_{t=1}^{T}\log\left(\alpha_{ij}\delta_{A_{ij}^t,A_{ij}^{t-1}} + (1-\alpha_{ij})(\chi_{ij})^{A_{ij}^t}(1-\chi_{ij})^{1-A_{ij}^t}\right) = 0, \tag{1.63}$$

or more explicitly,

$$\begin{cases} \sum_{t=1}^{T}\dfrac{\delta_{A_{ij}^t,A_{ij}^{t-1}} - (\chi_{ij})^{A_{ij}^t}(1-\chi_{ij})^{1-A_{ij}^t}}{\alpha\delta_{A_{ij}^t,A_{ij}^{t-1}} + (1-\alpha_{ij})(\chi_{ij})^{A_{ij}^t}(1-\chi_{ij})^{1-A_{ij}^t}} = 0 \\ \sum_{t=1}^{T}\dfrac{\delta_{A_{ij}^t,1} - \delta_{A_{ij}^t,0}}{\alpha\delta_{A_{ij}^t,A_{ij}^{t-1}} + (1-\alpha_{ij})(\chi_{ij})^{A_{ij}^t}(1-\chi_{ij})^{1-A_{ij}^t}} = 0. \end{cases} \tag{1.64}$$

Given the solution(s) of Eq. 1.64 in $(0,1)\times(0,1)$, we check what is the maximum among the possible candidates and the values on the bounds. Hence, we obtain the maximum likelihood estimators for α_{ij} and χ_{ij}.

When applying the DAR(p) model to a temporal network considered as independent sequences of links, we neglect the network as a whole. Another possibility

is considering $\alpha_{ij} = \alpha$ and $\chi_{ij} = \chi \ \forall i, j$ in Equation (1.60) with generic p-th order Markov dependence, as in (1.59). We refer to this model as DAR(p) network model. It has been studied by [Williams et al., 2019] with $\beta_1 = ... = \beta_p = 1/p$ to understand how memory can affect the properties of dynamical processes taking place over temporal networks, in particular to relate the p-th order of the Markovian behavior of link persistence with the dynamics of epidemic spreading processes.

Interestingly, we can analytically prove the ensemble equivalence between the DAR(p) network model and the p-th order Markovian generalization of TERG (1.58). Hence, this supports theoretically the equivalence of describing the network metric of link stability, see (1.57), with the mechanism of copying from the past of the discrete autoregressive models.

1.3.1.2 Ensemble equivalence between DAR(p) and TERG

The DAR(p) network model is equivalent to the $p-th$ order Markovian generalization of TERG, see (1.58), because of the existence of one-to-one mapping of parameters.

DAR(p) network model

Assume to observe a temporal network with N nodes and described by a time series of adjacency matrices $\boldsymbol{A}^t = \{A^t_{ij}\}_{i,j=1,...,N}$ with $A^t_{ij} \in \{0,1\}$ at $t = 0,1,...,T$. The transition probability of the network snapshot at time t given the information on the past p network snapshots is for the DAR(p) network model

$$\begin{aligned}&\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},...,\boldsymbol{A}^{t-p},\alpha,\chi) = \\ &= \prod_{i,j>i}\left[\alpha\left(\sum_{l=1}^{p}\beta_l\delta_{A^t_{ij},A^{t-l}_{ij}}\right) + (1-\alpha)\,(\chi)^{A^t_{ij}}(1-\chi)^{1-A^t_{ij}}\right],\end{aligned} \tag{1.65}$$

with $\alpha, \chi \in [0,1]$, $\beta_l \in [0,1]$ and $\sum_{l=1}^{p}\beta_l = 1$. Let us notice that the DAR(p) network model is described by $p+1$ parameters.

TERG for link stability

The metric of link stability (1.57) can be generalized to any lag order τ as

$$\Psi^{\tau}_{persistence}(\boldsymbol{A}^t,\boldsymbol{A}^{t-\tau}) = \sum_{i,j>i}\left[A^t_{ij}A^{t-\tau}_{ij} + (1-A^t_{ij})(1-A^{t-\tau}_{ij})\right]. \tag{1.66}$$

By considering link density and link stability up to order p as metrics to preserve on average over the dynamic graph ensemble, the TERG model for link stability with p-lags is described by the following transition probability

$$\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},...,\boldsymbol{A}^{t-p},\theta,\boldsymbol{\mu}) = \\ = \frac{1}{Z_{ij}}\prod_{i,j>i} e^{\theta A^t_{ij}+\mu_1\left[A^t_{ij}A^{t-1}_{ij}+(1-A^t_{ij})(1-A^{t-1}_{ij})\right]+...+\mu_p\left[A^t_{ij}A^{t-p}_{ij}+(1-A^t_{ij})(1-A^{t-p}_{ij})\right]}, \quad (1.67)$$

with $\boldsymbol{\mu} \equiv \{\mu_1,...,\mu_p\}$. By computing the partition function, we obtain the following explicit formula for the transition probability,

$$\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},...,\boldsymbol{A}^{t-p},\theta,\boldsymbol{\mu}) = \prod_{ij}\frac{e^{A^t_{ij}(\theta+\sum_{l=1}^p \mu_l(2A^{t-l}_{ij}-1))}}{1+e^{\theta+\sum_{l=1}^p \mu_l(2A^{t-l}_{ij}-1)}}. \quad (1.68)$$

Let us notice that the TERG model for link stability with p lags is described by $p+1$ parameters.

Ensemble equivalence

It is trivial to show that we can map the model (1.58) to the model (1.65) with $p=1$ by means of the following transformation of the parameters

$$\theta = \frac{1}{2}\log\left(\frac{\frac{\chi}{1-\chi}+\alpha}{\frac{1}{\chi}-(1-\alpha)}\right), \quad (1.69)$$

$$\mu = \frac{1}{2}\log\left(1+\frac{\alpha}{(1-\alpha)^2\chi(1-\chi)}\right). \quad (1.70)$$

In the general case of p lags, we can map the model (1.68) to DAR(p) (1.65) as long as conditional probabilities do not fall on the bounds 0 or 1 and there exists the inverse of a given matrix M of ones and minus ones, see below.

For instance, when $p=2$ we can map the parameters of (1.68) to the ones of (1.65) when there exists the solution of the following problem

$$\begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix}\begin{pmatrix}\mu_1 \\ \mu_2 \\ \theta\end{pmatrix} = \begin{pmatrix} \log\left(\frac{1}{(1-\alpha)\chi}-1\right) \\ \log\left(\frac{1}{\alpha(1-\beta_1)+(1-\alpha)\chi}-1\right) \\ \log\left(\frac{1}{\alpha+(1-\alpha)\chi}-1\right)\end{pmatrix}. \quad (1.71)$$

For general p, the problem reads as

$$M\cdot\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \vdots\\ \mu_p\\ \theta\end{pmatrix}\equiv\begin{pmatrix}1 & 1 & ... & 1 & 1 & -1\\ 1 & 1 & ... & 1 & -1 & -1\\ 1 & 1 & ... & -1 & -1 & -1\\ 1 & ... & ... & ... & ... & -1\\ 1 & -1 & ... & -1 & -1 & -1\\ -1 & -1 & ... & -1 & -1 & -1\end{pmatrix}\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \vdots\\ \mu_p\\ \theta\end{pmatrix}=\begin{pmatrix}\log\left(\frac{1}{(1-\alpha)\chi}-1\right)\\ \vdots\\ \vdots\\ \vdots\\ \log\left(\frac{1}{\alpha+(1-\alpha)\chi}-1\right)\end{pmatrix}. \tag{1.72}$$

In practice, we can associate the probability $\mathbb{P}(A_{ij}^t=1|A_{ij}^{t-1},...,A_{ij}^{t-p})$ of the model (1.68) to the ones of the model (1.65) for $p+1$ different conditions[4], *i.e.*

$$\begin{cases}\frac{1}{1+e^{-\theta-\mu_1-...-\mu_p}}=\alpha+(1-\alpha)\chi & \text{if } A_{ij}^{t-1}=1,A_{ij}^{t-2}=1,...,A_{ij}^{t-p}=1;\\ ... & ...\\ \frac{1}{1+e^{-\theta+\mu_1+...+\mu_p}}=(1-\alpha)\chi & \text{if } A_{ij}^{t-1}=0,A_{ij}^{t-2}=0,...,A_{ij}^{t-p}=0.\end{cases} \tag{1.73}$$

Finally, (1.72) is obtained by noticing that

$$\frac{1}{1+e^x}=a\rightarrow x=\log(\frac{1}{a}-1)$$

as long as $a\neq 0,1$.

Hence, the two models can be mapped one to each other as long as there exists the inverse of the matrix M. Let us define $n=p+1$ and M_n as the matrix M associated with the problem (1.72) for given p.

Proposition 1.1. *Given the determinant of the matrix M_{n-1}, then the determinant of the matrix M_n is*

$$\det(M_n)=\begin{cases}2\det(M_{n-1}) & \textit{if } n \textit{ is even},\\ (-2)\det(M_{n-1}) & \textit{if } n \textit{ is odd}.\end{cases} \tag{1.74}$$

By means of the *minor expansion formula* (by using the minors associated with the elements of the first row), the determinant of M_n can be computed as

[4] Among the $2p$ different possibilities for $A_{ij}^{t-1},...,A_{ij}^{t-p}$, we select the $p+1$ conditions resulting in the matrix M. The other ones are fixed by using considerations on the probability distribution.

$$\det(M_n) = (+1)1\begin{vmatrix} 1 & \dots & 1 & -1 & -1 \\ 1 & \dots & -1 & -1 & -1 \\ \dots & \dots & \dots & \dots(-1) & -1 \\ -1 & \dots & -1 & -1 & -1 \\ -1 & \dots & -1 & -1 & -1 \end{vmatrix} + (-1)1\begin{vmatrix} 1 & \dots & 1 & -1 & -1 \\ 1 & \dots & -1 & -1 & -1 \\ 1 & \dots & \dots & \dots(-1) & -1 \\ 1 & \dots & -1 & -1 & -1 \\ -1 & \dots & -1 & -1 & -1 \end{vmatrix} +$$
$$+ \dots + (-1)^n(1)\begin{vmatrix} 1 & 1 & \dots & 1 & -1 \\ 1 & 1 & \dots & -1 & -1 \\ 1 & \dots & \dots & \dots & -1 \\ 1 & -1 & \dots & -1 & -1 \\ -1 & -1 & \dots & -1 & -1 \end{vmatrix} + (-1)^{n+1}(-1)\begin{vmatrix} 1 & 1 & \dots & 1 & -1 \\ 1 & 1 & \dots & -1 & -1 \\ 1 & \dots & \dots & \dots & -1 \\ 1 & -1 & \dots & -1 & -1 \\ -1 & -1 & \dots & -1 & -1 \end{vmatrix}, \tag{1.75}$$

where the first $n-2$ minors of the sum in (1.75) are zero because the last two columns of each $n-1 \times n-1$ matrix are the same (two $n-1 \times 1$ vectors of -1), thus resulting in a null determinant because of linear dependence of two columns. Hence, (1.75) becomes

$$\det(M_n) = (-1)^n 2\begin{vmatrix} 1 & 1 & \dots & 1 & -1 \\ 1 & 1 & \dots & -1 & -1 \\ 1 & \dots & \dots & \dots & -1 \\ 1 & -1 & \dots & -1 & -1 \\ -1 & -1 & \dots & -1 & -1 \end{vmatrix} = (-1)^n 2\det(M_{n-1}), \tag{1.76}$$

where we notice that the last two minors of (1.75) are equal to each other and correspond to the determinant of M_{n-1}.

Proposition 1.2. *There exists a solution of the problem (1.72) for any $p \in \mathbb{N}$ and this solution is unique.*

For $p = 1$, the solution can be explicitly computed as in (1.69) and (1.70). For $p = 2$ the problem (1.72) is equivalent to (1.71) and $\det(M_3) = 4$, thus there exists the inverse of the matrix M_3 and the solution is uniquely determined by solving the linear system of equations (1.71). Because of Proposition 1.1, the determinant of M_n is different from zero, *i.e.* $\det(M_n) = (-1)^{\sum_{l=4}^{n} l}(2^{n-3})\det(M_3) = (-1)^{\sum_{l=4}^{n} l}(2^{n-3})4$ $\forall n > 3$ ($\forall p > 2$), thus resulting in the existence of the inverse matrix of M_n. Hence, the solution of the problem (1.72) can be uniquely determined by solving the linear system of equations. This completes the proof.

1.3.2 Dynamic network models with latent variables

The key aspects in the description of time-varying parameter models are how we specify the latent process and the mapping between the latent variables and the link probability. However, for any possible choice, the common characteristic is the conditional independence for the probability of the network snapshot from previous

observations, see (1.54). In other words, at each time snapshot the network is generated by re-sampling links according to the unobserved dynamics of latent variables. This approach aims to capture the time-varying patterns of the network structure, *e.g.* how the community structure changes over time, thus resulting in the creation of new links and the disruption of old ones, according to the new node memberships.

The scientific literature on dynamic network models with latent variables is very rich and models are usually grouped in latent space models and stochastic block models, see [Kim et al., 2018] for a review.

A milestone work is represented by [Sarkar and Moore, 2006] that generalized the latent space model introduced in [Hoff et al., 2002] to dynamic networks. The dynamics of the network structure is modelled through random effects in a latent space. In particular, it assumes, as in the static case, that the link probability depends on the distance between latent positions of nodes, but latent positions move over time in the latent space with standard Markovian assumptions. That is, the link probability between node i and node j is proportional to the logistic function of the distance d_{ij}^t in the latent space, *i.e.*

$$\mathbb{P}(A_{ij}^t = 1 | d_{ij}^t, r_{ij}) \propto \text{logit}^{-1}(r_{ij} - d_{ij}^t), \tag{1.77}$$

if the latent coordinates of nodes are within the interaction distance r_{ij}. Outside this range, there is a constant link probability. Then, the latent coordinates of nodes evolve independently in the latent Euclidean space, subject to Gaussian perturbations with zero mean and given variance. Hence, within this model, we keep track of how much it is likely that one node connects to other nodes (depending on the distance from them in the latent space) and the probability of becoming an hub (depending on the number of neighbors within the interaction distance).

Further generalizations of latent position models have been introduced to capture more general features. For instance, [Sarkar et al., 2007] extends the model to be applicable for dynamic bipartite networks. [Sewell and Chen, 2015] propose a latent position model for directed networks, by describing the dynamics of latent positions with a Markov process, similarly to [Sarkar and Moore, 2006], but introducing also additional features to capture the effect of popularity and activity of nodes. Then, [Friel et al., 2016] develop a statistical model for bipartite dynamic networks by extending [Sarkar and Moore, 2006, Sarkar et al., 2007]. They describe the evolution of links through three Markov processes: one for the latent positions, one for the links, and one for the parameters.[5] In particular, the Markovian behavior for the links tend to capture the global persistence, by distinguishing persistence of links from no-links. Thus, the link probability within the model is

$$\begin{aligned} &\mathbb{P}(A_{ij}^t = 1 | A_{ij}^{t-1}, d_{ij}^t, \tilde{\alpha}_{link}^t, \tilde{\alpha}_{no-link}^t) = \\ &= \text{logit}^{-1}(\tilde{\alpha}_{link}^t A_{ij}^{t-1} + \tilde{\alpha}_{no-link}^t (1 - A_{ij}^{t-1}) + d_{ij}^t), \end{aligned} \tag{1.78}$$

[5] Note here that the conditional independence from the previous network observations does not hold anymore, because of the Markovian behavior of links. To the best of our knowledge this is the closest statistical approach to the problem investigated in the next chapter. We discuss better this point below.

where d_{ij}^t is the distance between latent positions of i and j, $\tilde{\alpha}_{link}^t$ is the link persistence parameter, and $\tilde{\alpha}_{no-link}^t$ is the no-link persistence parameter. Both persistence parameters evolve themselves in time by following a Markov process. Using the model, the authors analyze the dynamic evolution of the leading Irish companies and their directors from 2003 to 2013. Mainly focused on understanding the (global) persistence of links and the heterogeneity in the latent positions, the analysis reveals an increasing level of interlocking board behavior before and during the financial crisis, and stabilization thereafter.

Within this stream of literature, some approaches move from the description of latent positions in order to capture the evolution of some network effects, such as reciprocity and transitivity. This is, *e.g.* , the case of [Durante and Dunson, 2014]. They proposed a dynamic extension of the model of [Hoff, 2005], by describing the dynamics of some latent factors capturing the second and third order metrics via Gaussian processes. They introduced also an efficient MCMC algorithm for Bayesian inference to learn model parameters, see [Durante et al., 2016]. Furthermore, the latent factor models [Tsay, 2005] have been applied to the case of dynamic networks in several directions. [Hoff, 2011, Hoff et al., 2011] develops a general modeling framework for array data via reduced-rank decompositions, in which the adjacency matrix describing the network snapshot is coupled with the latent dynamics of some factors. Depending on the number of factors, this approach allows to reduce the dimensionality of the problem. In this works, Hoff et al. aim to describe the dynamics of the hierarchical structure of a networked system. [Heaukulani and Ghahramani, 2013] have studied a similar approach, but introducing a dependence of the latent factors from the previous network realization, thus describing how observed relationships from the past affect future unobserved structure in social networks. Finally, [Bräuning and Koopman, 2016] have combined the latent factor model with the stochastic block model, where the unobserved node membership determines which factor the node is following during the evolution. This model allows to capture cross-sectional dependencies in network data.

Close to the approach with latent factor models, [Giraitis et al., 2016] propose a methodology for the econometric modelling of dynamic networks which describes the link dynamics with a Tobit-like model, where latent variables are mapped to the entries of the adjacency matrix by means of the unit step function. Then, the evolution of the latent variables is described by VARX(1)-like model, *i.e.* a vector autoregressive process which includes also exogenous variables [Pesaran et al., 2004, Pesaran, 2015]. In order to reduce the problem dimensionality, Giraitis et al. restrict the space of parameters in a similar fashion of the factor decomposition. However, differently from latent space models, the authors do not specify further the latent dynamics, but they exploit kernel-based local maximum likelihood estimator of the time-varying parameters. This approach allows to capture structural changes in the dynamics as well as including exogenous factor driving the evolution of the network. For example, the application to the daily overnight money market network in the UK 2003-2012 shows that the model captures the several structural breaks arising from changes in the monetary policy.

Recently, [Lee et al., 2017] proposed a generalization of ERG which goes in the direction of latent space models. While TERG describes the Markovian behavior of dynamic network metrics, *i.e.* metrics involving consecutive network snapshots such as link stability, Lee et al. focus on static network metrics, such as link density, reciprocity, triangles, and so on, but introducing a latent dynamics for the Lagrange multipliers of ERG. Thus, the latent variables describe how the probability of observing such metrics changes over time. In mathematical terms, Lee et al. describe the conditional probability of the sequence of graph snapshots $\boldsymbol{\mathcal{A}} \equiv \{\boldsymbol{A}^1, ..., \boldsymbol{A}^T\}$ as

$$\mathbb{P}(\boldsymbol{\mathcal{A}}|\boldsymbol{\Theta}) = \prod_{t=1}^{T} \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{\theta}^t) = \prod_{t=1}^{T} \frac{e^{(\boldsymbol{\theta}^t)'\boldsymbol{\Psi}(\boldsymbol{A}^t)}}{Z^t(\boldsymbol{\theta}^t)}, \tag{1.79}$$

where $\boldsymbol{\Theta} \equiv \{\boldsymbol{\theta}^t\}^{t=1,...,T}$ is the set of latent variables, $\boldsymbol{\Psi} : \mathbb{R}^{N\times N} \to \mathbb{R}^n$ maps each network snapshot to n static network metrics, and $Z^t(\boldsymbol{\theta}^t)$ is the partition function at time t. The approach (1.79) is named Varying-Coefficient Exponential Random Graph (VCERG) models. Hence, the authors propose to describe the latent dynamics of $\boldsymbol{\theta}^t$ as a linear combination of basis functions, *i.e.* basis splines (b-splines) [Eilers and Marx, 1996], and provide a maximum pseudo-likelihood method for model estimation, see [Lee et al., 2017] for further details.

Let us notice that a dynamic generalization of the fitness model (1.20) can be obtained by using the degree sequence at time t as network metrics $\boldsymbol{\Psi}(\boldsymbol{A}^t)$ in (1.79). This is presented in the next chapter and, at the best of our knowledge, it is the first time the fitness model is thus generalized within the class of latent space models. A similar approach has been recently introduced in [Jung et al., 2018] where the authors study the dynamical evolution of a network driven by two mechanisms: fitness and popularity of nodes. However, the node fitness is considered as constant in time and no latent dynamics describes its evolution.

Networks with time-varying community structure are described by stochastic block models with node membership changing over time. The first generalization in this direction has been introduced by [Yang et al., 2011] which propose a model capturing the evolution of communities by explicitly modeling the transition of community memberships for individual nodes in the network. In particular, if k is the number of communities, they use a $k \times k$ transition matrix to model the probability for node i with label (membership) g_i^{t-1} at the previous snapshot to change community (or remaining in the same) at time t. Thus, the authors describe a Markovian behavior for the evolution of communities. Then, they propose two estimation methods, *i.e.* a point estimation based on a EM algorithm and an interval estimation based on Bayesian inference approach, and show how the Bayesian approach outperforms the proposed point estimation. However, the EM algorithm proposed by [Yang et al., 2011] is based on some approximation for the posterior distribution of node memberships. In the next subsection, we review [Ghasemian et al., 2016] which introduce the optimal EM algorithm for the inference of SBM with a Markovian behavior for node memberships.

The dynamic extension of SBM by [Yang et al., 2011] does not account for time-varying affinity matrix, *i.e.* the link probabilities between groups and within

the group itself are constant in time. Both time-varying communities and connectivity parameters are described in two different works by [Xu and Hero, 2014, Matias and Miele, 2017]. In the first one, Xu and Hero do not specify any latent dynamics for the node memberships but infer them by means of label-switching methods. However, they describe a Markovian behavior for the evolution of the affinity matrix $\boldsymbol{p}^t \equiv \{p^t_{ab}\}_{a,b=1,\dots,k}$, in particular the generic entries evolve according to a multivariate Gaussian random walk

$$\text{logit}(\text{vec}(\boldsymbol{p}^t)) = \text{logit}(\text{vec}(\boldsymbol{p}^{t-1})) + \boldsymbol{s}^t \tag{1.80}$$

where $\text{vec}(\boldsymbol{p}^t)$ represents the vectorization of the affinity matrix obtained by stacking the columns of the matrix on top of one another, and $\boldsymbol{s}^t$ is a random vector of zero-mean Gaussian entries with generic covariance matrix. The model inference is based on an iterative estimation procedure alternating two steps: (i) a label-switching method used to explore the space of node group configurations, thus maximizing the posterior probability of node memberships according to static SBM (by conditioning on the estimated affinity matrix for the considered network snapshot); (ii) then, the extended Kalman filter is used to estimate (1.80) and the other model parameters. In particular, this step allows to take into account the presence of some observation noise, if any. This is the case of SBM where the observed link density between groups is a proxy for the affinity matrix, whose estimated entries are approximately Gaussian with mean equal to observed link densities, in the asymptotic limit $N \to \infty$, see [Xu and Hero, 2014]. However, this estimation method strictly relies on persistence of node memberships because of the efficiency of the label-switching method for community inference.

To improve the performance of the label switching method across the different time steps, [Matias and Miele, 2017] focus on detecting communities characterized by a stable within-group connectivity behavior, thus considering the corresponding elements of the affinity matrix as fixed. They further study the problem of identifiability for the considered model and propose a novel variational expectation-maximization method to overcome the issues observed in [Yang et al., 2011].

As for the static case, there exist other generalizations of dynamic SBM to capture, *e.g.* , the multiple roles of nodes with mixed-memberships. For example, [Xing et al., 2010, Ho et al., 2011] propose dynamic extensions of the static mixed membership stochastic blockmodel using a state space model for the time-varying parameters, both for the mixed membership vector of a node and the connectivity behavior. Here, each node enters with a dynamic mixed membership vector that allows multiple interactions as well as linkage behavior changing in time.

Finally, [Xu, 2015] extends further [Xu and Hero, 2014] by introducing a novel version of dynamic SBM, namely *Stochastic Transition Block Model* (STBM), which takes into account for the Markovian behavior of link persistence in order to capture how past links (or no-links) influence directly the future link probability, but preserving the description with the stochastic block model of the community structure of a network. However, the role of the Markovian behavior of link persistence in the inference of communities for the stochastic block model has not be

assessed in this work. Not surprisingly, it can be shown that link persistence makes the inference of communities harder. In Chapter 4, we assess analytically this problem, thus pointing out how the persistence of both links and groups gives rise to a new phenomenon in the inference of SBM, namely *time-lagged inference*: the identification of past communities rather than present ones.

1.3.3 Community detection in dynamic networks

While a number of models of dynamic SBM and methods to estimate them have been introduced in the literature, the theoretical analysis and ultimate limits on when communities can be detected in dynamic networks have been studied in [Ghasemian et al., 2016]. Here, we briefly review this work and the key aspects of the detectability condition, which are functional to the research presented in Chapter 4.

By extending the BP inference method introduced for static SBM by [Decelle et al., 2011a], [Ghasemian et al., 2016] propose to study a special case of dynamic SBM introduced in [Yang et al., 2011] which is analytically tractable and, at the same time, is able to capture the dynamic (Markovian) behavior of time-varying community memberships. They study a generalization of the static stochastic block model (1.29) where the dynamic of each node label g_i^t is an independent Markov process with transition probability

$$\mathbb{P}(g_i^t|g_i^{t-1},\boldsymbol{q},\eta) = \eta\,\delta_{g_i^t,g_i^{t-1}} + (1-\eta)q_{g_i^t}, \tag{1.81}$$

where η is the persistence parameter and q_a is the marginal probability of belonging to group a, thus $\boldsymbol{q} \equiv \{q_a\}_{a=1,\dots,k}$ describes the marginal multinomial distribution for k groups.[6] As in the static case, we have a network with N nodes and each network snapshots is generated according to a static SBM given the node labels $\boldsymbol{g}^t \equiv \{g_i^t\}_{i=1,\dots,N}$ at the current time t, *i.e.*

$$\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{g}^t,\boldsymbol{p}) = \prod_{i,j>i} p_{g_i^t g_j^t}^{A_{ij}^t}(1-p_{g_i^t g_j^t})^{1-A_{ij}^t}, \tag{1.82}$$

where $\boldsymbol{p} = \{p_{ab}\}_{a,b=1,\dots,k}$ is the constant affinity matrix describing the link probability between two groups or within the same group. The likelihood of the sequence of graphs $\boldsymbol{\mathcal{A}} = \{\boldsymbol{A}^1,\dots,\boldsymbol{A}^T\}$ and labels $\vec{\boldsymbol{g}} \equiv \{\boldsymbol{g}^0,\boldsymbol{g}^1,\dots,g^T\}$ is then

$$\begin{aligned}\mathbb{P}(\boldsymbol{\mathcal{A}},\vec{\boldsymbol{g}}|\boldsymbol{p},\boldsymbol{q},\eta) = &\prod_{t=1}^{T}\left[\prod_{i=1}^{N}\left(\eta\,\delta_{g_i^t,g_i^{t-1}} + (1-\eta)q_{g_i^t}\right)\right]\mathbb{P}(\boldsymbol{g}^0|\boldsymbol{q})\times \\ &\times\prod_{t=1}^{T}\left(\prod_{i,j>i} p_{g_i^t g_j^t}^{A_{ij}^t}(1-p_{g_i^t g_j^t})^{1-A_{ij}^t}\right),\end{aligned} \tag{1.83}$$

[6] The process (1.81) is in effect a DAR(1) model (1.60) for categorical variables.

where we assume that labels $\boldsymbol{g}^0$ are sampled according to the marginal distribution.

For the same reasons explained before for the static case, let us focus on the difficult case of inference characterized by uniform prior, *i.e.* $q_a = 1/k \; \forall a = 1,...,k$, and affinity matrix $p_{ab} = c_{ab}/N$ describing an assortative planted partition model, *i.e.* $c_{aa} = c_{in} \; \forall a = 1,...,k$ and $c_{ab} = c_{out} \leqslant c_{in}$ when $a \neq b$. We are interested in the regime of sparse networks, *i.e.* $p_{ab} \sim O(1/N)$ or, equivalently, $c_{ab} \sim O(1)$.

Hence, the problem of community detection consists in inferring the node memberships $\vec{\boldsymbol{g}} = \{g_i^t\}_{i=1,...,N}^{t=1,...,T}$, given the model parameters $\boldsymbol{\pi} \equiv \{\boldsymbol{p}, \boldsymbol{q}, \eta\}$.[7] In the dynamic case, we can expect that the detectability condition depends strictly on the group persistence η. In fact, when communities change slowly in time and node labels are persistent, we can exploit the information on many network snapshots, resulting in better inference of the community structure of the network. On the contrary, if node labels at successive steps are uncorrelated, we cannot do better than the static BP inference. We thus expect that the detectability threshold in dynamic networks interpolates between its static value for $\eta = 0$ and zero for $\eta = 1$.

Similarly to the static case, when we aim to obtain the (*posterior*) *marginal* $\nu_i(g_i^t) = \sum_{\vec{\boldsymbol{g}} \setminus g_i^t} \mathbb{P}(\vec{\boldsymbol{g}} | \boldsymbol{\mathcal{A}}, \boldsymbol{\pi})$ in terms of 'messages', we need to consider the 'interactions' coming from the neighbors at time t, but also the 'interactions' of node i with both its past and future copies in time. The 'spatial interaction' between two neighbors is mediated by the affinity matrix describing the link (or no-link) probability, whereas the 'temporal interaction' between two copies of the same node at successive snapshots is mediated by the transition probability (1.81).

Finally, when we consider the case of sparse networks, short loops in the spatiotemporal network of 'interactions' are rare too, implying that we can adopt the treelike approximation, similarly to the static case.

Hence, the BP equations (1.40) can be generalized as

$$\begin{aligned} \psi_{g_i^t}^{i(t)\to j(t)} = & \frac{1}{Z^{i(t)\to j(t)}} \left(\sum_a \tau_{g_i^t a} \psi_a^{i(t-1)\to i(t)} \right) \left(\sum_a \tau_{a g_i^t} \psi_a^{i(t)\to i(t+1)} \right) \times \\ & \times \prod_{k\neq i,j} \left(\sum_{g_k^t} c_{g_i^t g_k^t}^{A_{ik}^t} \left(1 - \frac{c_{g_i^t g_k^t}}{N} \right)^{1-A_{ik}^t} \psi_{g_k^t}^{k(t)\to i(t)} \right), \end{aligned} \tag{1.84}$$

$$\begin{aligned} \psi_{g_i^t}^{i(t)\to i(t+1)} = & \frac{1}{Z^{i(t)\to i(t+1)}} \left(\sum_a \tau_{g_i^t a} \psi_a^{i(t-1)\to i(t)} \right) \times \\ & \times \prod_{k\neq i} \left(\sum_{g_k^t} c_{g_i^t g_k^t}^{A_{ik}^t} \left(1 - \frac{c_{g_i^t g_k^t}}{N} \right)^{1-A_{ik}^t} \psi_{g_k^t}^{k(t)\to i(t)} \right), \end{aligned} \tag{1.85}$$

[7] This corresponds to the dynamic generalization of the E-step of the MAP estimation of static SBM. The M-step, *i.e.* solving the Nishimori conditions (1.36-1.38) together with the one for η, can be obtained by simple computations, similarly to the static case.

$$\psi_{g_i^t}^{i(t)\to i(t-1)} = \frac{1}{Z^{i(t)\to i(t-1)}} \left(\sum_a \tau_{a g_i^t} \psi_a^{i(t)\to i(t+1)} \right) \times \\ \times \prod_{k\neq i} \left(\sum_{g_k^t} c_{g_i^t g_k^t}^{A_{ik}^t} \left(1 - \frac{c_{g_i^t g_k^t}}{N} \right)^{1-A_{ik}^t} \psi_{g_k^t}^{k(t)\to i(t)} \right), \tag{1.86}$$

where $Z^{i(t)\to j(t)}$ and $Z^{i(t)\to i(t\pm 1)}$ are normalization factors, and τ is the transition matrix associated with (1.81), *i.e.* $\tau = \eta \mathbb{1} + (1-\eta)\frac{\mathbf{1}}{k}$ with $\mathbb{1}$ the identity matrix and $\mathbf{1}$ the $k \times k$ matrix of ones.

Then, once the messages reach the fixed point $\{\bar{\psi}_{g_i^t}^{i(t)\to j(t)}, \bar{\psi}_{g_i^t}^{i(t)\to i(t+1)}, \bar{\psi}_{g_i^t}^{i(t)\to i(t-1)}\}$, the posterior marginals can be computed as

$$\psi_{g_i^t}^{i(t)} = \frac{1}{Z^{i(t)}} \left(\sum_a \tau_{g_i^t a} \bar{\psi}_a^{i(t-1)\to i(t)} \right) \left(\sum_a \tau_{a g_i^t} \bar{\psi}_a^{i(t)\to i(t+1)} \right) \times \\ \times \prod_{k\neq i} \left(\sum_{g_k^t} c_{g_i^t g_k^t}^{A_{ik}^t} \left(1 - \frac{c_{g_i^t g_k^t}}{N} \right)^{1-A_{ik}^t} \bar{\psi}_{g_k^t}^{k(t)\to i(t)} \right), \tag{1.87}$$

with $Z^{i(t)}$ the normalization factor.[8] Hence, we have $\nu_i(g_i^t) \equiv \psi_{g_i^t}^{i(t)}$.

In Equations (1.84-1.86), we have $O(TN^2)$ messages. However, as in the static case, we can approximate the 'interactions' mediated by no-links with an external field, such that we only need to keep track of $O(TN)$ messages in the sparse regime. Thus, the resulting estimation method scales linearly with the number of nodes. For further details and for the approximated BP equations see [Ghasemian et al., 2016].

As in [Decelle et al., 2011a], the BP equations have a trivial fixed point $\psi_{g_i^t}^{i(t)\to j(t)} = \psi_{g_i^t}^{i(t)\to i(t\pm 1)} = 1/k$ for all i, j, t, thus resulting in uniform posterior marginals. This is equivalent to say that the inference algorithm performs no better than chance. When the trivial fixed point is stable, communities are not detectable. However, there exists a detectability threshold which determines a phase transition for the inference problem. Above this transition, the trivial fixed point becomes unstable, and BP converges to a nontrivial fixed point. When the network is generated according to dynamic SBM and we know the static parameters, the nontrivial fixed point represents a method to reconstruct the communities better than chance. In particular, if we assign to each node its most likely label at each time, *i.e.* $\hat{g}_i^t = \arg\max_{g_i^t} \nu_i(g_i^t)$, this assignment maximizes the fraction of correct labels and we can quantify this by means of the overlap (1.27).

The detectability threshold for the dynamic case can be found with similar considerations adopted for the static case, or by noticing that the BP inference can be thought as a reconstruction problem of labels on a spatiotemporal tree [Janson et al., 2004], see [Ghasemian et al., 2016] for further details. Then, the cri-

[8] When $t = 1$ we can assume the messages coming from $t = 0$ as the ones of the static SBM. When $t = T$, the terms that should come from $T + 1$ need to be removed.

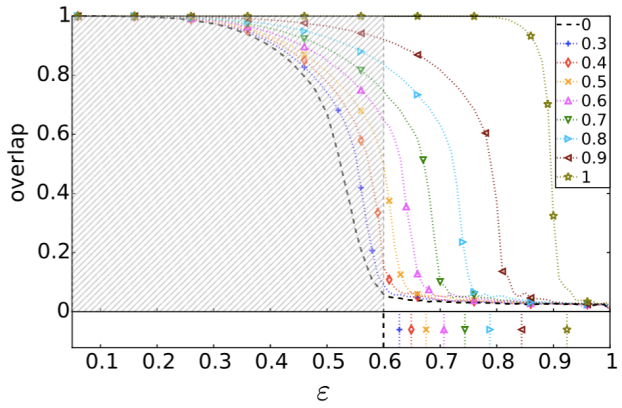


Fig. 1.3 Overlap (1.27) between the original assignment and its most likely estimate obtained with the dynamic BP algorithm, as a function of the assortativity parameter $\varepsilon = c_{out}/c_{in}$. Each graph is generated according to the dynamic SBM (1.81-1.82) for different value of the group persistence parameter η (given in the legend). For each η, the critical value of ε (with finite size corrections because of finite T) is shown as a vertical line in the lower panel, and the hatched area shows the region of detectability for static networks, see [Decelle et al., 2011a]. Graphs are generated using $N = 512$ nodes, 2 groups of the same size, $T = 40$, and average degree $c = 16$. The plot is taken from [Ghasemian et al., 2016].

terion for detectability in the asymptotic limit $N, T \to \infty$ is thus

$$|c_{in} - c_{out}| > k\sqrt{c\frac{1-\eta^2}{1+\eta^2}}, \tag{1.88}$$

which ranges from the static threshold $k\sqrt{c}$, with c the average degree, when $\eta = 0$, see (1.52), to zero when $\eta = 1$.

The phase diagram associated with the detectability problem for this specific stochastic block model is shown in Figure 1.3. Notice that by increasing group persistence we outperform the single snapshot inference, *i.e.* the static BP algorithm, by accounting for both spatial and temporal information. Finally, notice that the detectability threshold does not become zero for $\eta = 1$ because of finite T. Finite size corrections can be computed, *e.g.* the detectability threshold is $|c_{in} - c_{out}| = k\sqrt{c/T}$ for $\eta = 1$, see [Ghasemian et al., 2016].

Conclusions

In this chapter, we have reviewed several statistical network models and we have described the problem of community detection in stochastic block models, for both static and dynamic networks.

In particular, we have shown how the dynamic generalizations can take two directions: (i) static parameter models describing temporal network metrics, *i.e.* measurements involving consecutive network snapshots such as link persistence, and (ii) time-varying parameter models capturing the time-varying patterns of the network structure, *i.e.* the description of the latent dynamics which determines how some topological features, such as communities, change over time.

In the next chapters, we combine the two approaches and we show how disentangling the two associated patterns in dynamic networks. Moreover, we show how the introduction of link direction in stochastic block models modifies the problem of community detection. More specifically,

1. in Chapter 2, we introduce a dynamic generalization of the fitness model, then we combine it with the Markovian behavior of link persistence. We point out how the two dynamic patterns may coexist, how to infer them, and how link persistence is associated with the preferential linkage mechanism in the interbank market (*i.e.* the electronic Market of Interbank Deposit e-MID);
2. in Chapter 3, we show that link direction affects crucially the problem of community detectability in static SBM, thus resulting in a detectability condition depending on the asymmetry of the network as well as the degree of assortativity. Then, we study the problem of identifying macroscopic structures in networks, pointing out what are the difficult cases of inference for the directed SBM;
3. in Chapter 4, we combine the dynamic SBM described here, with the Markovian behavior of link persistence. The two dynamic components lead to time-lagged inference: the identification of past communities rather than present ones. By considering maximum likelihood inference from single snapshot observations of the network, we show, both analytically and numerically, that link persistence makes the inference of communities harder, decreasing the detectability threshold, while community persistence tends to make it easier.

Chapter 2
Link persistence in dynamic fitness model

Abstract In this chapter, we introduce a dynamic network model where two mechanisms control the probability of a link between two nodes: (i) the existence or absence of this link in the past, and (ii) node-specific latent variables (dynamic fitnesses) describing the propensity of each node to create links. In particular, we describe a discrete autoregressive dynamics for the mechanism of copying from the past, thus combining it with a dynamic generalization of the fitness model where the (latent) node fitness evolves in time by following an autoregressive process. Then, we propose an Expectation-Maximization algorithm for model estimation and inference of the latent variables. The estimated parameters and fitnesses can be used to forecast the presence of a link in the future. Finally, we apply our methodology to the e-MID interbank network for which the two linkage mechanisms are associated with two different trading behaviors in the process of network formation, namely preferential trading and trading driven by node-specific characteristics. The empirical results allow recognizing preferential lending in the interbank market and indicate how a method that does not account for time-varying network topologies tends to overestimate preferential linkage.
Almost all results in this chapter previously appeared in [Mazzarisi et al., 2019a].

Introduction

When we model dynamic networks, there are several mechanisms that can lead to link formation and destruction. The presence of a link may depend on current node properties of the system but also on previous network states. Consider as an example a trading network such as the interbank network studied in the empirical application below. The probability of a link between two nodes representing a transaction between the corresponding entities depends generically on the current supply and demand of the two entities, as well as on the existence (or absence thereof) of a link in the past between the two entities. The former driver is associated with node-specific properties (supply and demand) which evolve in time with their own

dynamics, possibly dependent also on the network state. The latter driver, instead, is associated with a link specific property, namely the persistence of links which describes the tendency to interact with whom we have interacted with in the past.

The objective of this chapter is to introduce a dynamic network model where both mechanisms are present and to propose a statistical estimation technique which allows disentangling the importance of the two mechanisms for each link in the network. The estimation method is based on an Expectation-Maximization scheme for Maximum A Posteriori (MAP) estimation. As we will show, the estimation of the model allows also forecasting the probability of the existence of a specific link in the future when the model parameters and the past network history are known.

More specifically, in our model we define a Markovian process on link dynamics combined with an autoregressive model for the latent variables governing the link probability. The Markovian behavior for each link is described in a similar fashion to the DAR(1) model (1.60), where the marginal is given now by a dynamic generalization of the fitness model (1.20). In fact, there is a latent variable in each node called the fitness of the node. The latent dynamics of the fitness is described by an autoregressive process AR(1). Thus at each time-step a link can be created — or not — either as a consequence of a copying process of the past link state or as a consequence of a random sampling whose probability depends on the current value of the latent variables of the two considered nodes. From the point of view of generative network models, link stability tends to capture preferential linkage mechanism between the nodes of the network while the fitness dynamics accounts for the evolving network topology. Clearly both mechanisms give rise to time correlation of the link state, even if its origin is quite different in the two cases. Being able to disentangle the role in link persistence due to explicit copy from the past or to fitness dynamics allows to identify genuine patterns of preferential linkage.

As a specific application, we study the interbank market, which is an important infrastructure of the financial system. Banks borrow and lend money in the interbank market to meet liquidity shortages or to allocate liquidity surpluses on a daily basis. The decision of whom to trade with is complex but certainly two aspects play an important role: first, the internal state (e.g. balance sheet, liquidity available or needed) of the bank and second the knowledge of the counterpart. Concerning this last aspect, all else being equal, a bank will prefer to trade with someone who was a counterpart in the past, since lending money requires some trust on the borrower's solvability. This behavior is known as preferential trading [Weisbuch et al., 2000], and has been documented in many empirical papers [Cocco et al., 2009]. Our model is able to assess the importance of preferential trading between two banks when the (possibly time-varying) internal states of the two banks are taken into account. In fact, we apply our methodology to the financial network of electronic Market of Interbank Deposit (e-MID) where the two linkage mechanisms, *i.e.* copying the link from the past or sampling it according to the latent dynamics, are associated with two different trading behaviors, i.e. random and preferential trading (quantified as in [Hatzopoulos et al., 2015]). Finally, it is important to stress that the same argument can be made for social networks where the copying mechanism

favors extant links due to a minor social cost of entertaining new relationships, see [Hellmann and Staudigl, 2014] for a review on the evolution of social networks.

Concerning the related literature, this work combines the two principal approaches in the description of dynamic networks, *i.e.* (i) description of the graph dynamics using (generalized) Markov chains on network observables, and (ii) description of the graph dynamics using models of latent variables whose dynamics determines the evolution of the network topology. In particular, from one side, we describe link persistence coming from the mechanism of copying from the past by modeling a Markov chain for link stability, *i.e.* the tendency of a link that does (or does not) exist at time $t-1$ to continue existing (or not existing) at time t, similarly to [Hanneke et al., 2010]. From the other side, we describe the stochastic dynamics of node-specific latent variables that we call fitnesses, with a similar aim of [Bräuning and Koopman, 2016, Lee et al., 2017]. The node fitness describes the tendency of a node in creating links and its evolution determines how the degree of the node changes in time. Recently, [Friel et al., 2016] study the persistence in company boards of directors of Irish companies by introducing a bipartite model which utilises two Latent Euclidean spaces for the two types of nodes, i.e. directors and boards, and two persistence parameters for conditioning future probabilities on both absent and present past links. To the best of our knowledge this is the closest statistical approach to the problem investigated in the present study. In particular, the work by Friel et al. differs in the Markovian behavior of links. In fact, the mechanism described by Friel et al. captures the average link persistence of the network, without distinguishing different persistent behaviors of links. In our approach, though, the local behavior captured by a link-specific persistence parameter is of fundamental importance when we aim to disentangle the two persistence patterns associated with both the node fitness and the links incident to the node. At difference with the aforementioned study, we are focusing on a model of dynamic networks with latent fitness variables, where past links are directly influencing the probability of future links, proposing an original Expectation-Maximisation procedure to learn all the parameters of the model, both latent variables and persistence parameters. Further, we test this model on a benchmark set of interbank networks, demonstrating the explanatory power of the persistence variables in a financial case. In particular, we exploit explicitly the link-specific persistence in order to assess properly what are the preferential lending relations between banks, distinguished from random trading mediated by node-specific characteristics which can be persistent, though.

The remainder of this chapter is organized as follows. In Section 2.1 we introduce our model(s) of dynamic networks. In Section 2.2 we present a novel Expectation-Maximization algorithm for model estimation and in Section 2.3 we run a Monte Carlo exercise to assess the goodness of fit of our estimation method. In Section 2.4 we apply our methodology to the network of the electronic Market of Interbank Deposit. Finally, we conclude with a discussion of our method and open areas for future research.

2.1 The model(s)

In this Section, we describe a benchmark model for link stability and introduce two novel models of dynamic networks: (i) in the first one, the presence or absence of a link can be a copy of the past with a given probability or can be sampled according to a Bernoulli marginal distribution; (ii) in the second model, each graph snapshot does not have an explicit dependence from the past snapshots but the link probability depends on node-specific latent dynamical variables, i.e. the node fitnesses, which evolve stochastically in time with memory of past information; (iii) the third network model combines the copying mechanism of the first model with dynamic node fitnesses of the second model.

We define a temporal network as a time series of graphs, that is the set $(V, \boldsymbol{\mathcal{A}} = \{\boldsymbol{A}^t\}^{t=0,1,\dots,T})$ with $|V| = N$ nodes and a sequence of adjacency matrices $\boldsymbol{\mathcal{A}} = \{\boldsymbol{A}^t\}^{t=0,1,\dots,T}$. A network snapshot is the observed graph at a given time t and is described by the adjacency matrix $\boldsymbol{A}^t$ which has entry $A^t_{ij} = 1$ if the edge from node i to node j is present at time t and zero otherwise. In our models, we exclude graphs with self loops, i.e. the diagonal of $\boldsymbol{A}^t$ is null for all t. The adjacency matrix can be symmetric (undirected graphs) or not (directed ones). In the following, we refer to the undirected case for notational simplicity. The generalization is straightforward and is indeed used in the empirical analysis of the interbank market of Section 2.4.

In our framework, a temporal network is the observable of the following hidden Markov chain:

$$\begin{cases} \mathbb{P}(\boldsymbol{\Theta}^t|\boldsymbol{\Theta}^{t-1}, \boldsymbol{\Phi}) & = h(\boldsymbol{\Theta}^t, \boldsymbol{\Theta}^{t-1}, \boldsymbol{\Phi}) \\ \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1}, \boldsymbol{\Theta}^t, \boldsymbol{\beta}) & = g(\boldsymbol{A}^t, \boldsymbol{A}^{t-1}, \boldsymbol{\Theta}^t, \boldsymbol{\beta}) \end{cases} \tag{2.1}$$

where $\{\boldsymbol{\Theta}^t\}^{t=0,1,\dots,T}$ represents the set of time-varying parameters, which are also called latent variables of the Markov chain. Their dynamics is determined by the one-step transition probability h, whereas g represents the likelihood for the network snapshot at time t given the information about the previous network snapshot, as well as the latent variables $\boldsymbol{\Theta}^t$. Finally $\boldsymbol{\pi} \equiv \{\boldsymbol{\beta}, \boldsymbol{\Phi}\}$ represents the set of static parameters.

Since the Markov chain in Eq. 2.1 has a high-dimensional set of parameters, we reduce the dimensionality by assuming that the node-specific latent variables evolve independently and that there are no explicit spatial correlations among links. Nevertheless, spatial correlations between links are implicitly induced by the latent dynamics.

As mentioned above, in the following, we consider three different specifications of (2.1).

2.1.1 DAR(1) network model for link stability

Link stability can be modeled with the DAR(1) process (1.60) introduced in Section 1.3.1. That is, each link is copied from the past with probability α_{ij}, otherwise it is generated according to a Bernoulli marginal with probability of success χ_{ij}. As analytically proved in Subsection 1.3.1.2, the mechanism of copying from the past in the DAR(1) network model is equivalent to link stability modeled in TERG (1.57).

Networks characterized by high values of $\boldsymbol{\alpha} \equiv \{\alpha_{ij}\}_{i=1,\dots,N,\, j>i}$ tend to preserve the past structure through time, as described before. In particular, notice that each link (i, j) is associated with one persistence parameter α_{ij}, thus resulting in the characterization of the link-specific persistence pattern.

2.1.2 Dynamic fitness model

The second model is a generalization of the fitness network model (1.20) to a dynamic setting that accounts for time evolving node fitness. Fitness is a node property determining its capability of creating links. We assume that each node i is characterized by the fitness θ_i which evolves in time by following a covariance stationary autoregressive process AR(1),

$$\theta_i^t = \phi_{0,i} + \phi_{1,i}\theta_i^{t-1} + \varepsilon_i^t\,, \quad \forall i = 1,\dots,N \tag{2.2}$$

where $\phi_{0,i} \in \mathbb{R}$, $|\phi_{1,i}| < 1$ and the i.i.d. variables $\varepsilon_i^t \sim \mathcal{N}(0, \sigma_i^2)$. This choice is consistent with the Markovian assumption in (2.1). Moreover, the latent node state θ_i^t evolves in $\mathbb{R}$ between timesteps, but large changes are unlikely because of the Gaussian transition probabilities. This is consistent with the assumption that the network topology changes smoothly in time. Finally, assuming a Gaussian transition probability represents a simplification for model estimation.

The conditional probability for the network at time t is

$$\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{\Theta}^t) = \prod_{i,j>i} \frac{e^{A_{ij}^t(\theta_i^t+\theta_j^t)}}{1+e^{(\theta_i^t+\theta_j^t)}}, \tag{2.3}$$

where $\boldsymbol{\Theta}^t \equiv \{\theta_i^t\}_{i=1,\dots,N}$ is the vector of time-varying parameters. In (2.3) we assume that each link is independently sampled and the probability of a link between node i and node j at time t is determined by the corresponding θ_i^t and θ_j^t. The larger is θ_i^t, the larger is the probability for all links incident to node i.

We refer to this model also as Temporally Generalized Random Graphs (TGRG)[1] and the specification of (2.1) for TGRG is the following,

[1] [Park and Newman, 2004] refer to the (static) fitness model as Generalized (w.r.t. the Erdös-Rényi model) Random Graphs, hence the name of the temporal generalization.

$$\begin{cases} \mathbb{P}(\theta_i^t|\theta_i^{t-1},\boldsymbol{\Phi}_i) & = f(\theta_i^t|\phi_{0,i}+\phi_{1,i}\theta_i^{t-1},\sigma_i^2)\ \forall i=1,...,N \\ \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{\Theta}^t) & = \prod_{i,j>i}\mathbb{P}(A_{ij}^t|\theta_i^t,\theta_j^t) = \prod_{i,j>i}\frac{e^{A_{ij}^t(\theta_i^t+\theta_j^t)}}{1+e^{(\theta_i^t+\theta_j^t)}} \end{cases} \tag{2.4}$$

and $\mathbb{P}(\boldsymbol{\Theta}^t|\boldsymbol{\Theta}^{t-1},\boldsymbol{\Phi}) = \prod_{i=1}^N \mathbb{P}(\theta_i^t|\theta_i^{t-1},\boldsymbol{\Phi}_i)$ according to the hypothesis of independence, where $f(\theta_i^t|\phi_{0,i}+\phi_{1,i}\theta_i^{t-1},\sigma_i^2)$ is the density of normal variable with mean $\phi_{0,i}+\phi_{1,i}\theta_i^{t-1}$ and variance σ_i^2. The set of static parameters is $\boldsymbol{\Phi} \equiv \{\boldsymbol{\Phi}_i\}_{i=1,...,N}$ with $\boldsymbol{\Phi}_i \equiv \{\phi_{0,i},\phi_{1,i},\sigma_i\}$.

The TGRG model is fully determined by the $3\times N$ static parameters $\boldsymbol{\Phi}$. In the next Section we propose an Expectation-Maximization scheme based on MAP inference to estimate the model parameters and time-varying parameters. It alternates between an Expectation step where we infer the time-varying parameters $\{\boldsymbol{\Theta}^t\}^{t=0,1,...,T}$ and the Maximization step where we maximize the posterior of the static parameters conditional on the expectations $\{\hat{\boldsymbol{\Theta}}^t\}^{t=0,1,...,T}$.

Regarding the consistency of the estimator, [Chatterjee et al., 2011, Yan and Xu, 2013] show that, in the asymptotic limit of dense (undirected) networks with diverging number of nodes, maximum likelihood estimator of the (static) fitness model is consistent, thus the model is identifiable, see [Gabrielsen, 1978]. However, identifiability of the dynamic fitness model in the sparse regime is still an open question. In Section 2.3, we show numerically that the proposed EM method leads to unbiased estimation for both static and dynamic parameters.

Time-autocorrelated node fitnesses may induce link persistence. In fact, the probability of a link between two specific nodes $e^{(\theta_i^t+\theta_j^t)}/(1+e^{(\theta_i^t+\theta_j^t)})$ is persistent if θ_i^t and θ_j^t are autocorrelated. Note that link persistence occurs as a consequence of node properties. For TGRG, the two-point probability mass function for lagged links and the ACF of link state can be semi-analytically computed (see the section in appendix of this chapter).

The generalization of the TGRG model to directed networks can be accomplished by distinguishing between the out-degree and the in-degree and by introducing two fitnesses for each node i, i.e. $\theta_i^{t,out}$ and $\theta_i^{t,in}$. The probability of a link from node i to node j at time t given the latent variables $\theta_i^{t,out}$ and $\theta_j^{t,in}$ is $\mathbb{P}(A_{ij}^t|\theta_i^{t,out},\theta_j^{t,in}) = \frac{e^{A_{ij}^t(\theta_i^{t,out}+\theta_j^{t,in})}}{1+e^{(\theta_i^{t,out}+\theta_j^{t,in})}}$. Then, everything follows similarly to the undirected case with the exception that $\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{\Theta}^t)$ is invariant under a linear transformation for the latent node states: $\theta_i^{t,out} \mapsto \theta_i^{t,out}+c_t\ \forall i=1,...,N$, $\theta_j^{t,in} \mapsto \theta_j^{t,in}-c_t\ \forall j=1,...,N$, where $\{c_t\}_{t=0,1,...,T} \in \mathbb{R}^{T+1}$. This symmetry arises because the total number of outgoing links has to be equal to the total number of incoming links at each time. It makes the model as non-identifiable. However, when we are interested in the temporal patterns captured by the model, results remain unchanged as long as we preserve the sum of any $\theta_i^{t,out}$ and any $\theta_j^{t,in}$. Hence, we can take one of the fitnesses as constant in time and infer the other ones. Notice that this identifiability problem is well known for the exponential family of probability distributions for directed static graphs since the milestone work of [Holland and Leinhardt, 1981] and several studies have been focused on the same parameterization we adopt here, *e.g.*

see [Rinaldo et al., 2013, Luo et al., 2017]. In this case, any comparison between inferred fitnesses and any observed quantity has to be properly considered.

Finally, let us notice that we can interpret TGRG as an extension of Exponential Random Graphs (ERG) by [Park and Newman, 2004] to the dynamic case. ERG ensembles are probability distributions of networks obtained by maximizing the Shannon entropy under some constraints on the average value of a set of network observables. If this set is the degree sequence, the Lagrange multipliers of the entropy constrained optimization can be directly linked to the latent variables of our model. Differently from other dynamic extension of ERG (see for example [Hanneke et al., 2010]) where dynamical (*i.e.* two-time) observables are used as constraints, here we choose a dynamical model for the latent variable, namely the AR(1) process, and introduce an estimation method for them.

2.1.3 Dynamic fitness model with link stability

The persistence pattern associated with the copying mechanism described by (1.60) can coexist with the node fitnesses evolving in time according to (2.2). This can be captured by the following specification of the model (2.1),

$$\begin{cases} \mathbb{P}(\theta_i^t|\theta_i^{t-1},\boldsymbol{\Phi}_i) & = f(\theta_i^t|\phi_{0,i}+\phi_{1,i}\theta_i^{t-1},\sigma_i^2)\ \ \forall i=1,...,N \\ \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\Theta^t,\boldsymbol{\alpha}) & = \prod_{i,j>i}\left(\alpha_{ij}\delta_{A_{ij}^t,A_{ij}^{t-1}}+(1-\alpha_{ij})\frac{e^{A_{ij}^t(\theta_i^t+\theta_j^t)}}{1+e^{(\theta_i^t+\theta_j^t)}}\right) \end{cases} \tag{2.5}$$

and $\mathbb{P}(\Theta^t|\Theta^{t-1},\boldsymbol{\Phi})=\prod_{i=1}^N\mathbb{P}(\theta_i^t|\theta_i^{t-1},\boldsymbol{\Phi}_i)$ according to the hypothesis of independence, where $f(\theta_i^t|\phi_{0,i}+\phi_{1,i}\theta_i^{t-1},\sigma_i^2)$ is the density of a normal variable with mean $\phi_{0,i}+\phi_{1,i}\theta_i^{t-1}$ and variance σ_i^2, $\alpha_{ij}\in[0,1]$ and $\boldsymbol{\alpha}\equiv\{\alpha_{ij}\}\ \forall i,j=1,...,N$ with $\alpha_{ij}=\alpha_{ji}$ for undirected networks, $\boldsymbol{\Phi}\equiv\{\phi_{0,i},\phi_{1,i},\sigma_i\}_{i=1,...,N}$ with $\phi_{0,i}\in\mathbb{R}$, $|\phi_{1,i}|<1$ and $\sigma_i>0\ \forall i$, and $\theta_i^t\in\mathbb{R}\ \forall i,t$.

This model can be interpreted as a mixture of the two mechanisms, i.e. the one of copying the presence or absence of a link from the past with probability α_{ij} and the one of time evolving marginals described by the TRGR model with probability $1-\alpha_{ij}$. Let us stress that the temporal pattern generated by the fitness dynamics does not concern a specific link but it is a node property. Thus under this mechanism, links incident on the same node tend to have similar persistence properties. On the contrary, the persistence of the copying mechanism is a link property, and links incident on the same node can have very different persistence properties. The parameter α_{ij} disentangles the importance of these two effects in determining the dynamics of the link (i,j).

The model (2.5), here referred as DAR-TGRG, is fully determined by the $\binom{N}{2}$ parameters $\boldsymbol{\alpha}$ and the $3\times N$ parameters $\boldsymbol{\Phi}$ and it can be estimated by the Expectation-Maximization algorithm we propose in the next Section.

DAR-TGRG is a generalization of the TGRG model, then the same distinction between undirected and directed networks needs to be considered and similar con-

clusions can be obtained for the identifiability of the model. In particular, in the case of undirected networks, if TGRG is identifiable, then this is a strong evidence in favor of the identifiability of the DAR-TGRG model. In fact, a symmetry of the likelihood involving both the parameters associated with the mechanism of copying from the past and the parameters associated with the fitness dynamics, should hold independently from the adjacency matrices of consecutive network snapshots, in particular it should do not depend on the set of Kronecker deltas $\{\delta_{A_{ij}^{t-1},A_{ij}^{t}}\}_{i,j=1,\dots,N}^{t=0,1,\dots,T}$.

2.2 Estimation method

We now describe the procedure for the estimation of the DAR-TGRG model. We propose a Marginal Maximum a Posteriori (MMAP) approach together with an Expectation-Maximization like (EM) procedure to estimate the static parameters and learn the latent variables dynamics. The estimation method for TGRG model is simply obtained by setting parameters $\alpha_{ij}\ \forall i,j=1,\dots,N$ equal to zero in the following equations.

Let denote $\boldsymbol{\Theta} \equiv \{\Theta^t\}^{t=1,\dots,T}$, $\boldsymbol{\mathcal{A}} \equiv \{\boldsymbol{A}^t\}^{t=0,1,\dots,T}$ and $\boldsymbol{\pi} \equiv \{\boldsymbol{\Phi},\boldsymbol{\alpha}\}$. The marginal posterior distribution of the static parameters reads as

$$\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\mathcal{A}}) = \frac{\mathbb{P}(\boldsymbol{\pi})}{\mathbb{P}(\boldsymbol{\mathcal{A}})}\int [d\Gamma]\mathbb{P}(\boldsymbol{\mathcal{A}},\Gamma|\boldsymbol{\pi}) \propto \mathbb{P}(\boldsymbol{\pi})Z_{\boldsymbol{\pi}} \tag{2.6}$$

where $[d\Gamma]$ represents the measure over the probability space for $\boldsymbol{\Theta}$ and $Z_{\boldsymbol{\pi}} = \int [d\Gamma]\mathbb{P}(\boldsymbol{\mathcal{A}},\Gamma|\boldsymbol{\pi})$. It is used to learn the most likely set of parameters $\hat{\boldsymbol{\pi}}$ given the data.

Using smooth priors $\mathbb{P}(\boldsymbol{\pi})$, $\hat{\boldsymbol{\pi}}$ is obtained by extremizing over $\boldsymbol{\pi}$ the log-posterior $l(\boldsymbol{\pi}) \equiv \log\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\mathcal{A}})$, i.e. by solving the equations

$$\begin{aligned}\partial_{\boldsymbol{\pi}}\log Z_{\boldsymbol{\pi}} = \partial_{\boldsymbol{\pi}}\log\int [d\boldsymbol{\Gamma}]\mathbb{P}(\boldsymbol{\mathcal{A}},\boldsymbol{\Gamma}|\boldsymbol{\pi}) = \frac{\int [d\boldsymbol{\Gamma}]\partial_{\boldsymbol{\pi}}\mathbb{P}(\boldsymbol{\mathcal{A}},\boldsymbol{\Gamma}|\boldsymbol{\pi})}{\int [d\boldsymbol{\Gamma}]\mathbb{P}(\boldsymbol{\mathcal{A}},\boldsymbol{\Gamma}|\boldsymbol{\pi})} = \\ = \langle\partial_{\boldsymbol{\pi}}\log(\mathbb{P}(\boldsymbol{\mathcal{A}},\Gamma|\boldsymbol{\pi}))\rangle_{Z_{\boldsymbol{\pi}}^{-1}\mathbb{P}(\boldsymbol{\mathcal{A}},\boldsymbol{\Theta}|\boldsymbol{\pi})} = -\partial_{\boldsymbol{\pi}}\log(\mathbb{P}(\boldsymbol{\pi})).\end{aligned} \tag{2.7}$$

Since maximizing (2.6), *i.e.* solving (2.7), needs computing expectations with respect to the posterior distribution of the time-varying parameters (latent variables)

$$\mathbb{P}(\boldsymbol{\Theta}|\boldsymbol{\mathcal{A}},\boldsymbol{\pi}) = \frac{\mathbb{P}(\boldsymbol{\mathcal{A}},\boldsymbol{\Theta}|\boldsymbol{\pi})}{\int [d\Gamma]\mathbb{P}(\boldsymbol{\mathcal{A}},\boldsymbol{\Gamma}|\boldsymbol{\pi})} = Z_{\boldsymbol{\pi}}^{-1}\mathbb{P}(\boldsymbol{\mathcal{A}},\boldsymbol{\Theta}|\boldsymbol{\pi}), \tag{2.8}$$

the latter requiring in turn an estimate of $\boldsymbol{\pi}$, one typically repeats the two steps until convergence starting from an initial guess, in a standard Expectation-Maximization framework, see [Friedman et al., 2001]. Given $\hat{\boldsymbol{\pi}}$, the dynamics of the fitnesses $\hat{\boldsymbol{\Theta}}$ is obtained by maximizing $\mathbb{P}(\boldsymbol{\Theta}|\boldsymbol{\mathcal{A}},\hat{\boldsymbol{\pi}})$.

Since we have a large number of latent variables this approach is preferred with respect to the maximization of the joint posterior distribution (Joint Maximum A Posteriori (JMAP) approach)

$$\mathbb{P}(\boldsymbol{\Theta}, \boldsymbol{\pi} | \boldsymbol{\mathcal{A}}) \propto \mathbb{P}(\boldsymbol{\mathcal{A}} | \boldsymbol{\Theta}, \boldsymbol{\pi}) \mathbb{P}(\boldsymbol{\Theta} | \boldsymbol{\pi}) \mathbb{P}(\boldsymbol{\pi}), \tag{2.9}$$

that suffers inherent problems first recognized by [Little and Rubin, 1983, Neyman et al., 1948]: they showed that when static parameters are estimated simultaneously with the latent variables, the maximum likelihood estimates of the former need not be consistent as sample size increases.

The drawback of MMAP is that the integral over the posterior distribution of latent variables is typically intractable and cannot be computed exactly, thus the expectation step needs some approximations. In the following we explain in full detail each step of the proposed procedure.

A last remark concerns the choice of the prior $\mathbb{P}(\boldsymbol{\pi})$: in the following we always assume flat distributions, *i.e.* the right hand side of (2.7) is put to zero. However we find high stability with respect to the choice of the prior, as soon as it is a smooth distribution. This is not surprisingly because, since the parameters are not time-varying, the l.h.s of (2.7) is typically T times larger than the r.h.s., meaning that the contribution of the prior asymptotically must disappear as the sample size increases.

2.2.1 Inference of time-varying parameters

The inference of the latent variables according to the MMAP approach refers to the maximization of the posterior (2.8) of $\boldsymbol{\Theta} = \{\Theta^t\}^{t=1,\dots,T}$ (time-varying parameters) by conditioning on the value of the static parameters $\boldsymbol{\pi}$.

We do not solve the inference problem for the time series of dynamic parameters $\boldsymbol{\Theta}$ overall, *i.e.* by maximizing (2.8). Instead, we infer step by step the parameters Θ^t by conditioning on the expectations $\hat{\Theta}^{t-1}$, that are the one step backward estimates[2] for Θ^{t-1}.

Let us focus on the inference at the generic time $t \neq 0$ when the previous network snapshot is observed and let $\mathscr{F}^t \equiv \{\boldsymbol{A}^{t-1}, \boldsymbol{\pi}\}$ be the information set for the considered problem. From the Bayes' theorem, it is

$$\mathbb{P}(\Theta^t | \boldsymbol{A}^t, \Theta^{t-1}, \mathscr{F}^t) = \frac{\mathbb{P}(\boldsymbol{A}^t | \Theta^t, \mathscr{F}^t) \mathbb{P}(\Theta^t | \Theta^{t-1}, \mathscr{F}^t) \mathbb{P}(\Theta^{t-1} | \mathscr{F}^t)}{\mathbb{P}(\boldsymbol{A}^t, \Theta^{t-1} | \mathscr{F}^t)}. \tag{2.10}$$

Hence, by conditioning on the expectation for Θ^{t-1}, i.e. $\hat{\Theta}^{t-1}$, the inference problem can be solved by maximizing

$$\mathbb{P}(\Theta^t | \boldsymbol{A}^t, \boldsymbol{A}^{t-1}, \hat{\Theta}^{t-1}, \Pi) \propto \mathbb{P}(\boldsymbol{A}^t | \boldsymbol{A}^{t-1}, \Theta^t, \boldsymbol{\alpha}) \mathbb{P}(\Theta^t | \hat{\Theta}^{t-1}, \boldsymbol{\Phi}), \quad \forall t = 1, \dots, T.$$

[2] Here, we are assuming to know the expectation for Θ^0, *i.e.* $\hat{\Theta}^0$. Below, we explain how to infer the initial point for the latent dynamics.

(2.11)

Maximizing (2.11) is equivalent to solve the following problem

$$\hat{\Theta}^t = \underset{\Theta_t}{\operatorname{argmax}} \left(\log \mathbb{P}(\boldsymbol{A}^t | \boldsymbol{A}^{t-1}, \Theta_t, \boldsymbol{\alpha}) + \log \mathbb{P}(\Theta_t | \hat{\Theta}^{t-1}, \boldsymbol{\Phi}) \right), \tag{2.12}$$

where $\mathbb{P}(\Theta^t | \hat{\Theta}^{t-1}, \boldsymbol{\Phi}) \equiv \prod_{i=1}^{N} f(\theta_i^t | \phi_{0,i} + \phi_{1,i} \hat{\theta}_i^{t-1}, \sigma_i^2)$ is the Gaussian probability density function associated with the transition probability for the latent variables. (2.12) is equivalent to the following system of non linear equations,

$$\begin{aligned} &\left[\sum_{j \neq i} \left(\frac{(1-\alpha_{ij}) \frac{e^{A_{ij}^t (\theta_i^t + \theta_j^t)}}{1+e^{(\theta_i^t + \theta_j^t)}}}{\alpha_{ij} \delta_{A_{ij}^t, A_{ij}^{t-1}} + (1-\alpha_{ij}) \frac{e^{A_{ij}^t (\theta_i^t + \theta_j^t)}}{1+e^{(\theta_i^t + \theta_j^t)}}} \right) \left(-A_{ij}^t + \frac{e^{(\theta_i^t + \theta_j^t)}}{1+e^{(\theta_i^t + \theta_j^t)}} \right) \right] + \\ &- \frac{\theta_i^t - \phi_{0,i} - \phi_{1,i} \hat{\theta}_i^{t-1}}{\sigma_i^2} = 0, \ \forall i = 1, ..., N. \end{aligned} \tag{2.13}$$

It can be solved by the following iterative proportional fitting procedure:

1. assume any starting point $\hat{\theta}_i^t \ \forall i = 1, ..., N$ for the node fitness[3];
2. then, solve one by one the equations in Eq. 2.13 by conditioning on $\hat{\theta}_j^t \ \forall j \neq i$;
3. update the value for $\hat{\theta}_i^t$ with the solution of the corresponding equation;
4. repeat until convergence.

2.2.1.1 Link prediction

The proposed method for the inference of the time-varying parameters is a statistical filtering algorithm. Filtering is an operation that involves the extraction of information about a latent quantity of interest at time t by using data measured up to and including t, like in Kalman filter and its extensions [Chen et al., 2003]. However, differently from Kalman filter, we study the case of a Hidden Markov Model with continuous-valued state space, i.e. continuous-valued state vector Θ^t, but binary measurement matrix, i.e. $\boldsymbol{A}^t$. Notice that the proposed method can be used for on-line inference: once the off-line learning of the static parameters is completed, we can solve the filtering problem for Θ^t (2.12) whenever the new measurement $\boldsymbol{A}^t$ is available.

On-line inference is particularly useful for link prediction: let $\boldsymbol{A}^t$ be the observation at the current time and we want to construct the one-step-ahead forecast, i.e. $\mathbb{E}[\boldsymbol{A}^{t+1} | \boldsymbol{A}^t]$. Once $\hat{\Theta}^t$ is inferred on-line by solving (2.12), the one-step-ahead forecast is constructed by averaging over the probability distribution

$$\mathbb{P}(\boldsymbol{A}^{t+1} | \boldsymbol{A}^t, \hat{\Theta}^t, \Pi) = \int [d\Theta^{t+1}] \mathbb{P}(\boldsymbol{A}^{t+1} | \boldsymbol{A}^t, \Theta^{t+1}, \boldsymbol{\alpha}) \mathbb{P}(\Theta^{t+1} | \hat{\Theta}^t, \boldsymbol{\Phi})$$

[3] A possible choice is $\hat{\theta}_i^t = \phi_{0,i} + \phi_{1,i} \hat{\theta}_i^{t-1}$

obtained by projecting the latent state. In Section 2.4 we show an application of this procedure.

2.2.2 Learning the static parameters

The inference of the static parameters according to the MMAP approach refers to the maximization of the posterior (2.6) of the static parameters $\boldsymbol{\pi}$ by using the information on time-varying parameters $\boldsymbol{\Theta}$. However, as claimed before, the posterior distribution of latent variables is intractable because of strict nonlinearity of the distribution function. Hence, we use an approximated posterior, where the transition probabilities for the latent variables (Γ_t) are conditioned on the expectations at the previous step ($\hat{\boldsymbol{\Theta}}_{t-1}$), *i.e.*

$$\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\mathcal{A}}) \propto \int [d\boldsymbol{\Gamma}]\mathbb{P}(\boldsymbol{\mathcal{A}},\boldsymbol{\Gamma}|\boldsymbol{\pi}) \approx \prod_t \int [d\Gamma_t]\mathbb{P}(\boldsymbol{A}^t,\Gamma_t|\hat{\boldsymbol{\Theta}}_{t-1},\boldsymbol{A}^{t-1},\boldsymbol{\pi}). \tag{2.14}$$

In the next Section, we verify numerically the consistency of this approximation by showing how the MMAP approach leads to unbiased estimation of the parameters.

2.2.2.1 Learning $\boldsymbol{\alpha}$

Let us assume to know the static parameters $\boldsymbol{\Phi}$ and consider the problem of learning $\boldsymbol{\alpha}$. The most likely estimate for $\boldsymbol{\alpha}$ is obtained by maximizing the associated posterior in (2.6), that is

$$\mathbb{P}(\boldsymbol{\alpha}|\boldsymbol{\mathcal{A}}) \propto \int [d\boldsymbol{\Theta}]\,\mathbb{P}(\boldsymbol{\mathcal{A}},\boldsymbol{\Theta}|\boldsymbol{\alpha},\boldsymbol{\Phi}) = \int \prod_{t=1}^{T} \left[\, d\Theta^t \right] \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\Theta^t,\boldsymbol{\alpha}) f(\Theta^t|\Theta^{t-1},\boldsymbol{\Phi}). \tag{2.15}$$

The integral in (2.15) is infeasible because of the nonlinearity of the probability distribution. Hence, we approximate the likelihood of parameters $\boldsymbol{\alpha}$ as

$$\begin{aligned} l_{\boldsymbol{\alpha}} &\equiv \int \prod_{t=1}^{T} \left[\, d\Theta^t \right] \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\Theta^t,\boldsymbol{\alpha}) f(\Theta^t|\Theta^{t-1},\boldsymbol{\Phi}) \\ &\approx \prod_{t=1}^{T} \int \left[\, d\Theta^t \right] \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\Theta^t,\boldsymbol{\alpha}) f(\Theta^t|\hat{\Theta}^{t-1},\boldsymbol{\Phi}) \equiv \tilde{l}_{\boldsymbol{\alpha}}, \end{aligned} \tag{2.16}$$

where $\hat{\Theta}^{t-1} = \{\hat{\theta}_i^{t-1}\}_{i=1,\dots,N}$ is the expectation of latent variables at time $t-1$ that we obtained by solving (2.13).

Let us focus on the learning of parameter α_{ij}. When we aim to obtain the solution for α_{ij}, the only time-varying parameters that are involved in the learning are the

ones associated with node i and node j, i.e. $\{\theta_i^t\}^{t=1,\dots,T}$ and $\{\theta_j^t\}^{t=1,\dots,T}$. Hence, the most likely estimate for α_{ij} is the value that maximizes

$$\tilde{S}_{\alpha_{ij}} = \log \tilde{l}_{\alpha_{ij}} = \sum_{t=1}^{T} \log \int dxdy \left(\alpha_{ij} \delta_{A_{ij}^t, A_{ij}^{t-1}} + (1-\alpha_{ij}) \frac{e^{A_{ij}^t (x+y)}}{1+e^{(x+y)}} \right) \times \\ \times f(x|\phi_{0,i} + \phi_{1,i} \hat{\theta}_i^{t-1}, \sigma_i^2) f(y|\phi_{0,j} + \phi_{1,j} \hat{\theta}_j^{t-1}, \sigma_j^2). \tag{2.17}$$

In the learning procedure, the following double integral is involved,

$$\mathscr{I}_{A_{ij}^t}(\hat{\theta}_i^{t-1}, \hat{\theta}_j^{t-1}, \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j) = \tag{2.18}$$
$$\int dxdy \frac{e^{A_{ij}^t (x+y)}}{1+e^{(x+y)}} f(x|\phi_{0,i} + \phi_{1,i} \hat{\theta}_i^{t-1}, \sigma_i^2) f(y|\phi_{0,j} + \phi_{1,j} \hat{\theta}_j^{t-1}, \sigma_j^2),$$

which can be solved numerically. However, we propose to apply the following integral identity proposed by [Polson et al., 2013]

$$\frac{(e^{\psi})^a}{(1+e^{\psi})^b} = 2^{-b} e^{(a-\frac{b}{2})\psi} \int_0^{\infty} e^{-\frac{\omega \psi^2}{2}} p_{PG}(\omega) d\omega \tag{2.19}$$

where $b > 0$, $a, \psi \in \mathbb{R}$, and $p_{PG} : [0,\infty) \mapsto [0,1)$ is the density of the Pólya-Gamma distribution. There is no a closed-form expression for p_{PG} but we evaluate it numerically. The double integral in (2.18) is therefore equivalent to the following integral,

$$\mathscr{I}_{A_{ij}^t}(\hat{\theta}_i^{t-1}, \hat{\theta}_j^{t-1}, \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j) \equiv \int_0^{\infty} \frac{d\omega}{2} p_{PG}(\omega) \mathscr{K}_{A_{ij}^t}(\omega, \hat{\theta}_i^{t-1}, \hat{\theta}_j^{t-1}, \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j) \tag{2.20}$$

where

$$\mathscr{K}_{A_{ij}^t}(\omega, \hat{\theta}_i^{t-1}, \hat{\theta}_j^{t-1}, \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j) = \\ = \frac{\exp\left(\frac{\sigma_i^2 + \sigma_j^2 + 4(\phi_{0,i} + \phi_{1,i} \hat{\theta}_i^{t-1} + \phi_{0,j} + \phi_{1,j} \hat{\theta}_j^{t-1})(2A_{ij}^t - 1 - \omega(\phi_{0,i} + \phi_{1,i} \hat{\theta}_i^{t-1} + \phi_{0,j} + \phi_{1,j} \hat{\theta}_j^{t-1}))}{8(1+\omega(\sigma_i^2 + \sigma_j^2))} \right)}{\sqrt{1+\omega(\sigma_i^2 + \sigma_j^2)}}. \tag{2.21}$$

We propose to evaluate numerically the integral in (2.20). This gives the advantage of computing a single integral.

Then α_{ij} is estimated by solving the equation $\partial_{\alpha_{ij}} \tilde{S}_{\alpha_{ij}} = 0$, which can be explicitly rewritten as

$$\sum_{t=1}^{T} \frac{\delta_{A_{ij}^t, A_{ij}^{t-1}} - \mathscr{I}_{A_{ij}^t}(\hat{\theta}_i^{t-1}, \hat{\theta}_j^{t-1}, \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j)}{\alpha_{ij} \delta_{A_{ij}^t, A_{ij}^{t-1}} + (1-\alpha_{ij}) \mathscr{I}_{A_{ij}^t}(\hat{\theta}_i^{t-1}, \hat{\theta}_j^{t-1}, \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j)} = 0. \tag{2.22}$$

The solution of (2.22) represents the most likely estimate $\hat{\alpha}_{ij}$ given the data (and the approximation of the posterior).

2.2.2.2 Learning $\boldsymbol{\Phi}$

Let us assume to know the static parameters $\boldsymbol{\alpha}$ and consider the problem of learning $\boldsymbol{\Phi}$. Similarly to the previous Subsection, we use conditions on the expectations for the latent variables to obtain an approximated log-likelihood for $\boldsymbol{\Phi}$,

$$\begin{aligned}\tilde{S}_{\boldsymbol{\Phi}} = \sum_{t=1}^{T} \log \int \left[\prod_{k=1}^{N} dx_k \, f(x_k|\phi_{0,k} + \phi_{1,k}\hat{\theta}_k^{t-1}, \sigma_k^2) \right] \times \\ \times \left[\prod_{i,j>i} \alpha_{ij}\delta_{A_{ij}^t, A_{ij}^{t-1}} + (1-\alpha_{ij}) \frac{e^{A_{ij}^t (x_i + x_j)}}{1 + e^{(x_i + x_j)}} \right].\end{aligned} \tag{2.23}$$

Let us focus on the learning of parameters $\boldsymbol{\Phi}_i \equiv \{\phi_{0,i}, \phi_{1,i}, \sigma_i\}$. Because of the marginal distribution, each time-varying parameter θ_i^t is coupled with all the others and this prevents the valuation of the multiple integral in (2.23). Hence, we adopt the following approximation for the probability measure,

$$\begin{aligned}&\prod_{k=1}^{N} dx_k \, f(x_k|\phi_{0,k} + \phi_{1,k}\hat{\theta}_k^{t-1}, \sigma_k^2) \approx \\ &\approx dx_i \, f(x_i|\phi_{0,i} + \phi_{1,i}\hat{\theta}_i^{t-1}, \sigma_i^2) \prod_{k \neq i} dx_k \, \delta(x_k - \hat{\theta}_k^t) f(x_k|\phi_{0,k} + \phi_{1,k}\hat{\theta}_k^{t-1}, \sigma_k^2),\end{aligned} \tag{2.24}$$

i.e. we condition on the expectations at time t for all the latent variables with the exception of θ_i^t. Then, we maximize the following approximated likelihood,

$$\begin{aligned}\tilde{S}_{\boldsymbol{\Phi}_i} = \sum_{t=1}^{T} \log \int_{-\infty}^{\infty} dx_i \, f(x_i|\phi_{0,i} + \phi_{1,i}\hat{\theta}_i^{t-1}, \sigma_i^2) \times \\ \times \left(\prod_{j \neq i} \alpha_{ij}\delta_{A_{ij}^t, A_{ij}^{t-1}} + (1-\alpha_{ij}) \frac{e^{A_{ij}^t (x_i + \hat{\theta}_j^t)}}{1 + e^{(x_i + \hat{\theta}_j^t)}} \right),\end{aligned} \tag{2.25}$$

or, equivalently, we solve the system of equations $\partial_{\boldsymbol{\Phi}_i} \tilde{S}_{\boldsymbol{\Phi}_i} = 0$.

Let us define the following partition function $\forall t = 1, ..., T$,

$$Z_{\boldsymbol{\Phi}_i}^t \equiv \int_{-\infty}^{+\infty} dx \, f(x_i|\phi_{0,i} + \phi_{1,i}\hat{\theta}_i^{t-1}, \sigma_i^2) \left(\prod_{j \neq i} \alpha_{ij}\delta_{A_{ij}^t, A_{ij}^{t-1}} + (1-\alpha_{ij}) \frac{e^{A_{ij}^t (x + \hat{\theta}_j^t)}}{1 + e^{(x + \hat{\theta}_j^t)}} \right) \tag{2.26}$$

and let $\mu^t_{\boldsymbol{\Phi}_i}$ and $\Sigma^t_{\boldsymbol{\Phi}_i}$ be the first and the second moment of the distribution, respectively.

The system of equations $\partial_{\boldsymbol{\Phi}_i}\tilde{S}_{\boldsymbol{\Phi}_i} = 0$ reads explicitly as

$$\begin{cases} \langle \boldsymbol{\mu}_{\boldsymbol{\Phi}_i} \rangle - \phi_{0,i} - \langle L\hat{\boldsymbol{\theta}}_i \rangle \phi_{1,i} & = 0 \\ \frac{1}{T}(L\hat{\boldsymbol{\theta}}_i^{\intercal}\, \boldsymbol{\mu}_{\boldsymbol{\Phi}_i}) - \langle L\hat{\boldsymbol{\theta}}_i \rangle \phi_{0,i} - \frac{1}{T}(L\hat{\boldsymbol{\theta}}_i^{\intercal}\, L\hat{\boldsymbol{\theta}}_i)\phi_{1,i} & = 0 \\ \sigma_i^2 - \Big(\langle \boldsymbol{\Sigma}_{\boldsymbol{\Phi}_i} \rangle + \phi_{0,i}^2 + \frac{1}{T}(L\hat{\boldsymbol{\theta}}_i^{\intercal}\, L\hat{\boldsymbol{\theta}}_i)\phi_{1,i}^2 - 2\langle \boldsymbol{\mu}_{\boldsymbol{\Phi}_i} \rangle \phi_{0,i} - 2\frac{1}{T}(L\hat{\boldsymbol{\theta}}_i^{\intercal}\, \boldsymbol{\mu}_{\boldsymbol{\Phi}_i})\phi_{1,i} + 2\langle L\hat{\boldsymbol{\theta}}_i \rangle \phi_{0,i}\phi_{1,i} \Big) & = 0 \end{cases} \tag{2.27}$$

where bold symbols represent T-dimensional vectors, *e.g.* $\boldsymbol{x} = (x^1, x^2, ..., x^T)'$, angle brackets denote time average, *e.g.* $\langle \boldsymbol{x} \rangle \equiv \frac{1}{T}\sum_{t=1}^{T} x^t$, and L is the *lag operator*, *i.e.* $Lx^t = x^{t-1}$.[4]

The system of nonlinear equations can be solved with the following iterative proportional fitting procedure:

1. assume any starting point $\boldsymbol{\Phi}_i^0$;
2. compute $\mu^t_{\boldsymbol{\Phi}_i^0}$ and $\Sigma^t_{\boldsymbol{\Phi}_i^0}$ $\forall t = 1, ..., T$;
3. solve the system of equations in (2.27) by substituting $\mu^t_{\boldsymbol{\Phi}_i^0} \to \mu^t_{\boldsymbol{\Phi}_i}$ and $\Sigma^t_{\boldsymbol{\Phi}_i^0} \to \Sigma^t_{\boldsymbol{\Phi}_i}$ $\forall t = 1, ..., T$;
4. update the values for $\boldsymbol{\Phi}_i^0$ and continue until convergence.

2.2.3 Expectation-Maximization algorithm

The point estimation of the model by means of the Expectation-Maximization (EM) algorithm consists in alternating the inference of the latent variables (Expectation step) and the learning of the static parameters (Maximization step) up to convergence.

2.2.3.1 Single Snapshot Inference (SSI)

As a starting point of the method, the time-varying parameters $\{\Theta^t\}^{t=0,1,...,T}$ can be estimated by single snapshot inference, *i.e.* given the network snapshot at time t and by assuming $\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{\gamma}) = \prod_{i,j>i} \frac{e^{A^t_{ij}(\gamma_i+\gamma_j)}}{1+e^{(\gamma_i+\gamma_j)}}$, we solve snapshot by snapshot the problem of inference of the (static) fitness model, *i.e.*

$$\tilde{\Theta}^t = \underset{\boldsymbol{\gamma}}{\text{argmax}}\ \log \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{\gamma}) \quad \forall t = 0, 1, ...T \tag{2.28}$$

[4] Notice that $L\hat{\theta}_i^1 = \hat{\theta}_i^0$ represents the latent state at the initial time. It is estimated as described in Section 2.2.3.

and we obtain a naive estimation $\tilde{\boldsymbol{\Theta}} \equiv \{\tilde{\Theta}^t\}^{t=0,1,...,T}$ of the latent states of the Markov chain in (2.5). In particular, we infer the latent state at the initial time, i.e. $\hat{\Theta}^0 \equiv \tilde{\Theta}^0$.

Then, we estimate the autoregressive process AR(1) in (2.2) for each naively inferred fitness, *i.e.* $\{\tilde{\theta}_i^t\}^{t=0,1,...,T} \ \forall i = 1,...,N$, to obtain a naive estimation of the static parameters $\tilde{\boldsymbol{\Phi}}$. Finally, a naive estimation of $\tilde{\boldsymbol{\alpha}}$ can be obtained by solving (2.22) with naively inferred $\{\tilde{\Theta}^t\}^{t=0,1,...,T}$ and $\tilde{\boldsymbol{\Phi}}$. We refer to this naive estimation method as the Single Snapshot Inference (SSI) of the DAR-TGRG model.

2.2.3.2 The EM algorithm for MMAP

The Expectation-Maximization algorithm based on the MMAP approach alternates the following steps up to convergence:

1. Assume SSI as starting point, *i.e.* $\tilde{\boldsymbol{\Theta}}$ and $\tilde{\boldsymbol{\pi}} = \{\tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\alpha}}\}$ obtained as explained in Subsection 2.2.3.1;
2. infer $\hat{\boldsymbol{\Theta}} \equiv \{\hat{\Theta}^t\}^{t=1,...T}$ by solving (2.13) with $\tilde{\boldsymbol{\pi}}$;
3. learn $\hat{\boldsymbol{\alpha}}$ by solving (2.22) for each possible couple of nodes with previously inferred $\hat{\boldsymbol{\Theta}}$ and $\tilde{\boldsymbol{\Phi}}$;
4. learn $\hat{\boldsymbol{\Phi}}$ by solving (2.27) for each i with previously inferred $\hat{\boldsymbol{\Theta}}$ and $\hat{\boldsymbol{\alpha}}$;
5. update $\tilde{\boldsymbol{\Theta}} \leftarrow \hat{\boldsymbol{\Theta}}$;
6. update $\tilde{\boldsymbol{\pi}} \leftarrow \hat{\boldsymbol{\pi}}$;
7. repeat up to convergence.

This is in many aspects an Expectation-Maximization learning algorithm (see [Dempster et al., 1977]). In the latent variables inference step (line 2) we also use a generalization of the RAS algorithm [Bacharach, 1965]. The RAS algorithm is usually adopted to solve the problem of estimating nonnegative matrices from marginal data[5] and is preferred to other methods due to its computational speed, numerical stability and algebraic simplicity. In Subsection 2.2.1 we generalize the RAS algorithm to the case of time-varying parameters. The main cycle of the algorithm takes $O(N \times T)$ time. The number of iterations needed for the generalized RAS algorithm to converge is not deterministic, similarly to the original algorithm. However, we observe numerically it takes $O(10^{\log_{10} N-1})$ iterations when $N \in [10^2, 10^3]$. The number of operations needed for learning static paramenters (lines 3 and 4) is, in general, a more complicated question. Learning $\boldsymbol{\alpha}$ takes $O(N^2)$ steps, one for each α_{ij}, and each step takes $T+1$ operations, the numerical evaluation of T single integrals and finding the zero of a function. Learning $\boldsymbol{\Phi}$ takes N steps, one for each $\boldsymbol{\Phi}_i$, but each step takes a non deterministic number of cycles in order to solve the system of integral equations in Eq. 2.27. On average, each step takes $O(T)$ cycles. Each cycle takes $3 \times T$ operations, i.e. the numerical evaluation of $3 \times T$ single integrals. Finally, the number of iterations for the algorithm to converge is not deterministic

[5] The problem in Eq. 2.28 can be solved with the RAS algorithm where the generic entry of the matrix is $\frac{e^{\gamma_i+\gamma_j}}{1+e^{\gamma_i+\gamma_j}}$ and the marginal data are represented by the degree sequence.

but we observe numerically that it is quite constant in the size of the system. Table

	$N = 100$	$N = 250$	$N = 500$	$N = 1000$
time (h)	2.8(5)	11(1)	45(4)	151(12)

Table 2.1 The average time of convergence for the EM algorithm based on MMPA applied to the DAR-TGRG model in the case of undirected networks with $T = 200$. In the model simulations, the parameters are randomly determined as explained in Section 2.3. The simulations were performed using a Matlab code executed on an ordinary dual-core Intel Core i5, with 8 GB RAM. The number in parenthesis is the uncertainty on the last digit.

2.1 shows how much time the EM algorithm takes in average to converge.

2.2.3.3 JMAP approach

For the sake of completeness, we briefly describe the Joint Maximum A Posteriori (JMAP) approach for point estimation of the model in order to compare it with MMAP. JMAP inference consists in the maximization of the joint posterior (2.9) for both $\boldsymbol{\Theta}$ and $\boldsymbol{\pi}$, thus treating the latent variables as the static parameters. By assuming uniform prior, the posterior of both static and time-varying parameters (2.9) for DAR-TGRG reads more explicitly as

$$\mathbb{P}(\boldsymbol{\Theta}, \boldsymbol{\alpha}, \boldsymbol{\Phi}|\boldsymbol{\mathcal{A}}) \propto \mathbb{P}(\boldsymbol{\mathcal{A}}|\boldsymbol{\Theta}, \boldsymbol{\alpha})\mathbb{P}(\boldsymbol{\Theta}|\boldsymbol{\Phi}). \tag{2.29}$$

In a similar fashion of what we propose in Subsection 2.2.1, the inference of the time-varying parameters at time t is obtained by maximizing (2.29) with respect to Θ^t. By conditioning on the expectation for Θ^{t-1}, *i.e.* $\hat{\Theta}^{t-1}$, and the static parameters, this is in effect equivalent to (2.12), thus JMAP is equivalent to MMAP regarding the estimation of the latent variables.

Then, the static parameters are estimated by maximizing

$$\max_{\boldsymbol{\alpha}} \log \mathbb{P}(\boldsymbol{\mathcal{A}}|\boldsymbol{\Theta}, \boldsymbol{\alpha}), \tag{2.30}$$

and

$$\max_{\boldsymbol{\Phi}} \log \mathbb{P}(\boldsymbol{\Theta}|\boldsymbol{\Phi}). \tag{2.31}$$

By conditioning on the expectations for the latent variables, the maximization problem (2.30) corresponds to solve the following equation for α_{ij},

$$\sum_{t=1}^{T} \frac{\delta_{A_{ij}^t, A_{ij}^{t-1}} - \left(1 + e^{-(\hat{\theta}_i^t + \hat{\theta}_j^t)}\right)^{-1}}{\alpha_{ij}\delta_{A_{ij}^t, A_{ij}^{t-1}} + (1 - \alpha_{ij})\left(1 + e^{-(\hat{\theta}_i^t + \hat{\theta}_j^t)}\right)^{-1}} = 0. \tag{2.32}$$

Finally, the maximization problem (2.31) is equivalent to the maximum likelihood estimation of a AR(1) process when we use conditions over the latent variables, *i.e.* we consider the inferred $\hat{\boldsymbol{\Theta}}$ as observable variables in the inference of parameters $\boldsymbol{\Phi}$. See [Tsay, 2005] for details about MLE of the AR(1) model.

Then, we use the following Expectation-Maximization scheme for the JMAP point estimation of the DAR-TGRG model:

1. Assume SSI as starting point, *i.e.* $\tilde{\boldsymbol{\Theta}}$ and $\tilde{\boldsymbol{\pi}} = \{\tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\alpha}}\}$ obtained as explained in Subsection 2.2.3.1;[6]
2. infer $\hat{\boldsymbol{\Theta}} \equiv \{\hat{\Theta}^t\}^{t=1,\dots T}$ by solving (2.13) with $\tilde{\boldsymbol{\pi}}$;
3. learn $\hat{\boldsymbol{\alpha}}$ by solving (2.32) for each possible couple of nodes with previously inferred $\hat{\boldsymbol{\Theta}}$ and $\tilde{\boldsymbol{\Phi}}$;
4. learn $\hat{\boldsymbol{\Phi}}$ by MLE of the N independent AR(1) processes, each one describing one time series $\{\hat{\theta}_i^t\}^{t=0,1,\dots,T}$;
5. update $\tilde{\boldsymbol{\Theta}} \leftarrow \hat{\boldsymbol{\Theta}}$;
6. update $\tilde{\boldsymbol{\pi}} \leftarrow \hat{\boldsymbol{\pi}}$;
7. repeat up to convergence.

2.3 Monte Carlo simulations

Before applying our methodology to real data, we run Monte Carlo simulations to study the performance of the proposed estimation method based on MMAP approach when applied both to undirected and to directed networks. First, we point out that using the EM algorithm based on MMAP approach outperforms the single snapshot inference of the model: in fact naive estimation of the latent dynamics is a pretty common procedure in dynamic network models, thus we aim to quantify how much we earn in adopting a more complex inference procedure. Second, we compare MMAP with both SSI and JMAP in order to show numerically the consistency of the proposed estimation method with respect to other approaches which lead to unbiased estimates of the parameters.

In the simulations, data are generated according to the described models with randomly chosen static parameters. In the case of undirected networks, the DAR(1) model parameters are sampled uniformly in the unit interval. For TGRG we sample $\phi_{1,i} \sim U(-1,1)$, $\sigma_i \sim U(0,1)$ and $\phi_{0,i} \sim \mathcal{N}(0,1)$. For DAR-TGRG, $\alpha_{ij} \sim U(0,1)$. For both models, time-varying parameter θ_i^t follows the stationary AR(1) process (2.2). We estimate the models with the proposed Expectation-Maximization (EM) algorithm based on MMAP and compare the results with both the Single Snapshot Inference (SSI) and JMAP approach. For each simulation, we estimate $\boldsymbol{\Phi}_i$ for each node i. For DAR-TGRG model we obtain also $\binom{N}{2}$ estimates for α_{ij}, one for each possible couple of nodes (i,j). For both models, we infer the time series of the latent

[6] In the SSI inference of $\boldsymbol{\alpha}$, we use (2.32) instead of (2.22), but conditioning on $\tilde{\boldsymbol{\Theta}}$ as described in Subsection 2.2.3.1.

variables $\{\theta_i^t\}_{i=1,\dots,N}^{t=0,1,\dots,T}$. We simulate each model 100 times. In evaluating the goodness of fit of the proposed estimation method, we report the mean absolute relative error for the estimate of parameters. The mean is obtained by averaging over the nodes and the number of simulations. For the time-varying parameters we consider also the time average of the absolute relative errors. A similar study is performed for the case of directed networks, with the exception that for each node we have two fitnesses, $\theta_i^{t,out}$ and $\theta_i^{t,in}$, and as a consequence two sets of static parameters $\boldsymbol{\Phi} \equiv \{\boldsymbol{\Phi}^{out}, \boldsymbol{\Phi}^{in}\}$. For DAR-TGRG model, we obtain $N(N-1)$ estimates for α_{ij}, one for each possible couple of ordered nodes.

	N	θ_i^t	$\phi_{0,i}$	$\phi_{1,i}$	σ_i
SSI	100	0.30	0.58	0.46	0.69
MMAP	100	0.22	0.13	0.13	0.06
SSI	200	0.20	0.31	0.27	0.31
MMAP	200	0.10	0.10	0.10	0.05

Table 2.2 The mean absolute relative error of the estimates of parameters for the TGRG model in the case of undirected networks. We set $T = 200$ and simulate the model 100 times. We compare the proposed Expectation-Maximization (EM) algorithm based on MMAP with the Single Snapshot Inference (SSI).

	N	$\theta_i^{t,out(in)}$	$\phi_{0,i}^{out(in)}$	$\phi_{1,i}^{out(in)}$	$\sigma_i^{out(in)}$
SSI	100	0.31	0.59	0.47	0.71
MMAP	100	0.23	0.12	0.12	0.06
SSI	200	0.21	0.33	0.29	0.33
MMAP	200	0.10	0.11	0.10	0.05

Table 2.3 The mean absolute relative error of the estimates of parameters for the TGRG model in the case of directed networks. We set $T = 200$ and simulate the model 100 times. We compare the proposed Expectation-Maximization (EM) algorithm based on MMAP with the Single Snapshot Inference (SSI).

A comparison between the proposed EM method based on MMAP and the SSI inference is shown in Tables 2.2, 2.3, 2.4 and 2.5, where we show that the EM method greatly outperforms the single snapshot inference SSI. Furthermore, we find that the mean absolute relative error for both EM and SSI declines with the number of nodes N since the number of observations increases as N^2, while the number of parameters increases linearly with N. Furthermore, Table 2.6 shows how the mean absolute relative error of the parameters (estimated by MMAP) of the DAR-TGRG model of a dynamic undirected graph decreases with the length of the time series.

	α_{ij}	θ_i^t	$\phi_{0,i}$	$\phi_{1,i}$	σ_i
SSI	0.22	0.29	0.27	0.18	0.22
MMAP	0.18	0.14	0.15	0.10	0.06

Table 2.4 The mean absolute relative error of the estimates of parameters for the DAR-TGRG model in the case of undirected networks. We compare the EM algorithm based on MMAP with the single snapshot inference SSI. We set $N = 200$, $T = 200$ and simulate the model 100 times.

	α_{ij}	$\theta_i^{t,out(in)}$	$\phi_{0,i}^{out(in)}$	$\phi_{1,i}^{out(in)}$	$\sigma_i^{out(in)}$
SSI	0.22	0.30	0.28	0.18	0.23
MMAP	0.17	0.14	0.14	0.10	0.05

Table 2.5 The mean absolute relative error of the estimates of parameters for the DAR-TGRG model in the case of directed networks. We compare the EM algorithm based on MMAP with the single snapshot inference SSI. We set $N = 200$, $T = 200$ and simulate the model 100 times.

	$T = 300$	$T = 500$	$T = 1000$
θ_i^t	0.13	0.13	0.12
α_{ij}	0.13	0.10	0.08
$\phi_{0,i}$	0.10	0.09	0.07
$\phi_{1,i}$	0.09	0.08	0.07
σ_i	0.05	0.04	0.04

Table 2.6 Mean absolute relative error of the parameters for the DAR-TGRG model estimated via the EM algorithm based on the MMAP approach. The network is undirected. We set $N = 200$ and the number of simulations is equal to 100.

	$T = 100$	$T = 200$	$T = 500$	$T = 1000$
χ_{ij}	0.28	0.21	0.16	0.10
α_{ij}	0.29	0.19	0.12	0.08

Table 2.7 The mean absolute relative error of the estimates of parameters of the DAR(1) model as a function of the length T of time series. We simulate the DAR(1) model 100 times.

For completeness, in Table 2.7 we show the results for the maximum likelihood estimation of the DAR(1) process (1.60).

Hence, we compare the three different estimation methods in order to show numerically the consistency of the MMAP approach. According to Monte Carlo simulations of the model DAR-TGRG, the three estimation methods are quite equivalent in inferring the set of parameters $\boldsymbol{\alpha}$ (when sampled uniformly in the unit interaval), see Figure 2.1. Furthermore, we show in Figure 2.2 the robustness of the EM algorithm (based on MMAP approach) to the choice of the prior distribution of static

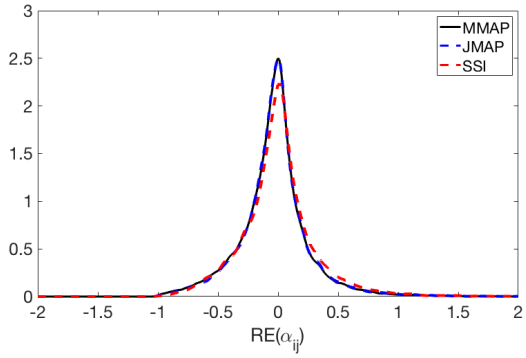


Fig. 2.1 Distribution of relative errors *RE* for parameters α_{ij}, inferred according to MMAP (black solid line), JMAP (blue dashed line), and SSI (red dotted line). Data are simulated according to DAR-TGRG model with $N = 300$ nodes, $T = 200$, and model parameters sampled uniformly in the domain of definition.

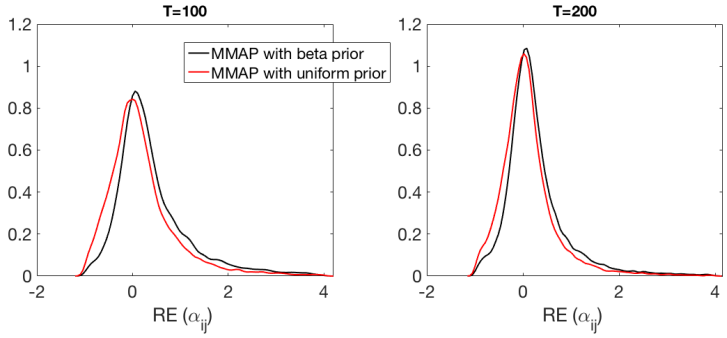


Fig. 2.2 Distribution of relative errors *RE* for α_{ij} inferred according to the MMAP approach, by specifying the correct prior distribution (black line) or by assuming uniform prior (red line). Data are simulated according to DAR-TGRG model with Beta distribution ($Beta(\alpha = 2, \beta = 5)$) for α_{ij} in the case of $N = 200$ nodes, $T = 100$ (left panel) or $T = 200$ (right panel).

parameters, *i.e.* $\mathbb{P}(\boldsymbol{\alpha})$ for this specific case, as long as it is a smooth distribution. In fact, the contribution of the prior to the estimate is vanishing asymptotically as the sample size T increases (this strongly depends on the static nature of the parameters that are not time-varying as explained in Section 2.2).

However, the three estimation methods differ significantly in the estimation of the latent dynamics. In Figure 2.3, we show a typical example of the inferred fitness according to both JMAP and MMAP, and compared with data simulated according to DAR-TGRG. Notice that the JMAP method underestimates largely the variability of data.

In Figure 2.4, we show the relative error in the inference of parameters $\Phi \equiv \{\phi_{0,i}, \phi_{1,i}, \sigma_i\}_{i=1,\dots,N}$ of the latent dynamics of both TGRG (left panels) and DAR-TGRG (right panels) models for the three estimation methods, *i.e.* SSI, MMAP, and

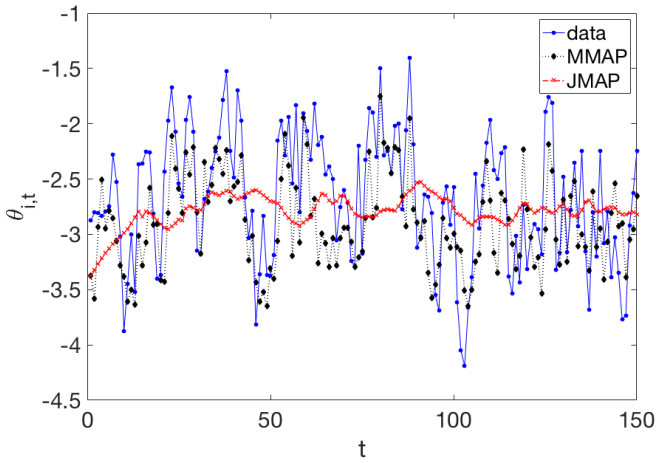


Fig. 2.3 An example of the fitness dynamics following a AR(1) process for a generic fitness θ_i^t generated according to the DAR-TGRG model (blue dots). Black dots represent the fitness inferred according to MMAP, while red crosses represent the fitness inferred according to JMAP.

JMAP. From numerical simulations, we can notice that both SSI and JMAP lead to a bias in the inference of parameters, whereas MMAP outperforms the others:

1. in the case of DAR-TGRG, both SSI and JMAP infer $\boldsymbol{\Phi}$ by solving (2.31), without using explicitly the information about link persistence. On the contrary, MMAP exploits this information, specifically the information carried by each link (i, j) is weighted with its memory parameter α_{ij} , see (2.25). Persistence of links incident to node i tends to increase the persistence of the node degree. If the estimation method is not able to disentangle the origin of persistence, *i.e.* link persistence of the Markovian behavior for the edges or degree persistence because of the node fitness autocorrelated in time, the result is an overestimation of the autoregressive coefficient $\phi_{1,i}$ associated with the node fitness θ_i, as it happens for both SSI and JMAP (see the middle right panel of Figure 2.4). The overestimation of fitness persistence is related to the underestimated variability of the inferred latent dynamics of Figure 2.3.
   As a consequence of the overestimation of $\phi_{1,i}$, the parameter $\phi_{0,i}$ is underestimated in order to preserve the overall mean of the latent autoregressive dynamics AR(1) (*i.e.* $\phi_{0,i}/(1-\phi_{1,i})$), see the top right panel of Figure 2.4.
   Finally, since the variance of a AR(1) process is $\sigma_i^2/(1-\phi_{1,i}^2)$, an overestimation of $\phi_{1,i}$ leads to an underestimation of σ_i (once the observed variance is fixed). This is what we observe for the naive estimation method SSI, see the red dotted line in the bottom right panel of Figure 2.4. For JMAP, this behavior is even more evident because of a feedback effect in the algorithm. In fact, the maximization problem (2.12) for the inference of the latent variables is

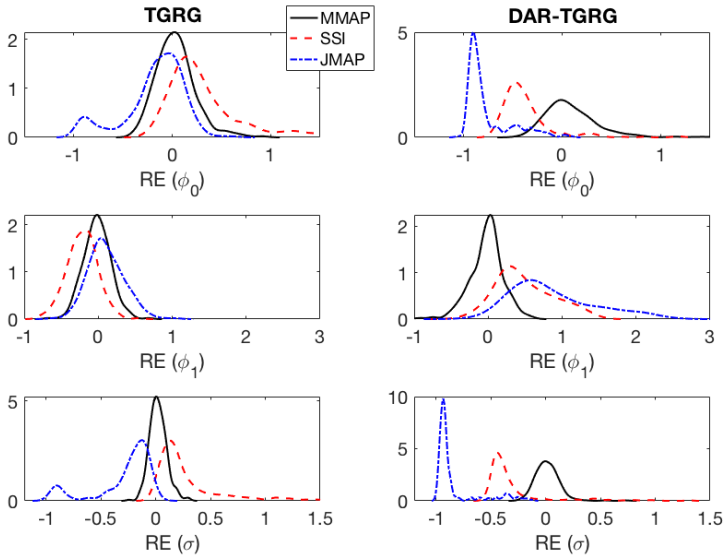


Fig. 2.4 Distribution of relative errors RE in the inference of parameters $\Phi_i = \{\phi_{0,i}, \phi_{1,i}, \sigma_i\}$ of the latent AR(1) dynamics for both TGRG (left panels) and DAR-TGRG (right panels) models according to the three estimation methods: MMAP (black), SSI (red), and JMAP (blue). Both TGRG and DAR-TGRG models are simulated by sampling the static parameters uniformly in the domain. The convergence conditions are the same for both MMAP and JMAP expectation-maximization schemes.

$$\frac{\partial \log \mathbb{P}(A^t|A^{t-1}, \boldsymbol{\alpha}, \boldsymbol{\Theta}^t)}{\partial \theta_i^t} - \frac{\theta_i^t - \phi_{0,i} - \phi_{1,i}\theta_i^{t-1}}{\sigma_i^2} = 0, \ \forall t = 1, ..., T, \tag{2.33}$$

thus an underestimation of σ_i in the M-step has the consequence of weighting less the information about the network and more the (biased) information about the latent dynamics in the previous step, resulting in a further amplification of the bias.

2. in the case of TGRG (left panels of Figure 2.4), the bias of both SSI and JMAP estimators is less evident. Nevertheless, it is still present and by comparing the performances of the three methods, MAP outperforms the others.

Finally, we point out that a model for link stability not accounting for time-varying node fitness does not capture correctly the patterns of link persistence. In fact, when the dynamics of the link is affected both by link persistence and by dynamic fitness, neglecting the last one can lead to an overestimation of the importance of the persistence. To show this we simulate a DAR-TGRG model for undirected networks taking $\phi_{1,i}$ equal for all time-varying parameters θ_i^t[7]. Then we estimate

[7] $\phi_{1,i}$ determines the autocorrelation of node fitness and as a consequence the link persistence associated with the time-varying marginal.

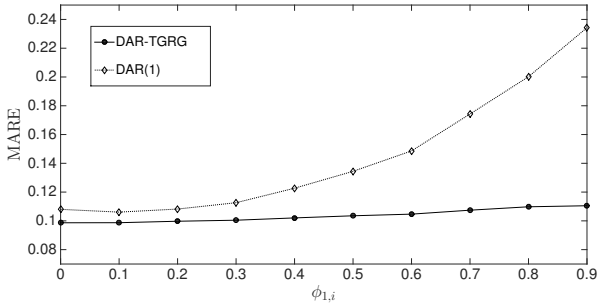


Fig. 2.5 Mean Absolute Relative Error (MARE) of the estimates of α_{ij} as a function of the autoregressive coefficient $\phi_{1,i}$ for the time-varying parameters θ_i^t. In the simulation of the DAR-TGRG model for the case of undirected networks, α_{ij}, $\phi_{0,i}$ and σ_i are randomly sampled while $\phi_{1,i}$ are equal for all i. We compare the goodness of fit of the estimates of α_{ij} via the EM method for the DAR-TGRG model (solid line) with the Maximum Likelihood (ML) estimates of α_{ij} according to the DAR(1) model (dotted line). We set $N = 200$ and $T = 400$.

α_{ij} according to a DAR(1) model (which neglects dynamic fitness) and to a DAR-TGRG model. Figure 2.5 shows the mean absolute error of α_{ij} for the two estimations as a function of $\phi_{1,i}$. When this parameter is small both the DAR(1) model and the DAR-TGRG model perform quite equivalently. On the contrary, when the dynamic fitness has a significant persistence due to a high value of $\phi_{1,i}$, the DAR(1) model wrongly imputes this to a link persistence which now has a large bias with respect to the DAR-TGRG model.

2.4 Empirical application: understanding link persistence in the interbank market

Trading and credit networks are a natural application case for dynamic networks with persistence, like the one described by our model. Financial institutions lend mutually money on a daily basis and interbank markets are considered an important channel of propagation of systemic risk. While there is a vast literature on the static case, only few papers deal with the dynamic property of interbank networks. The static fitness model has been proved to characterize quantitatively several topological properties of the e-MID network [De Masi et al., 2006, Musmeci et al., 2013], to outperform other network models in the problem of reconstructing the e-MID network from limited information [Gabrielli et al., 2014, Mazzarisi and Lillo, 2017] and to give useful insights for systemic risk analysis of the interbank market [Cimini et al., 2015]. The ability of the fitness model to describe the static interbank network indicates that the size of two banks correlates with the existence of a credit between them. However it has been documented [Iori et al., 2015] the presence of memory effects in the process of network formation for interbank markets,

according to the idea that a borrower, having asked for a loan many times to a lender in the past, is more likely to borrow from that lender again in the future than from other lenders, with which the borrower has never (or infrequently) interacted, see also [Di Maggio et al., 2017]. [Finger and Lux, 2017] considered a similar problem in a behavioral model of network formation, finding that link persistence is very important in e-MID.

In this section we estimate our dynamic model on data of an interbank market to disentangle the relative importance of fitness and link persistence in determining the future state of the network. This will allow also to perform a forecasting exercise to predict the existence of a credit relation between two banks.

2.4.1 Data

We investigate data from the electronic Market of Interbank Deposit (e-MID), a market where banks extend loans to one another for a specified term and/or collateral. A significant fraction of interbank loans are for maturities of one week or less, the majority being overnight. The e-MID is an electronic market in the Euro Area and it was founded in Italy in 1990 for Italian Lira transactions and denominated in Euros in 1999. According to the "Euro Money Market Study 2006" published by the European Central Bank in February 2007, e-MID accounted for 17% of total turnover in unsecured money market in the Euro Area. More recently the amount of overnight lending in e-MID has significantly declined, especially around the sovereign debt crisis, see [Barucca and Lillo, 2018]. The e-MID network has been thoroughly studied to understand bank liquidity management, as for instance in [Iori et al., 2008, Finger et al., 2013].

The dataset contains the edge list of all credit transactions in each day from March 9^{th}, 2012 to February 27^{th}, 2015. In our analysis, we investigate the interbank network aggregated weekly. Each network snapshot of interbank deposits is constructed from the list of transactions where a bank, the lender, extends a loan to another bank, the borrower, that repays the loan in seven days, at most. Hence, we exclude loans with a term larger than a week. However, we account approximately for the 92% of all the traded volume in the market since there are few credit relations with longer maturity. Then, we describe the e-MID weekly network with the unweighted and directed adjacency matrix $\boldsymbol{A}^t$: a generic element A^t_{ij} is 1 if the bank i lends money at least once to bank j during the week t, 0 otherwise. We do not consider banks that interact less than 5% of times in the considered period, i.e. in a period of $T = 156$ weeks a bank has at least a credit relation for more than 7 weeks. Hence, the credit network is formed by $N = 98$ banks.

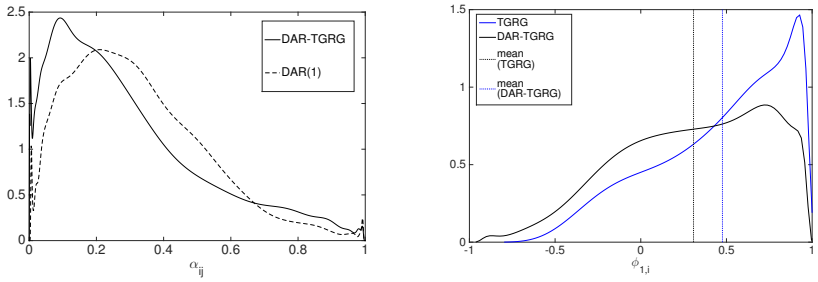


Fig. 2.6 Left panel: distributions of α_{ij} estimated via EM on e-MID data. The solid line refers to the DAR-TGRG model while the dotted line refers to the DAR(1) model. Right panel: distribution of parameters $\phi_{1,i}$ estimated via EM. The black line refers to DAR-TGRG while the blue line to TGRG. The dotted lines represent the mean of the two distributions.

2.4.2 Estimated fitness and link persistence in e-MID

We estimate the three models on the time series of e-MID networks. More specifically, we consider maximum likelihood estimation of the DAR(1) model and marginal maximum a posteriori inference of TGRG and DAR-TGRG. Figure 2.6 shows the estimated density of the α_{ij} link parameters (left panel) and of the $\phi_{1,j}$ node parameters (right panel) for the different model. We see that the DAR(1) model estimates larger α_{ij} parameters, i.e. larger link persistence, than the DAR-TGRG model. Similarly, the TGRG model estimates larger $\phi_{1,j}$, i.e larger fitness persistence, than the complete DAR-TGRG model. Thus the full model balances the relative role of the two persistence mechanisms.

Node fitness is a latent variable whose time evolution is not observed but inferred according to models of temporal networks. However it is interesting to ask if there exists an observable quantity correlated with it. We show that for the considered dataset, node fitness is correlated with the bank exposure in the e-MID interbank market[8].

In fact we observe that the quantity $x_i^{t,out(in)} \equiv e^{\theta_i^{t,out(in)}} \quad \forall i = 1,...,N$ estimated on data for both TGRG and DAR-TGRG models is strongly correlated with the corresponding bank's exposure in e-MID for the considered week t, see the top left panel in Figure 2.7. This result suggests that, at a given time, banks with larger exposures are the nodes with larger fitnesses $\theta_i^{t,out(in)}$ or equivalently with larger degrees. Furthermore, the time-varying fitness of a node is correlated significantly with its bank exposure (see the top right panel of Figure 2.7). Finally, in the bottom panel of Figure 2.7 we show an example of this behavior for node 3 whose correlation coefficient is ~ 0.90. Thus the dynamic fitness model can be seen as a procedure

[8] Exposure of bank i is defined as the strength of node i in the weighted network. We refer to it as $s_i^{t,out(in)}$ for generic node i. The node out-strength corresponds to the bank asset exposure in e-MID while the node in-strength to the liability.

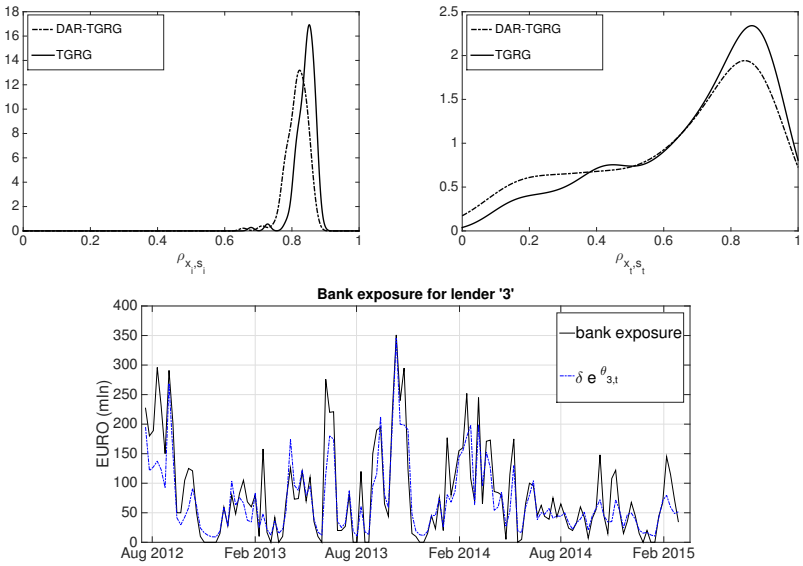


Fig. 2.7 Density estimation of cross sectional (top left panel) and temporal (top right panel) Spearman correlation between the inferred $x_i^{t,out(in)} \equiv e^{\theta_i^{t,out(in)}}$ and the corresponded bank exposure s_i^t in e-MID. Bottom panel: an example of time-varying fitness compared with the bank exposure for node '3'. The parameter δ is chosen in such a way that the maxima of the two time series correspond.

allowing to have some insights on bank exposures having only information on the binary network.

2.4.3 Link stability and preferential trading in e-MID

For credit networks such as e-MID, the preferential linkage mechanism reflects the presence of banks which trade preferentially each others. Preferential trading between banks can be detected by comparing empirically observed trading relationships with a null hypothesis that assumes random trading. [Hatzopoulos et al., 2015] have introduced a statistical test to assess the statistical significance of the observed interbank transactions in order to reveal preferential credit relationships among banks. We apply the same statistical test to show that preferential trading relations in e-MID are associated with link stability.

Following [Hatzopoulos et al., 2015], we apply the test to the weekly aggregated e-MID data split in time windows of 3-maintenance periods[9] In each time window

[9] The period of time in which credit institutions have to comply with the minimum reserve requirements is called the reserve maintenance period. Each reserve maintenance period is equivalent to

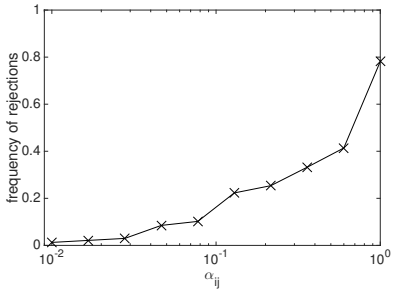


Fig. 2.8 Fraction of statistically validated links according to the test in [Hatzopoulos et al., 2015] conditional to the value of the estimated α_{ij} parameter measuring the link persistence in the TGRG-DAR model.

and for each link (i, j) we count the number of times n_{ij}^{lb} bank j borrowed money from bank i. Then, let n_i^l be the number of times bank i lent money to any other banks and let n_j^b be the number of times bank j borrowed money from any other bank. Finally, let us define N_T as the total number of trades among banks in the system for the considered 3-maintenance period. Under the null hypothesis of random trading, n_{ij}^{lb} follows the hypergeometric distribution $H\left(n_{ij}^{lb}|n_i^l, n_j^b, N_T\right)$. Hatzopoulos et al. associate preferential trading with over-expressed number of links with respect to the null hypothesis of random trading, i.e. they use the hypergeometric distribution to associate a p-value with the observed number n_{ij}^{lb}. Preferential trading relations $i \to j$ are the ones rejected according to the statistical test, i.e. with a p-value smaller than the threshold value $\frac{0.05}{a}$ where a is the Bonferroni correction to avoid a large number of false positive validated links because of the multiple hypothesis testing, see [Hatzopoulos et al., 2015] for more details.

Figure 2.8 shows the frequency of rejection for the statistical test by [Hatzopoulos et al., 2015] conditional to the estimated α_{ij} parameter measuring the link persistence in the TGRG-DAR model. The clear monotonic behavior indicates that link stability is statistically associated with preferential trading detected according to [Hatzopoulos et al., 2015].

2.4.4 Link prediction

Finally we compare the proposed network models in their out-of-sample link forecasting performance. We use the first 106 weekly network observations for model estimation and the last 50 as our out-of-sample period. In the training phase we estimate the static parameters for the three models and then we adopt the following forecast scheme based on on-line inference. Rolling over the out-of-sample period, at each week t we use the new observed snapshot $\boldsymbol{A}^t$ to infer the expected $\hat{\boldsymbol{\Theta}}^t$ via

one calendar month and we aggregate the maintenance periods in groups of three. Hence, we consider 12 3-maintenance periods ranging from April 2^{nd}, 2012 to February 27^{th}, 2015.

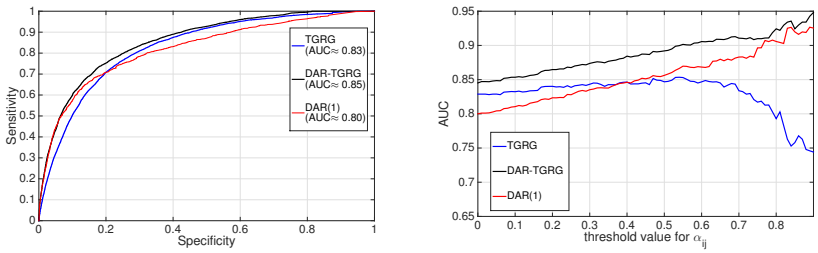


Fig. 2.9 Left panel: ROC curve drawn according to the out-of-sample forecasting exercise: TGRG (blue line), DAR-TGRG (black line) and DAR(1) (red line). Right panel: Area Under the Curve (AUC) of the three models as a function of the threshold value for α_{ij} inferred according to DAR-TGRG.

Eq. 2.12. Then, for DAR-TGRG model we produce the one-step-ahead forecast for each link as

$$\mathbb{E}[A_{ij}^{t+1}|A_{ij}^{t},\hat{\theta}_i^t,\hat{\theta}_j^t] = \int d\theta_i^{t+1} d\theta_j^{t+1} \mathbb{P}[A_{ij}^{t+1}=1|A_{ij}^{t},\theta_i^{t+1},\theta_j^{t+1}] n(\theta_i^{t+1}|\hat{\theta}_i^t) n(\theta_j^{t+1}|\hat{\theta}_j^t) =$$
$$= \alpha_{ij}A_{ij}^t + (1-\alpha_{ij}) \int_0^\infty \frac{d\omega}{2} p_{PG}(\omega) \frac{e^{\frac{-4\omega(\phi_{0,i}+\phi_{1,i}\hat{\theta}_i^t+\phi_{0,j}+\phi_{1,j}\hat{\theta}_j^t)^2+(\sigma_i^2+\sigma_j^2)+4(\phi_{0,i}+\phi_{1,i}\hat{\theta}_i^t+\phi_{0,j}+\phi_{1,j}\hat{\theta}_j^t)}{8(1+\omega(\sigma_i^2+\sigma_j^2))}}}{\sqrt{1+\omega(\sigma_i^2+\sigma_j^2)}}, \tag{2.34}$$

where we have applied the result of [Polson et al., 2013] as before. The one-step-ahead forecast for the TGRG model is simply obtained by putting α_{ij} equal to 0 in Eq. 2.34. The one-step-ahead forecast for DAR(1) model is a standard result of time series analysis given by

$$\mathbb{E}[A_{ij}^{t+1}|A_{ij}^t] = \alpha_{ij}A_{ij}^t + (1-\alpha_{ij})\chi_{ij}. \tag{2.35}$$

By comparing the forecasting performances of the models, we restrict the analysis to those links which are not always zero in the out-of-sample period. Finally, the forecasting perfomance of each model is assessed by constructing the Receiving Operating Characteristic (ROC) curve (see [Friedman et al., 2001] for the definition of ROC curve).

The results are summarized in Figure 2.9. In the left plot, we compare the three ROC curves and we can notice how the DAR-TGRG model (slightly) outperforms the other models. Furthermore, in the right plot we show the area under the curve (AUC) as a function of a threshold for $\hat{\alpha}_{ij}$ estimated according to DAR-TGRG model. In other words, we compare the AUC considering only the links for which the $\hat{\alpha}_{ij}$ estimated by the the DAR-TGRG model is larger than a threshold value. We find that taking into account both of fitness dynamics and preferential linkage better forecast links, i.e. DAR-TGRG always outperforms the other models. When we consider links with both high and low persistence, the TGRG model outper-

forms the DAR(1) network model, that is the evolution of the network topology is more important than preferential linkage in determining the average characteristics of the e-MID network. However, the link copying mechanism associated with the DAR(1) model characterizes better than the fitness dynamics the persistence pattern associated with a smaller set of links representing the preferential relations among banks. In fact, there exists a value of the threshold (around 0.4) after which the AUC associated with the DAR(1) model is larger than the one for TGRG.

Conclusions

In this chapter, we have introduced a novel state-of-the-art statistical methodology to describe link persistence and fitness dynamics in temporal networks. We have modeled a Markov dynamics for both observed and unobserved time-varying states which drive the evolution of the network. The analytic tractability of the autoregressive network ensemble we have proposed allows us to easily calibrate our parameters from the data with a general likelihood maximization iterative procedure. The introduction of the autoregressive dynamics permits link forecasting by taking account of memory properties of the network system. Then, the estimation method we have introduced allows online-inference of the time-varying parameters which is particularly useful from a computational point of view to face the problem of link prediction.

The contribution of this work is twofold. First, the introduction of autoregressive endogenous components displays the clear advantage of describing the network evolution via time-varying states which reproduce the network topology as well as capturing the local property of link persistence, thus going beyond a single snapshot analysis where parameters are chosen for each network snapshot, independently. Second, the analysis on real data from the eMID interbank network from 2012 to 2015 (weekly aggregated) displays the statistical equivalence between link stability, identified by positive value of the persistence parameter, and preferential trading, identified by over-expressed number of trades between counterparties. Hence, our methodology permits to disentangle preferential trading from random trading in dynamic trading networks such as the eMID money market. Finally, the forecasting performance of the model points out both fitness dynamics and link persistence as linkage mechanisms in the process of network formation for the credit market.

As future outlooks, the formalism discussed could also be applied to more general memory kernel function of the autoregressive model governing the evolution of the system as well as it could permit the introduction of exogenous factors driving the fitness dynamics or the local link probability. Furthermore, a challenging issue is the introduction of a dependence structure for the dynamic fitnesses. We also note that the estimation method we have introduced to obtain our results for dynamic networks is quite general, and could be used to obtain similar results for other types of fitness dynamics.

Appendix: Two-point probability mass function for TGRG

In the TGRG model, the node fitness is autocorrelated in time when $\phi_{i,1} \neq 0$. An autocorrelated fitness reflects the autocorrelation of the degree and ultimately of all links incident to the node[10]. A positive autocorrelated fitness is associated with persistence of links. This effect can be characterized by studying the two-point distribution function or equivalently the autocorrelation function. In the TGRG model we can compute semi-analytically the the two-point distribution function,

$$\begin{aligned}
&\mathbb{P}(A_{ij}^t = 1, A_{ij}^{t-\tau} = 1) = \\
&= \int d\theta_i^t d\theta_j^t d\theta_i^{t-\tau} d\theta_j^{t-\tau} \mathbb{P}(A_{ij}^t = 1|\theta_i^t, \theta_j^t)\mathbb{P}(A_{ij}^{t-\tau} = 1|\theta_i^{t-\tau}, \theta_j^{t-\tau}) p(\theta_i^t, \theta_i^{t-\tau}) p(\theta_j^t, \theta_j^{t-\tau}) = \\
&= \int d\theta_i^{t-\tau} d\theta_j^{t-\tau} \frac{1}{1+e^{-(\theta_i^{t-\tau}+\theta_j^{t-\tau})}} n(\theta_i^{t-\tau}) n(\theta_j^{t-\tau}) \times \\
&\times \int \left[\prod_{q=1}^{\tau-1} \prod_{a=i,j} f(\theta_a^{t-\tau+q}|\theta_a^{t-\tau+(q-1)}) d\theta_a^{t-\tau+q} \right] \int d\theta_i^t d\theta_j^t \frac{1}{1+e^{-(\theta_i^t+\theta_j^t)}} f(\theta_i^t|\theta_i^{t-1}) f(\theta_j^t|\theta_j^{t-1}) = \\
&= \int d\theta_i^{t-\tau} d\theta_j^{t-\tau} \frac{1}{1+e^{-(\theta_i^{t-\tau}+\theta_j^{t-\tau})}} n(\theta_i^{t-\tau}) n(\theta_j^{t-\tau}) \int_0^\infty \frac{d\omega}{2} p_{PG}(\omega) K^\tau(\omega|\theta_i^{t-\tau}, \theta_j^{t-\tau})
\end{aligned} \tag{2.36}$$

where we have applied the result of [Polson et al., 2013] as before and

$$K^\tau(\omega|\theta_i^{t-\tau}, \theta_j^{t-\tau}) = \frac{e^{\frac{-4\omega(\mu_i^\tau+\mu_j^\tau)^2+((\sigma_i^\tau)^2+(\sigma_j^\tau)^2)+4(\mu_i^\tau+\mu_j^\tau)}{8(1+\omega((\sigma_i^\tau)^2+(\sigma_j^\tau)^2))}}}{\sqrt{1+\omega((\sigma_i^\tau)^2+(\sigma_j^\tau)^2)}}$$

with

$$\mu_a^\tau = \phi_{0,a}\left(\sum_{t=0}^{\tau-1}(\phi_{1,a})^t\right) + (\phi_{1,a})^\tau \theta_a^{t-\tau} \;\; a = i, j$$

$$(\sigma_a^\tau)^2 = \sigma_a^2 (\sum_{t=0}^{\tau-1}(\phi_{1,a}^2)^t) \;\; a = i, j.$$

The last recursive formulas are obtained by integrating over the Gaussian transition probabilities in Eq. 2.36. Let us notice that μ_a^τ and $(\sigma_a^\tau)^2$ converge to the mean and the variance of the marginal distribution for θ_a^t in the limit $\tau \to \infty$ as we can expect for the standard AR(1) process.

Then, the two-point distribution function can be obtained by integrating over the Gaussian marginals, i.e. $n(\theta_i^{t-\tau})$ and $n(\theta_j^{t-\tau})$ with $\tilde{\mu}_a \equiv \frac{\phi_{0,a}}{1-\phi_{1,a}}$ and $\tilde{\sigma}_a^2 \equiv \frac{\sigma_a^2}{1-\phi_{1,a}^2}$ $a = i, j$ as the mean and the variance of the Gaussian marginal distribution for $\theta_a^{t-\tau}$, and finally by performing the numerical integration over the probability density function associated with the Polya-Gamma distribution. It is

[10] This is always true in the case of finite network size.

$$\mathbb{P}(A_{ij}^t=1,A_{ij}^{t-\tau}=1)=\int_0^\infty \frac{d\omega}{2}p_{PG}(\omega)\int_0^\infty \frac{d\zeta}{2}p_{PG}(\zeta)\times$$
$$\times \frac{e^{g(\omega,\zeta,\phi_{0,i},\phi_{0,j},\phi_{1,i},\phi_{1,j},\sigma_i,\sigma_j)}}{\sqrt{1+\zeta(\tilde{\sigma}_i^2+\tilde{\sigma}_j^2)+\omega(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2+(B_i^\tau)^2\tilde{\sigma}_i^2+(B_j^\tau)^2\tilde{\sigma}_j^2)+\zeta\omega(\tilde{\sigma}_i^2(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2)+\tilde{\sigma}_j^2(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2)+\tilde{\sigma}_i^2\tilde{\sigma}_j^2(B_i^\tau-B_j^\tau)^2)}} \tag{2.37}$$

where

$$g(\omega,\zeta,\boldsymbol{\Phi}_i,\boldsymbol{\Phi}_j)=\frac{h(\omega,\zeta,\boldsymbol{\Phi}_i,\boldsymbol{\Phi}_j)}{8\left(1+\zeta(\tilde{\sigma}_i^2+\tilde{\sigma}_j^2)+\omega(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2+(B_i^\tau)^2\tilde{\sigma}_i^2+(B_j^\tau)^2\tilde{\sigma}_j^2)+\zeta\omega(\tilde{\sigma}_i^2(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2)+\tilde{\sigma}_j^2(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2)+\tilde{\sigma}_i^2\tilde{\sigma}_j^2(B_i^\tau-B_j^\tau)^2)\right)}$$

and

$$\begin{aligned}h(\omega,\zeta,\boldsymbol{\Phi}_i,\boldsymbol{\Phi}_j)=&4(T_i^\tau+T_j^\tau)+4(1+B_i^\tau)\tilde{\mu}_i+4(1+B_j^\tau)\tilde{\mu}_j+C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2+(1+B_i^\tau)^2\tilde{\sigma}_i^2+(1+B_j^\tau)^2\tilde{\sigma}_j^2+\\
&+\zeta\ (\tilde{\sigma}_i^2\tilde{\sigma}_j^2(B_i^\tau-B_j^\tau)^2-4(\tilde{\mu}_i+\tilde{\mu}_j)^2+\\
&\quad+\tilde{\sigma}_i^2\ (4T_i^\tau+4T_j^\tau+4\tilde{\mu}_j(B_j^\tau-B_i^\tau)+C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2)+\\
&\quad+\tilde{\sigma}_j^2(4T_i^\tau+4T_j^\tau+4\tilde{\mu}_i(B_i^\tau-B_j^\tau)+C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2))+\\
&-\omega\ (4(T_i^\tau+T_j^\tau)^2+4(B_i^\tau\tilde{\mu}_i+B_j^\tau\tilde{\mu}_j)(2T_i^\tau+2T_j^\tau+B_i^\tau\tilde{\mu}_i+B_j^\tau\tilde{\mu}_j)+4(\tilde{\mu}_i+\tilde{\mu}_j)(-1+\zeta(\tilde{\mu}_i+\tilde{\mu}_j))(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2)+\\
&\quad+\tilde{\sigma}_i^2(4(T_i^\tau+T_j^\tau+(B_j^\tau-B_i^\tau)\tilde{\mu}_j)(B_i^\tau(1-\zeta\tilde{\mu}_j)+\zeta(T_i^\tau+T_j^\tau+B_j^\tau\tilde{\mu}_j))-(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2))+\\
&\quad+\tilde{\sigma}_j^2(4(T_i^\tau+T_j^\tau+(B_i^\tau-B_j^\tau)\tilde{\mu}_i)(B_j^\tau(1-\zeta\tilde{\mu}_i)+\zeta(T_i^\tau+T_j^\tau+B_i^\tau\tilde{\mu}_i))-(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2))+\\
&\quad-\tilde{\sigma}_i^2\tilde{\sigma}_j^2(B_i^\tau-B_j^\tau)^2)\end{aligned}$$

where we have defined for notational simplicity

$$T_a^\tau\equiv\phi_{0,a}\left(\sum_{t=0}^{\tau-1}(\phi_{1,a})^t\right),\ B_a^\tau\equiv(\phi_{1,a})^\tau,\ C_a^\tau\equiv(\sum_{t=0}^{\tau-1}(\phi_{1,a}^2)^t)\ \ a=i,j.$$

Finally, the ACF can be obtained by noticing that $\mathbb{E}[A_{ij}^tA_{ij}^{t-\tau}]\equiv\mathbb{P}(A_{ij}^t=1,A_{ij}^{t-\tau}=1)$ and the unconditional expectation for A_{ij}^t is

$$\begin{aligned}\mathbb{E}[A_{ij}^t]&=\int d\theta_{i,t}d\theta_{j,t}\mathbb{P}(A_{ij}^t=1|\theta_{i,t},\theta_{j,t})n(\theta_{i,t})n(\theta_{j,t})=\\
&=\int\frac{d\omega}{2}p_{PG}(\omega)\frac{e^{\frac{-4\omega(\bar{\mu}_i+\bar{\mu}_j)^2+(\tilde{\sigma}_i^2+\tilde{\sigma}_j^2)+4(\bar{\mu}_i+\bar{\mu}_j)}{8(1+\omega(\tilde{\sigma}_i^2+\tilde{\sigma}_j^2))}}}{\sqrt{1+\omega(\tilde{\sigma}_i^2+\tilde{\sigma}_j^2)}}.\end{aligned} \tag{2.38}$$

Chapter 3
Community detection in directed stochastic block models

Abstract In this chapter, we study the problem of community detectability in the difficult cases of the inference of stochastic block models when link direction is considered. We therefore identify what the macroscopic structures of interest for the problem are, thus generalizing the results obtained for undirected networks. To this end, building on the stochastic block model, we construct a class of hardly detectable directed networks. We find closed-form solutions by using belief propagation method showing how the transition line depends on the assortativity and the asymmetry of the network. Finally, we numerically identify the existence of a hard phase for detection close to the transition point.
Almost all results in this chapter previously appeared in [Wilinski et al., 2019].

Introduction

Simple undirected graphs are often just an approximation of real-world networks. Neglecting weights, multiple links or self loops may be justified in some cases but also implies a loss of information. An important generalization of a simple graph is a directed network. Community detection in such networks is already an active field [Malliaros and Vazirgiannis, 2013].

There is, however, no work, at least to our best knowledge, that describes the impact of introducing directions on the detectability phase transition. In this chapter we fill this gap by introducing belief propagation technique [Mézard and Parisi, 2001, Decelle et al., 2011a] to the directed case of the stochastic block model [Holland et al., 1983, Wang and Wong, 1987].

As in the standard symmetric case, *i.e.* when the affinity matrix describing the link probability between groups is symmetric because of no link direction, we first identify an entire class of difficult cases for community detection, where both the average in- and out-degree are the same across all groups. We characterize the phase transition associated with the detectability problem, thus identifying both first (*i.e.* hard) and second (*i.e.* easy) order transitions. Then, we concentrate on the case

of many groups, where varying the assortativity leads to a hard detectable phase. We show how the asymmetry affects the size of this hard phase and the existence of discontinuity. Finally, as expected, we show that by introducing asymmetry in the affinity matrix, we are able to significantly increase the range of the detectable phase.

This chapter is organized as follows: in Section 3.1 we define the stochastic block model when link direction is considered, then in Section 3.2 we study the inference of the model and point out what are the difficult cases of detectability which reveal the macroscopic structures of interest for the problem. Finally, in Section 3.3 we study the phase transitions associated with the problem of detectability threshold and in the last section we conclude.

3.1 Directed Stochastic Block Models (DiSBM)

Let us consider a network $G = (V,E)$ with $|V| = N$ nodes. Links between nodes can be described by the $N \times N$ adjacency matrix $\boldsymbol{A}$ whose generic element A_{ij} takes value 1 if a link goes from i to j, 0 otherwise. Since the network is directed, each couple of nodes may share two links, one for each direction, and, in general, the adjacency matrix is not symmetric.

A random directed graph is generated according to DiSBM as follows. Each node i has a label $g_i \in \{1,2,...,k\}$ indicating which group it belongs to among k possibilities. Node labels are chosen independently, where for each node i the probability that $g_i = a$ is q_a (normalized such that $\sum_{a=1}^{k} q_a = 1$). Once the groups are defined, for each pair of ordered nodes (i, j), we put a link from i to j ($A_{ij} = 1$) with probability $p_{g_i g_j}$, where p_{ab} are the entries of the $k \times k$ affinity matrix $\boldsymbol{p}$. Differently from the case of simple undirected graphs where the affinity matrix is symmetric, here $\boldsymbol{p}$ is in general not symmetric because of the introduction of link direction. Then, we are interested in the sparse regime where, *i.e.* $p_{ab} = O(1/N)$. In this case, let us define the rescaled affinity matrix $\boldsymbol{c}$ with $c_{ab} \equiv N p_{ab}$.

The total average degree of the graph is then $c = 2\sum_{ab} c_{ab} q_a q_b$, whereas the total average degree of a generic node in the group a is $c_a = \sum_b (c_{ab} + c_{ba}) q_b$, with $c_a^{out} = \sum_b c_{ab} q_b$ and $c_a^{in} = \sum_b c_{ba} q_b$ being respectively the average out- and in-degree in the same group.

Let N_a be the number of nodes in the group a and M_{ab} the number of links from the group a to the group b for a specific graph realization. Then $\lim_{N\to\infty} N_a/N = q_a$ and $\lim_{N\to\infty} NM_{ab}/N_a N_b = c_{ab}$.

3.2 Inference of DiSBM

In the detection setting, the adjacency matrix $\boldsymbol{A}$ is the only information available and the goal is to infer the true group assignments $\boldsymbol{g} = \{g_i\}_{i=1,\dots,N}$ together with the parameters $\boldsymbol{\pi} = (\boldsymbol{q}, \boldsymbol{c})$ for a given number k of groups.

Similarly to the undirected case, the posterior probability of $\boldsymbol{\pi}$, conditioned to the observed graph $\boldsymbol{A}$ is

$$P(\boldsymbol{\pi}|\boldsymbol{A}) = \frac{P(\boldsymbol{\pi})}{P(\boldsymbol{A})} \sum_{\{\boldsymbol{g}\}} P(\boldsymbol{A}, \boldsymbol{g}|\boldsymbol{\pi}), \tag{3.1}$$

where the prior $P(\boldsymbol{\pi})$ includes all the information about the parameters that does not depend on the graph and, in the absence of this, we assume it as uniform. Hence, the sum is over all possible assignments $\boldsymbol{g}$, where $g_i \in \{1, \dots, k\}$ for each node i, *i.e.* we need to consider the averages over the posterior distribution of group assignments. That is

$$P(\boldsymbol{g}|\boldsymbol{A}, \boldsymbol{\pi}) = \frac{P(\boldsymbol{A}, \boldsymbol{g}|\boldsymbol{\pi})}{\sum_{\{\boldsymbol{t}\}} P(\boldsymbol{A}, \boldsymbol{t}|\boldsymbol{\pi})}. \tag{3.2}$$

where the sum runs over all possible assignments $\{\boldsymbol{t}\}$.

The *Belief Propagation* approach consists in maximizing step-by-step both the posterior (3.2) of labels $\boldsymbol{g}$ and the posterior (3.1) of parameters $\boldsymbol{\pi}$, up to convergence. To this end, we need to compute the marginals $\psi^i_{g_i} = \sum_{\{g_j\}_{j\neq i}} P(\{g_j\}_{j\neq i}, g_i|\boldsymbol{A}, \boldsymbol{\pi})$ for each node i of the posterior, from which the most probable assignment is

$$\hat{g}_i = \text{argmax}_{a\in\{1,\dots,k\}}\, \psi^i_a. \tag{3.3}$$

Generalizing [Decelle et al., 2011a], marginals can be estimated by using the messages, similarly to (1.44) but generalized for directed graphs, as

$$\psi^i_{g_i} = \frac{q_a e^{-h_i}}{Z^i} \prod_{k\in\partial i} \sum_{g_k} c^{A_{ik}}_{g_i g_k} c^{A_{ki}}_{g_k g_i} \bar{\psi}^{k\to i}_{g_k}, \tag{3.4}$$

where $h_i = \frac{1}{N}\sum_k \sum_{g_k} (c_{g_k g_i} + c_{g_i g_k}) \psi^k_{g_k}$ and the messages $\bar{\psi}^{i\to j}_{g_i}$, for generic neighbors (i, j) and $g_i \in \{1, \dots, k\}$, are the fixed point of the BP equations

$$\psi^{i\to j}_{g_i} = \frac{q_a e^{-h_i}}{Z^{i\to j}} \prod_{k\in\partial i\setminus j} \sum_{g_k} c^{A_{ik}}_{g_i g_k} c^{A_{ki}}_{g_k g_i} \psi^{k\to i}_{g_k}. \tag{3.5}$$

In the previous equations Z^i and $Z^{i\to j}$ are normalization factors, since $\psi^i_{g_i}$ and $\psi^{i\to j}_{g_i}$ are probabilities, and the neighborhood of a node i is intended as those nodes for which $A_{ij} = 1$ or $A_{ji} = 1$. Thus it is important to stress that message information propagates along the undirected skeleton of the network, while directions only change the interaction between nodes.

Marginals and messages are used also to maximize (3.1), finding the optimal parameters

$$q_a = \frac{\sum_i \psi_a^i}{N}, \tag{3.6}$$

$$c_{ab} = \frac{1}{N}\frac{c_{ab}}{q_a q_b}\sum_{i,j}\frac{\psi_a^{i\to j}\cdot\psi_b^{j\to i}}{Z^{ij}}. \tag{3.7}$$

We use an Expectation-Maximization procedure of alternatively solving until convergence the BP equations (3.4)-(3.5) and the equations (3.6)-(3.7).

Finally, to measure the accuracy of the estimator (3.3) we use the measure of *overlap* as defined in (1.27).

Difficult cases for detectability of DiSBM

As highlighted in [Decelle et al., 2011a] for the case of undirected graphs, the problem of community detection is trivial when different groups have different total average degree c_a. In this case the total degree distribution is multimodal (linear combination of Poisson distributions with different means) and nodes can be classified according to their degree, since each mode of the distribution is correlated with the corresponding group. The addition of directions make the detection even easier, since nodes can be classified according to their in(out)-degree even when the groups have homogeneous total average degree. For example an affinity matrix $\begin{pmatrix} c/2 & c \\ 0 & c/2 \end{pmatrix}$ generates a directed network with a strong bipartite structure that becomes completely undetectable without the directions. Thus the problem becomes non trivial once

$$c_a^{in} = c_a^{out} = c_a/2 = c/2, \quad \forall a = 1,\dots,k. \tag{3.8}$$

Circulant models

Without losing generality we consider the case of equally sized group, i.e. $q_a = 1/k$, for which the conditions (3.8) means that the affinity matrix must be a multiple of a *doubly stochastic* matrix, *i.e.* with constant sums along rows and columns. This implies it can be represented as a linear combination of permutation matrices, *i.e.* there exist coefficients

$$c_{in}, c_{out}^{(1)}, \dots, c_{out}^{(d)},$$

and permutations

$$\sigma_1, \dots, \sigma_d; \sigma_i : \{1,\dots,k\} \to \{1,\dots,k\},$$

such that

$$\boldsymbol{c} = c_{in}\mathbb{1} + c_{out}^{(1)}P^{\sigma_1} + \ldots + c_{out}^{(d)}P^{\sigma_d}, \tag{3.9}$$

where P_{ij}^{σ} is defined as 1 if $j = \sigma(i)$ and 0 otherwise.

Of particular interest are the cycles of different order $\mathscr{C}_d(i) = i + d \bmod k$ because they bring to asymmetric affinity matrices where the role of the directions emerge. They produce the class of non trivial directed models where the affinity matrix $\boldsymbol{c} = c_{in}\mathbb{1} + c_{out}^{(1)}P^{\mathscr{C}_1} + \ldots + c_{out}^{(d)}P^{\mathscr{C}_d}$ is a circulant matrix, *i.e.* with constant diagonals of any order.

A special, yet important, case is what we call *asymmetric planted partition* model, in which the affinity matrix depends on just three parameters $c_{in}, c_{out}^{(1)}, c_{out}^{(2)}$ as

$$\boldsymbol{c} = \begin{pmatrix} c_{\text{in}} & c_{\text{out}}^{(1)} & c_{\text{out}}^{(2)} & \cdots & c_{\text{out}}^{(2)} \\ c_{\text{out}}^{(2)} & c_{\text{in}} & c_{\text{out}}^{(1)} & \cdots & c_{\text{out}}^{(1)} \\ c_{\text{out}}^{(1)} & c_{\text{out}}^{(2)} & c_{\text{in}} & \cdots & c_{\text{out}}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{\text{out}}^{(1)} & c_{\text{out}}^{(2)} & c_{\text{out}}^{(1)} & \cdots & c_{\text{in}} \end{pmatrix}. \tag{3.10}$$

A schematic representation of the simplest case of a network generated by such a model with three groups ($k = 3$) is shown in Fig. 3.1.

Note that from the perspective of asymmetry in the model, it is only interesting to analyze the above case when k is odd. Otherwise, we end up with a symmetric affinity matrix and the detectability is described only by the network assortativity. Let us define

$$\varepsilon = \frac{c_{\text{out}}^{(1)} + c_{\text{out}}^{(2)}}{2c_{\text{in}}}, \qquad \gamma = \frac{c_{\text{out}}^{(1)}}{c_{\text{out}}^{(2)}}. \tag{3.11}$$

to measure, respectively, the level of assortativity, similarly to the undirected case [Decelle et al., 2011b, Decelle et al., 2011a], and the level of asymmetry in the direction of connections between each pair of groups.

3.3 Detectability transitions

For simple undirected graphs, [Decelle et al., 2011a] show that the detectability transition is governed by the level of assortativity ε of the network: below a given threshold value, similarity in the connections within and between groups does not allow to distinguish what are the nodes belonging to a given community.

In the case of directed graphs, we can further exploit the information on link direction, *i.e.* the level of asymmetry γ. Indeed, let us notice that there exists a slight

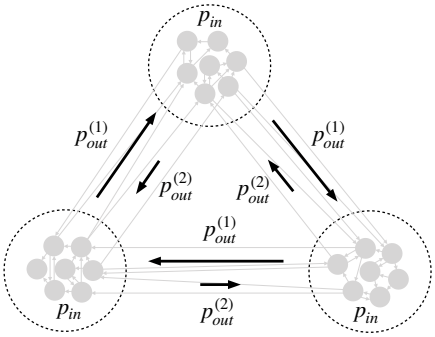


Fig. 3.1 Schematic plot showing a network generated with asymmetric planted partition in case of $k = 3$. Probabilities of given edges are described by p_{in}, $p_{\text{out}}^{(1)}$ and $p_{\text{out}}^{(2)}$.

but important difference with respect to the symmetric case of undirected graphs. Also in the difficult case of equal out-degrees and in-degrees, from asymmetric link densities for the two opposite directions between each pair of groups[1] we could learn more if we do not disregard information about direction. For instance, let us consider first the symmetric case of undirected graphs. Let us consider a subgraph of the network formed by nodes and links belonging to any pair of groups. In the difficult case for inference, the degree of the considered nodes is the same in average both in the whole network and in the subgraph. This symmetry, by contrast, is broken in the case of directed graphs when asymmetric link densities exist between any pair of groups. If we consider link direction, subparts are easier than the difficult case represented by the whole. This symmetry breaking can help detectability of network communities. This is confirmed by studying how the detectability threshold is modified in the case of the asymmetric planted partition model.

For the difficult case of inference of DiSBM specified by (3.8), the BP equations always have the so-called trivial fixed point $\psi_{g_i}^{i\to j} = q_{g_i}$, which does not carry any information on the group structure. When the trivial solution is locally stable the inference becomes impossible or at best extremely hard, as solving the hardest known optimization problems [Franz et al., 2001, Krzakala and Zdeborová, 2009].

To check the stability we study how small random perturbations of the trivial fixed point propagate as the BP equations are iterated. The detectability threshold is obtained as the condition for which the trivial fixed point becomes unstable, see the section in appendix at the present chapter for further technical details. Using the transfer matrix $T_{ab} = q_a \left(\frac{c_{ab}}{c/2} - 1 \right)$ we obtain the following detectability threshold,

$$c\lambda = 1, \tag{3.12}$$

[1] In the limit of zero link density in one direction, the subset of nodes and links for the two groups has a bipartite structure that is easier to recognize (in absence of other nodes) with respect to the corresponding symmetric case where links have no direction.

where c is the network average degree (3.8) and λ is the largest eigenvalue of matrix $T'T$.

1. As soon as $c\lambda < 1$, the perturbation vanishes as we go through the network and the trivial fixed point is locally stable: this is the undetectable phase, where the network is indistinguishable from a random Erdos-Renyi graph.
2. On the other hand, for $c\lambda > 1$ the perturbation is amplified exponentially: the trivial fixed point is unstable and BP solution shows a positive overlap between the inferred and the original communities: this is the detectable phase.

In the special case of the asymmetric planted partition model described by (3.10), the detectability condition is

$$ck^2 = (2c_{\text{in}} - (c_{\text{out}}^{(1)} + c_{\text{out}}^{(2)}))^2 + (c_{\text{out}}^{(1)} - c_{\text{out}}^{(2)})^2 \cot^2\left(\frac{\pi}{2k}\right). \tag{3.13}$$

The first term in the r.h.s. represents the signal carried by the symmetrized adjacency matrix, thus for $c_{\text{out}}^{(1)} = c_{\text{out}}^{(2)}$ we get the original equation for the largest eigenvalue in the undirected case of planted partition model [Decelle et al., 2011a]. Conversely, the second term represents the signal carried by the directions because of the asymmetry in the affinity matrix.

For low number of groups, for example $k = 3$ in the top left panel of Figure 3.2, there are only two distinct phases, the detectable phase with positive overlap and the undetectable phase with zero overlap, and the observed phase transition is continuous at the critical line (3.50). In this case BP recursion converges quickly to the globally attractive fixed point (many equivalent up to a permutation in the ordered phase) for any random initial conditions for messages and marginals: convergence time diverges at the transition. Interestingly, one can reach the detectable phase by either increasing the assortativity ε or by decreasing the symmetry γ. In both cases there is a range of values that guarantees detectability, regardless of the other parameter value.

From (3.50) one sees that increasing the average degree decreases the size of the undetectable phase. As shown on the top right panel of Figure 3.2, this behaviour works similarly for either assortativity or symmetry of connections. The bottom panel shows a different behaviour when varying the number of groups. The critical value of ε, below which the network is detectable regardless of γ, decreases by increasing k. This is somehow intuitive because higher number of groups increases the task complexity. Surprisingly, however, the critical value of γ, below which the network is detectable regardless of ε, does not change much. Moreover, although the change is small, this value is increasing, making even more important not to disregard the information about link direction. When $k \to \infty$ the assortativity becomes irrelevant and one finds that there is only critical value $\gamma_c = (\frac{2}{\pi}\sqrt{c} - 1)/(\frac{2}{\pi}\sqrt{c} + 1)$.

Similarly to the undirected case [Decelle et al., 2011a], increasing the number of groups, $k \geqslant 5$, the hard phase appears and the transition becomes of first order. The hard phase can be identified by studying the sensitivity of BP fixed points to initial conditions: in the hard phase the fixed point correlated with the original assignment coexists with the trivial one and has a very small basin of attraction. Thus,

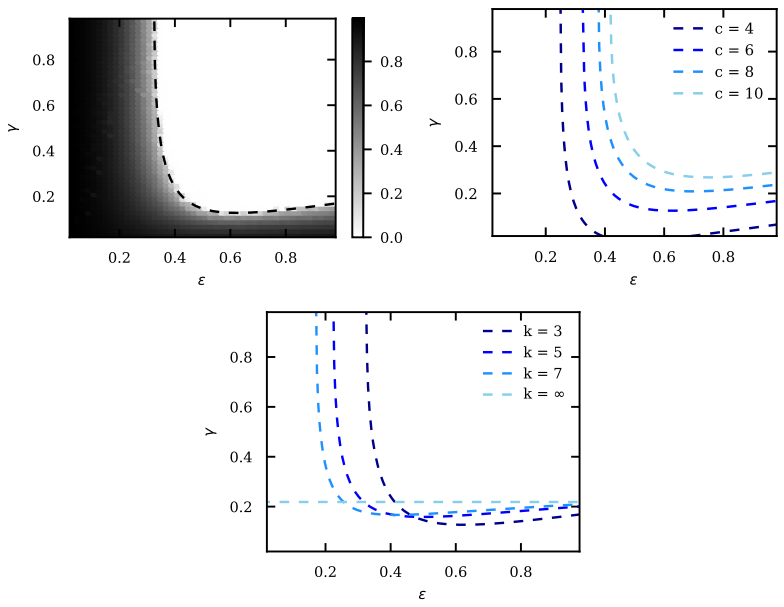


Fig. 3.2 Top left: Two dimensional phase diagram of an asymmetric planted partition model with $k = 3$ groups and average degree $c = 6$, obtained numerically for network with $N = 9 \cdot 10^4$ nodes. The colour scale corresponds to the overlap between the inferred and the real nodes' assignment. Dashed black line represents the analytical approximation of (3.50). Top right: analytical critical lines when number of groups equal to $k = 3$ and varying average degree c. Bottom: analytical critical lines when network average degree is equal to $c = 6$ and varying number of groups k.

starting the recursion from random messages and marginals, BP will never reach a solution with positive overlap. Nevertheless such a solution does exist and still can be found by initializing marginals and messages close enough to it, for example $\psi_a^i = \psi_a^{i\to j} = \delta_{a,g_i}$. The top panels of Figure 3.3 indicate that, in this case, not only the transition point moves towards lower assortativity with increasing asymmetry, but also the size of the hard phase is decreasing. Furthermore, looking at the phase diagram for the asymmetry parameter, (bottom panel of Figure 3.3) we find a different behaviour. Although for $\varepsilon = 0.8$ we can still observe a small jump of the overlap, the presence of the hard phase cannot be confirmed with numerical simulations. When the assortativity disappears completely, the jump seemingly vanishes as well.

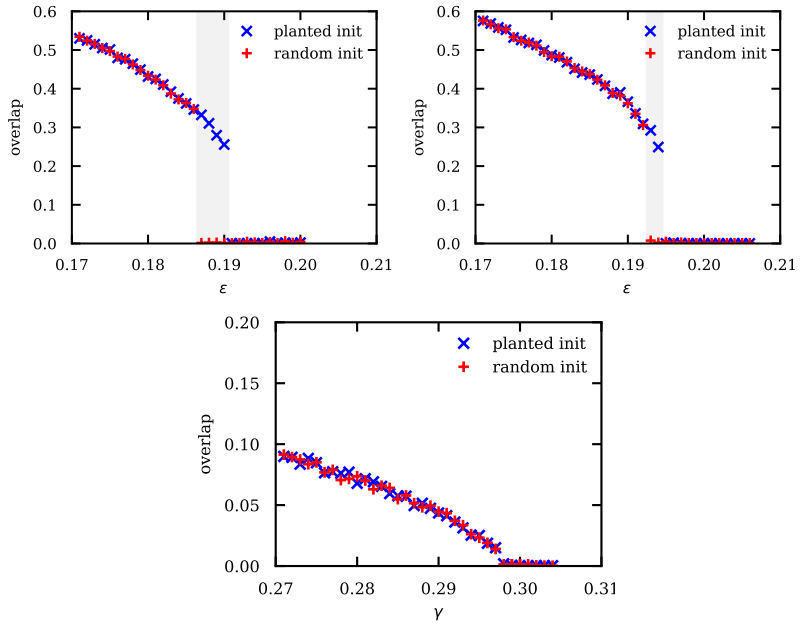


Fig. 3.3 Phase diagrams of an asymmetric planted partition with $k = 9$ groups and average degree $c = 9$. Results obtained numerically, for networks with $N = 3.6 \cdot 10^5$ nodes, using belief propagation with random initial conditions (red pluses) and initiated with the correct assignments (blue crosses). The gray area approximates the hard phase regime, according to numerical results. Left: Results for fixed asymmetry parameter $\gamma = 0.8$. Right: Results for fixed asymmetry parameter $\gamma = 0.6$. Bottom: Results for fixed assortativity parameter $\varepsilon = 0.8$.

Conclusions

In this chapter, we have studied the problem of community detection limits for the directed stochastic block model with an asymmetric affinity matrix. We have shown the detectability condition for a broad range of cases with commuting transfer matrices. We have proposed and focused on the asymmetric planted partition to obtain the detectability threshold as a function of the model parameters. The results show that positive overlap with the correct assignments can be achieved not only by increasing assortativity but also by increasing asymmetry. Both parameters, above a given threshold, lead to positive detectability regardless of the value of the other. This was also confirmed by extensive numerical simulations.

Similarly to the undirected case, the phase transition type depends on the number of groups. For small number of groups we have observed a second order phase transitions in both assortativity and asymmetry. When the number of groups is above five, the phase transitions becomes discontinuous. The observed jump is, however, decreasing with increasing asymmetry and so is the size of the hard phase. For small

values of the assortativity parameter, the phase transition in asymmetry is no longer distinguishable from a continuous phase transition.

Varying the number of groups also affects the critical line. Larger number of groups require a higher assortativity level in order for the network to have detectable community structure. Interestingly, it does not affect the critical level of symmetry much. A more intuitive effect is observed when we vary the average degree. In this case, both the critical assortativity and asymmetry do increase.

Appendix: Detectability threshold in DiSBM

BP equations for DiSBM

The probability that a directed stochastic block model, described by parameters $\boldsymbol{\pi} = \{\{q_a\}_{a=1,\dots,k}, \{p_{ab}\}_{a,b=1,\dots,k}\}$, generates a graph G, with adjacency matrix $\boldsymbol{A}$, is equal to

$$P(\boldsymbol{A}, \boldsymbol{g}|\boldsymbol{\pi}) = \prod_{i\neq j}\left(p_{g_i,g_j}^{A_{ij}}(1-p_{g_i,g_j})^{1-A_{ij}}\right)\prod_i q_{g_i}. \tag{3.14}$$

Using Bayes' rule, we can find the posterior probability of a given assignment $\boldsymbol{g}$, given the adjacency matrix $\boldsymbol{A}$ and model parameters $\boldsymbol{\pi}$, as

$$P(\boldsymbol{g}|\boldsymbol{A}, \boldsymbol{\pi}) = \frac{P(\boldsymbol{A}, \boldsymbol{g}|\boldsymbol{\pi})}{\sum_{\{\boldsymbol{t}\}} P(\boldsymbol{A}, \boldsymbol{t}|\boldsymbol{\pi})}. \tag{3.15}$$

In the directed case, above equations lead to belief propagation messages that differ from their undirected counterparts. In fact, we need to take into account both incoming and outgoing links which results with following equations

$$\psi_{g_i}^{i\to j} = \frac{q_{g_i}}{Z^{i\to j}}\prod_{\substack{k\neq i\\ k\neq j}}\left(\sum_{g_k} c_{g_i g_k}^{A_{ik}} c_{g_k g_i}^{A_{ki}}\left(1-\frac{c_{g_i g_k}}{N}\right)^{1-A_{ik}}\left(1-\frac{c_{g_k g_i}}{N}\right)^{1-A_{ki}}\psi_{g_k}^{k\to i}\right). \tag{3.16}$$

The marginal equations change accordingly

$$\psi_{g_i}^{i} = \frac{q_{g_i}}{Z^{i}}\prod_{k\neq i}\left(\sum_{g_k} c_{g_i g_k}^{A_{ik}} c_{g_k g_i}^{A_{ki}}\left(1-\frac{c_{g_i g_k}}{N}\right)^{1-A_{ik}}\left(1-\frac{c_{g_k g_i}}{N}\right)^{1-A_{ki}}\psi_{g_k}^{k\to i}\right). \tag{3.17}$$

For large sparse networks, when $N \gg 1$ and $c_{ab} = O(1)$ we can rewrite the message passing by neglecting terms of sub-leading order in N.

$$\psi_{g_i}^{i\to j} = \frac{q_{g_i} e^{-h_i}}{Z^{i\to j}}\prod_{k\in\partial i\setminus j}\left(\sum_{g_k} c_{g_i g_k}^{A_{ik}} c_{g_k g_i}^{A_{ki}}\psi_{g_k}^{k\to i}\right), \tag{3.18}$$

where

$$h_i = \frac{1}{N}\sum_k\sum_{g_k}(c_{g_k g_i} + c_{g_i g_k})\psi_{g_k}^{k}. \tag{3.19}$$

Similarly for marginals

$$\psi_{g_i}^i = \frac{q_{t_i} e^{-h_i}}{Z^i} \prod_{k \in \partial i} \left(\sum_{g_k} c_{g_i g_k}^{A_{ik}} c_{g_k g_i}^{A_{ki}} \psi_{g_k}^{k \to i} \right). \tag{3.20}$$

In this case the partition function can be written as a sum

$$Z^i = \sum_{g_i} q_{t_i} e^{-h_i} \prod_{k \in \partial i} \left(\sum_{g_k} c_{g_i g_k}^{A_{ik}} c_{g_k g_i}^{A_{ki}} \psi_{g_k}^{k \to i} \right). \tag{3.21}$$

Moreover, we can write the free energy estimate for BP as

$$f_{\mathrm{BP}}(k, \{q_a\}, \{c_{ab}\}) = \frac{1}{N} \sum_{(i,j) \in E} \log Z^{ij} - \frac{1}{N} \sum_i \log Z^i - \frac{c}{2}, \tag{3.22}$$

where E is the set of all directed edges, which may include separately edges (i, j) and (j, i). In the first term we have

$$Z^{ij} = \sum_{a,b} c_{ab} \, \psi_a^{i \to j} \, \psi_b^{j \to i} = \Psi^{i \to j} \, c \, \Psi^{j \to i}. \tag{3.23}$$

Using this form of the free energy, we can calculate the Nishimori conditions, thus obtaining

$$q_a = \frac{\sum_i \psi_a^i}{N}, \tag{3.24}$$

$$c_{ab} = \frac{1}{N} \frac{c_{ab}}{q_a q_b} \sum_{(i,j) \in E} \frac{\psi_a^{i \to j} \cdot \psi_b^{j \to i}}{Z^{ij}}. \tag{3.25}$$

By exploiting the previous equations, we can infer both the group assignments and the underlying directed stochastic block model parameters with a similar BP algorithm described in Subsection 1.2.5.2.

Detectability threshold for DiSBM

We are particularly interested in the difficult case of inference, *i.e.*

$$\sum_{b=1}^{k} c_{ab} q_b = \sum_{b=1}^{k} c_{ba} q_b = \frac{c}{2} \qquad \forall a = 1, 2, \ldots, k. \tag{3.26}$$

For such graphs, the BP equations always have the so called trivial fixed point

$$\psi_{g_i}^{i \to j} = q_{g_i}, \tag{3.27}$$

which does not carry any information on the group structure. In this difficult case, it holds that the group assignments are not detectable as soon as the trivial fixed point is stable. To check the stability we consider perturbations of it as follows, where, being in the sparse network regime, we can consider a treelike approximation. Let us consider the path from a leaf k_d to a root k_0 to be $k_d, k_{d-1}, \ldots, k_1, k_0$ and assume that on the leaves the trivial fixed point is perturbed as

$$\psi_a^{k_d} = q_a + s_a^{k_d}. \tag{3.28}$$

Following the idea of leaves perturbation, taken from the undirected case, we calculate the transfer matrix

$$T_i^{ab} = \left.\frac{\partial \psi_a^{k_i}}{\partial \psi_b^{k_{i+1}}}\right|_{\psi_a = q_t} = q_a \left(\frac{c_{ab}^{A_{k_i,k_{i+1}}} \cdot c_{ba}^{A_{k_{i+1},k_i}}}{c/2} - 1 \right). \tag{3.29}$$

Unlike in the undirected case, the transfer matrix does depend on i. Nevertheless, if we assume that two way links are neglectable, we only obtain either $T_{ab} = q_a \left(\frac{c_{ab}}{c/2} - 1 \right)$ or its transposition $T'_{de} = q_e \left(\frac{c_{de}}{c/2} - 1 \right)$, depending on the direction of the link. The perturbation $s_{g_0}^{k_0}$ on the root due to the perturbation $s_{g_d}^{k_d}$ on the leaf k_d can then be written as

$$s_{s_0}^{k_0} = \sum_{\{g_i\}_{i=1,\ldots,d}} \left(\prod_{i=0}^{d-1} T_i^{g_i, g_{i+1}} \right) s_{g_d}^{k_d}, \tag{3.30}$$

or in matrix notation

$$s^{k_0} := T_{k_d \to k_0} s^{k_d} = \left(\prod_{i=0}^{d-1} T_i \right) s^{k_d}. \tag{3.31}$$

In particular the strength of the perturbation at the root given by the leaf is

$$\langle s^{k_0}, s^{k_0} \rangle = \langle T_{k_d \to k_0} s^{k_d}, T_{k_d \to k_0} s^{k_d} \rangle = \langle s^{k_d}, T'_{k_d \to k_0} T_{k_d \to k_0} s^{k_d} \rangle \leqslant \lambda_{k_d} \langle s^{k_d}, s^{k_d} \rangle, \tag{3.32}$$

where λ_{k_d} is the largest eigenvalue of the matrix $T'_{k_d \to k_0} T_{k_d \to k_0}$. If we consider the total perturbation induced to the root by all the leaves at distance d on the tree we obtain

$$s^{k_0} := \sum_{k_d} T_{k_d \to k_0} s^{k_d}, \tag{3.33}$$

whose strength, if we consider uncorrelated perturbations on the leaves, reads as

$$\langle s^{k_0}, s^{k_0} \rangle = \sum_{k_d} \langle s^{k_d}, T'_{k_d \to k_0} T_{k_d \to k_0} s^{k_d} \rangle \leqslant \sum_{k_d} \lambda_{k_d} \langle s^{k_d}, s^{k_d} \rangle. \tag{3.34}$$

In principle λ_{k_d} does depend on the path, *i.e.* the leaf k_d, but as soon as $[T,T']=0$ we have that $T'_{k_d\to k_0}T_{k_d\to k_0}=(T'T)^d$, independently from k_d, thus, assuming normalized perturbations on the leaves,

$$\langle s^{k_0},s^{k_0}\rangle \leqslant \lambda^d \sum_{k_d}\langle s^{k_d},s^{k_d}\rangle \sim \lambda^d c^d, \tag{3.35}$$

where λ is the largest eigenvalue of $T'T$ and c^d is the expected number of leaves at distance d. This leads to the following detectability threshold

$$c\lambda = 1. \tag{3.36}$$

In fact as soon as $c\lambda<1$, the perturbation vanishes as we go through the tree and the factorised fixed point is the only stable solution. In this case the detection is impossible. On the other hand, for $c\lambda>1$ the perturbation is amplified exponentially and the communities are detectable.

Asymmetric Planted Partition Case

Now let us consider the asymmetric planted partition matrix, defined by a following affinity matrix:

$$\boldsymbol{p}=\begin{bmatrix} p_{\text{in}} & p_{\text{out}}^{(1)} & p_{\text{out}}^{(2)} & \cdots & p_{\text{out}}^{(2)} \\ p_{\text{out}}^{(2)} & p_{\text{in}} & p_{\text{out}}^{(1)} & \cdots & p_{\text{out}}^{(1)} \\ p_{\text{out}}^{(1)} & p_{\text{out}}^{(2)} & p_{\text{in}} & \cdots & p_{\text{out}}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{\text{out}}^{(1)} & p_{\text{out}}^{(2)} & p_{\text{out}}^{(1)} & \cdots & p_{\text{in}} \end{bmatrix} \tag{3.37}$$

and equal group fractions. It is a special case of a circulant matrix, which is a matrix of a following form

$$D=\begin{bmatrix} d_0 & d_1 & d_2 & \dots & d_{k-1} \\ d_{k-1} & d_0 & d_1 & \dots & d_{k-2} \\ d_{k-2} & d_{k-1} & d_0 & \dots & d_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_1 & d_2 & d_3 & \dots & d_0 \end{bmatrix} \tag{3.38}$$

As a result, both T and $T'T$ matrices are also circulant. The transfer matrix, in this case, is parametrised by three values

$$d_n(T)=\begin{cases} X & \text{if} \quad n=0 \\ Y & \text{if} \quad n \text{ odd} \\ Z & \text{if} \quad n \text{ even} \end{cases} \tag{3.39}$$

where $n = 0, 1, 2, \ldots, k-1$ and

$$X = \frac{1}{k}\left(\frac{c_{\text{in}}}{c/2} - 1\right), \quad Y = \frac{1}{k}\left(\frac{c_{\text{out}}^{(1)}}{c/2} - 1\right), \quad Z = \frac{1}{k}\left(\frac{c_{\text{out}}^{(2)}}{c/2} - 1\right). \tag{3.40}$$

We can use the above parameters to describe the matrix $T'T$ as a circulant matrix

$$d_n(T'T) = \begin{cases} X^2 + \frac{k-1}{2}(Y^2+Z^2) & \text{if} \quad n = 0 \\ \frac{n-1}{2}(Y^2+Z^2) + (k-n-1)YZ + X(Y+Z) & \text{if} \quad n \text{ odd} \\ \frac{k-n-1}{2}(Y^2+Z^2) + (n-1)YZ + X(Y+Z) & \text{if} \quad n \text{ even} \end{cases} \tag{3.41}$$

The next step is to find the largest eigenvalue of the above matrix. For circulant matrices it is known that the eigenvalues are given by

$$\lambda_m = \sum_{j=0}^{k-1} d_j e^{-\frac{2\pi i m j}{k}}, \tag{3.42}$$

where $m = 0, 1, 2, \ldots, k-1$. Since $d_n(T'T) = s_{k-n}(T'T)$, all the eigenvalues are real and we can concentrate on the real part of the above equation.

$$\lambda_m = \sum_{j=0}^{k-1} d_j \cos\left(2\pi m \frac{j}{k}\right). \tag{3.43}$$

This form of the equation is still difficult though. In order to sum the elements, we will use a simple trigonometric identity

$$\cos\left(2\pi m \frac{j}{k}\right)\cos\left(2\pi m \frac{1}{2k}\right) = \frac{1}{2}\cos\left(2\pi m \frac{2j-1}{2k}\right) + \frac{1}{2}\cos\left(2\pi m \frac{2j+1}{2k}\right). \tag{3.44}$$

Combining (3.41), (3.43) and (3.44) leads to

$$\begin{aligned} \lambda_m = {} & X^2 + \frac{k-1}{2}(Y^2+Z^2) + (k-2)YZ + X(Y+Z) \\ & + \sum_{j=1}^{k-2}\left[\frac{k-2\pm 1}{2}(Y^2+Z^2) + (k-2\mp 1)YZ + 2X(Y+Z)\right] \cdot \frac{\cos\left(2\pi m \frac{2j+1}{2k}\right)}{2\cos\left(\pi \frac{m}{k}\right)}, \end{aligned} \tag{3.45}$$

where $\pm$ means $+$ when j is odd and $-$ when j is even, vice versa for $\mp$. To eventually get rid off the dependence on the parity of j we will need to use one more trigonometric identity

$$\cos\left(2\pi m\frac{2j+1}{2k}\right)\cos\left(2\pi m\frac{1}{2k}\right)=\frac{1}{2}\cos\left(2\pi m\frac{j}{k}\right)+\frac{1}{2}\cos\left(2\pi m\frac{j+1}{k}\right). \tag{3.46}$$

Applying it to Equation (3.45) give

$$\begin{aligned}\lambda_m &= X^2+\frac{k-1}{2}(Y^2+Z^2)+(k-2)YZ+X(Y+Z)\\ &+\left(\frac{k-3}{4}(Y^2+Z^2)+\frac{k-1}{2}YZ+X(Y+Z)\right)\frac{\cos\left(2\pi\frac{m}{k}\right)}{\cos^2\left(\pi\frac{m}{k}\right)}\\ &+\left[(k-2)(Y^2+Z^2)+2(k-2)YZ+4X(Y+Z)\right]\sum_{j=1}^{k-3}\frac{\cos\left(2\pi m\frac{j+1}{k}\right)}{4\cos^2\left(\pi\frac{m}{k}\right)}.\end{aligned} \tag{3.47}$$

The final sum can be calculated using

$$\sum_{j=0}^{n-1}\cos(2\pi j)=1, \tag{3.48}$$

where n is odd. Finally, after simplifying all the parts we get

$$\lambda_m=\frac{1}{c^2k^2}\left((2c_{\text{in}}-(c_{\text{out}}^{(1)}+c_{\text{out}}^{(2)}))^2+(c_{\text{out}}^{(1)}-c_{\text{out}}^{(2)})^2\tan^2\left(\pi\frac{m}{k}\right)\right). \tag{3.49}$$

This form makes it easy to get the two largest eigenvalues

$$\lambda_{\max}=\frac{1}{c^2k^2}\left((2c_{\text{in}}-(c_{\text{out}}^{(1)}+c_{\text{out}}^{(2)}))^2+(c_{\text{out}}^{(1)}-c_{\text{out}}^{(2)})^2\cot^2\left(\frac{\pi}{2k}\right)\right). \tag{3.50}$$

Note that for $c_{\text{out}}^{(1)}=c_{\text{out}}^{(2)}$ we get the original equation for the largest eigenvalue in the undirected case of planted partition.

Chapter 4
Group and link persistence in dynamic stochastic block model

Abstract In this chapter, we study the problem of community detection for dynamic networks whose evolution is governed by both a latent dynamics for the group memberships and a Markovian behavior for the link persistence. In particular, we study the inference of a dynamic stochastic block model in which both communities and links keep memory of previous network states. By considering maximum likelihood inference from single snapshot observations of the network, we show that link persistence makes the inference of communities harder, decreasing the detectability threshold, while community persistence tends to make it easier. We analytically show that communities inferred from single network snapshot can share a maximum overlap with the underlying communities of a specific previous instant in time. This leads to time-lagged inference: the identification of past communities rather than present ones. Finally we compute the time lag and propose a corrected algorithm, the Lagged Snapshot Dynamic (LSD) algorithm, for community detection in dynamic networks. We analytically and numerically characterize the detectability transitions of such algorithm as a function of the memory parameters of the model and we make a comparison with a full dynamic inference.
Almost all results in this chapter previously appeared in [Barucca et al., 2018].

Introduction

Community detection in time-evolving interacting systems is an open problem in data mining. Temporal networks [Holme and Saramäki, 2012] provide a framework to study the dynamic evolution of interacting systems, and can be analysed as a sequence of network snapshots. Here, we study the problem of learning the dynamic evolution of the community structure of a temporal network with link and community persistence. With a similar aim of Chapter 2, we focus on the problem of disentangling the two different sources of persistence in a dynamic framework.

Community detection is a long-standing problem that has been thoroughly studied in the static network case with various approaches: modularity maximization

[Newman, 2016], spectral methods [Hendrickson and Leland, 1995, Krzakala et al., 2013], Belief-Propagation (BP) [Decelle et al., 2011a], and other heuristic algorithms [Blondel et al., 2008]. The problem has been faced also in the dynamic case, by generalizing the BP algorithm [Ghasemian et al., 2016]. However, the effect of link persistence has not taken into account.

In this chapter, we focus on stochastic block models with dynamic community structure and link persistence, which introduce time correlations in the network structure. When time correlations are present, the information obtained from the inference on individual snapshots might be contaminated by the past history of the system. This is analogous to what happens in multilayer networks [Boccaletti et al., 2014], for which the analysis cannot be decomposed into the separate analysis over each layer if they are correlated.

Static stochastic block models have been shown to display a detectability transition [Decelle et al., 2011a, Mossel et al., 2014, Mossel et al., 2015b] when the ratio between the average degree within a block of nodes and the average degree towards different blocks, *i.e.* the assortativity parameter, becomes too low: below a critical value of assortativity, detection becomes computationally hard.

Recently the problem was also investigated in temporal networks [Mucha et al., 2010, Yang et al., 2011, Bassett et al., 2013, Bazzi et al., 2016] and in a specific case of Markovian community structures [Ghasemian et al., 2016]. In this latter dynamic network model, it was shown that persistence in communities can help detection, by decreasing the detectability threshold: a weaker assortativity is required to infer communities with respect to the static case. On the contrary, we show that persistence in relations can hinder detection, eventually causing the detection of old communities instead of the ones present at the time the detection is performed. We compute analytically the time lag in community detection and provide a dynamic community detection algorithms for the model under study. The method is built upon optimal static algorithms on individual snapshots combined with our analytic result to correct for the time lag.

This chapter is organized as follows: in Section 4.1 we define the dynamic stochastic block model where both communities and links are persistent in time. In Section 4.2 we study the single snapshot inference and we show how link persistence leads to time lagged inference, that is the detection of past communities rather than present ones. In Section 4.3 we introduce the lagged snapshot dynamic (LSD) algorithm, that corrects static detection algorithm for the time lag. In Section 4.4 we show that the LSD algorithm can be considered an interesting tradeoff between the accuracy of a full dynamic inference and the simplicity of a naive single snapshot inference. Then, we conclude with a discussion on possible directions of applications of the proposed LSD algorithm.

4.1 Dynamic Stochastic Block Model (DSBM) with link and group stability

We consider a Dynamic Stochastic Block Model (DSBM) with link persistence, *i.e.* at each time step the presence of a link between two nodes is copied from the previous time with probability α, while with probability $1-\alpha$ the link is generated according to a SBM where the community structure changes over time. In the proposed DSBM, the community structure changes over time though. Several models of DSBM were previously introduced for community detection in dynamic networks [Zhang et al., 2017, Xu and Hero, 2014, Xu, 2015, Ghasemian et al., 2016]. Differently from these works, our variant includes both link and community persistence.

A dynamic network generated according to DSBM with k groups consists of a set of nodes and a sequence of graph snapshots $(V,\boldsymbol{A}^t)_{t=1}^T$, each with its own community structure $\boldsymbol{g}^t$. As in the previous chapters, we indicate with $\boldsymbol{\mathcal{A}} = \{\boldsymbol{A}^0,...,\boldsymbol{A}^T\}$ the sequence of observed adjacency matrices and with $\vec{\boldsymbol{g}} = \{\boldsymbol{g}^0,...,\boldsymbol{g}^T\}$ the sequence of (latent) node labels. As in [Ghasemian et al., 2016], the dynamic of each node's label g_i^t is an independent Markov process with transition probability

$$\mathbb{P}(g_i^t|g_i^{t-1}) = \eta\,\delta_{g_i^t,g_i^{t-1}} + (1-\eta)q_{g_i^t},$$

see (1.81), meaning that with probability η a node remains in the same community, otherwise it changes randomly to a group a (including g_i^{t-1}) with probability q_a. We have $\boldsymbol{q} = \{q_a\}_{a=1,...,k}$ with $\sum_{a=1}^k q_a = 1$. We assume that at $t=0$ labels are assigned according to the static SBM (*i.e.* according to the marginal multinomial distribution specified by $\boldsymbol{q}$). Hence, it is

$$\mathbb{P}(\vec{\boldsymbol{g}}) = \prod_{i=1}^{N}\left[\prod_{t=1}^{T}\eta\,\delta_{g_i^t,g_i^{t-1}} + (1-\eta)q_{g_i^t}\right]q_{g_i^0} \tag{4.1}$$

Adding link persistence to the DSBM we obtain the persistent dynamic model (for undirected networks), see the flow in Figure 4.1,

$$\begin{aligned}\mathbb{P}(\boldsymbol{\mathcal{A}}|\vec{\boldsymbol{g}}) = &\prod_{i,j>i}^{N} p_{g_i^0 g_j^0}^{A_{ij}^0}(1-p_{g_i^0 g_j^0})^{1-A_{ij}^0}\times \\ &\times\prod_{t=1}^{T}\left(\alpha\;\delta_{A_{ij}^t,A_{ij}^{t-1}} + (1-\alpha)p_{g_i^t g_j^t}^{A_{ij}^t}(1-p_{g_i^t g_j^t})^{1-A_{ij}^t}\right),\end{aligned} \tag{4.2}$$

where at time t we copy a link from the past with probability α, otherwise it is sampled depending on the current community structure, and the network at $t=0$ is generated according to a static SBM from $\boldsymbol{g}^0$. Thus the two parameters η and α can be interpreted as, respectively, the persistence of communities and the persistence of links. Community persistence models the tendency of nodes to remain in the same group over time. Link persistence models the preference of nodes in keeping pre-existent relations over time, for example because of the cost of adding or removing links in socio-economic networks [Amaral et al., 2000].

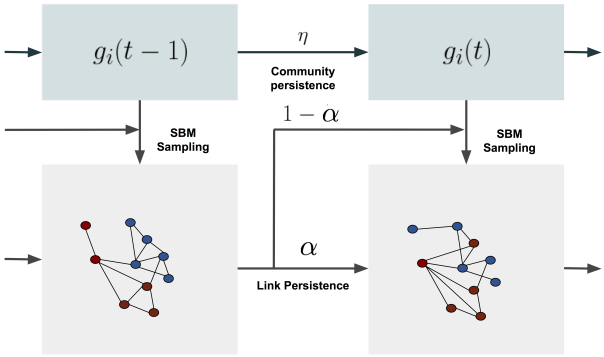


Fig. 4.1 Schematic representation of the persistent dynamic blockmodel. Label is kept with probability η and randomly changed with probability $1-\eta$, while the link relation, *i.e.* the presence or absence of a link, is conserved with probability α or resampled with the new block structure with probability $1-\alpha$.

Here we focus on the common choices of a uniform prior, *i.e.* $q_a = 1/k \ \forall a = 1,\dots,k$, and affinity matrix with a constant p_{in} on the diagonal and another constant $p_{out} \leq p_{in}$ off diagonal, the so called assortative *planted partition model* that is widely used as benchmark in the mathematics and computer science community detection literature [Decelle et al., 2011a, Krzakala et al., 2013, Dyer and Frieze, 1989, Condon and Karp, 2001]. Moreover we measure the level of assortativity with a parameter $a \in [0,1]$ such that

$$\boldsymbol{p} = a\, k\bar{p}\mathbb{1} + (1-a)\bar{p}\mathbf{1} \tag{4.3}$$

interpolating between a fully assortative $k\bar{p}\mathbb{1}$ (proportional to the identity matrix) and a fully random $\bar{p}\mathbf{1}$ (proportional to a matrix of ones) affinity matrix, with fixed mean degree $N/k^2 \sum_{ab} p_{ab} = N\bar{p}$.[1] We are interested in the sparse regime $\bar{p} = \bar{c}/N$, that is the most challenging from the inference perspective, since most of real networks of interest are sparse and because sparsity allows to carry out asymptotically optimal analysis.

The central problem is to study under which conditions we can detect, better than by chance, the correct labeling of the latent communities $\vec{\boldsymbol{g}}$ from the observation of $\boldsymbol{\mathcal{A}}$, together with the most likely model parameters $\boldsymbol{\pi} = (a, \alpha, \eta, \boldsymbol{p}, \boldsymbol{q})$.

For the static SBM, it was shown (and proved at least for $k = 2$ [Mossel et al., 2018]) that there exists a sharp threshold below which no algorithm can perform better than chance in recovering the planted community structure. This threshold occurs, in terms of the parametrization (4.3) at $a = a_c := \bar{c}^{-1/2}$ meaning that there is a necessary minimum signal to noise ratio, in terms of assortativity, under which a

[1] Notice that this measure of assortativity a differs from the definition of ε given in Chapters 1 and 3, *i.e.* $\varepsilon = p_{out}/p_{in}$. This is done here for notational simplicity. However, a simple relation connects the two definitions for the level of assortativity, that is $a = \frac{1+\varepsilon}{1-\varepsilon}$.

community structure may still exist but is undetectable, as previously described in Subsection 1.2.5.

The Maximum A Posteriori (MAP) approach for the estimation of the fully dynamical model DSBM is formally similar to the one described in Subsection 1.2.5.1 for static SBM. For DSBM, let us considers the posterior distribution of the latent node labels

$$\mathbb{P}(\vec{\boldsymbol{g}}|\boldsymbol{\mathcal{A}},\boldsymbol{\pi}) = \frac{\mathbb{P}(\boldsymbol{\mathcal{A}},\vec{\boldsymbol{g}}|\boldsymbol{\pi})}{\sum_{\{\boldsymbol{\gamma}^t\}^{t=0,\dots,T}}\mathbb{P}(\boldsymbol{\mathcal{A}},\{\boldsymbol{\gamma}^t\}^{t=0,\dots T}|\boldsymbol{\pi})} = Z(\boldsymbol{\mathcal{A}},\boldsymbol{\pi})^{-1}e^{-H(\vec{\boldsymbol{g}},\boldsymbol{\mathcal{A}},\boldsymbol{\pi})}, \quad (4.4)$$

where the sum runs over all possible assignments $\boldsymbol{\gamma}^t$ at time t, and we have defined $H(\vec{\boldsymbol{g}},\boldsymbol{\mathcal{A}},\boldsymbol{\pi}) = -\log\mathbb{P}(\boldsymbol{\mathcal{A}},\vec{\boldsymbol{g}}|\boldsymbol{\pi})$. Then, the posterior distribution of the static parameters is

$$\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\mathcal{A}}) = \frac{\mathbb{P}(\boldsymbol{\pi})}{\mathbb{P}(\boldsymbol{\mathcal{A}})}\sum_{\{\boldsymbol{\gamma}^t\}}P(\boldsymbol{\mathcal{A}},\{\gamma^t\}^{t=0,\dots,T}|\boldsymbol{\pi}) \propto \mathbb{P}(\boldsymbol{\pi})Z(\boldsymbol{\mathcal{A}},\boldsymbol{\pi}). \quad (4.5)$$

Maximizing (4.4) leads to the inference of a set of statistically significant communities $\hat{\boldsymbol{g}}^0,\hat{\boldsymbol{g}}^1,\dots,\hat{\boldsymbol{g}}^T$, whereas maximizing (4.5) means learning the most likely set of parameters $\hat{\boldsymbol{\pi}}$ given the data. Using smooth priors $\mathbb{P}(\boldsymbol{\pi})$, $\hat{\boldsymbol{\pi}}$ is obtained by solving the equations

$$\sum_{\vec{\boldsymbol{g}}}Z(\boldsymbol{\mathcal{A}},\boldsymbol{\pi})^{-1}e^{-H(\vec{\boldsymbol{g}},\boldsymbol{\mathcal{A}},\boldsymbol{\pi})}\partial_{\boldsymbol{\pi}}H(\vec{\boldsymbol{g}},\boldsymbol{\mathcal{A}},\boldsymbol{\pi}) = \mathbb{E}\left(\partial_{\boldsymbol{\pi}}H(\vec{\boldsymbol{g}},\boldsymbol{\mathcal{A}},\boldsymbol{\pi})\right) = 0, \quad (4.6)$$

where the expectation is obtained by averaging over the graph ensemble.

Since the maximization of the posterior of static parameters (4.5) requires computing expectations w.r.t. the posterior of latent node assignments (4.4), alternating the inference of latent variables and the learning of static parameters would give rise to an Expectation-Maximization (EM) procedure [Friedman et al., 2001], similarly to the case of the static Belief-Propagation algorithm. The criticality of this approach is in the summation over all possible labels whose number grows exponentially with N. Overtaking this problem for the static case is usually done by Monte Carlo (MC) sampling [Peixoto, 2013] or by using BP [Decelle et al., 2011a, Decelle et al., 2011b]. In the dynamic case, the problem has been solved by [Ghasemian et al., 2016] which generalize BP for dynamic SBM, but without considering link persistence, see Subsection 1.3.3.

The output of both BP and MC approaches is thus providing an estimate of the posterior (4.4) in terms of the marginals

$$\nu_i(g_i^t) = \sum_{\{g_j^s\}_{j\neq i}^{s\neq t}}\mathbb{P}(\vec{\boldsymbol{g}}|\boldsymbol{\mathcal{A}},\boldsymbol{\pi}).$$

From them, a partition is obtained by assigning each node to its most likely group

$$\hat{g}_i^t = \text{argmax}_{g_i^t}\,\nu(g_i^t).$$

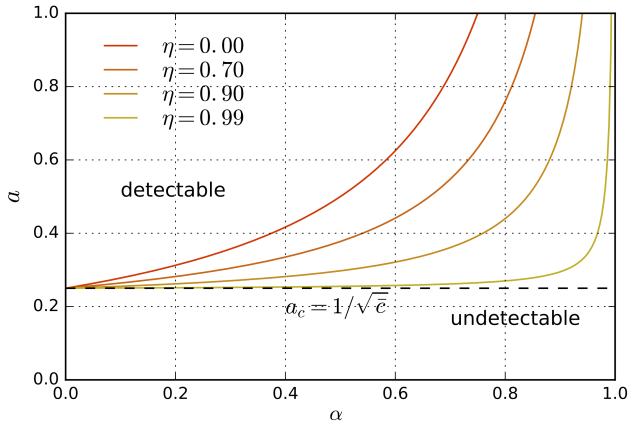


Fig. 4.2 Asymptotic phase space for single snapshot detectability as function of assortativity a and community/ link persistences, η and α, compared to the static threshold a_c, in the case of two equally sized planted groups $k = 2$.

This is known [Iba, 1999] to be an optimal estimator, maximising the overlap (1.27) with the original planted assignment.

Given the marginals, the equations (4.6) for each $\boldsymbol{\pi}_i$ can be solved to learn the most likely set $\hat{\boldsymbol{\pi}}$ of model parameters. To solve each equation, we need to consider averages over the posterior distribution of the node labels, similarly to the static SBM in Section 1.2.5. In particular, the equations for learning both q_a and p_{ab} $\forall a,b = 1,...,k$ are formally similar to (1.36-1.38), but considering also averages over time. Then, the most likely value of the assortativity parameter, *i.e.* $\hat{a}$, is obtained by fitting it on the inferred affinity matrix $\hat{\boldsymbol{p}}$.

However, the inference for the full dynamical model is complicated by the presence of both link and community persistence. In particular, link persistence creates short loops in the spatiotemporal network of 'interactions', *e.g.* if the link (i, j) is persistent from $t-1$ to t, we observe the loop $i(t-1) \to i(t) \to j(t) \to j(t-1) \to i(t-1)$ which breaks the treelike approximation, thus preventing the inference approach based on the replica symmetric cavity method [Mézard and Parisi, 2001].

In the following, we will adopt the EM procedure introduced in [Decelle et al., 2011a] where the static BP algorithm is used for the expectation step, *i.e.* the estimate of the posterior marginals, that is considering only the information on the current network snapshot for the inference of the latent node assignment (the E-step of the BP algorithm is described in Subsection 1.2.5.2).

4.2 Single snapshot inference of DSBM

In this section, we ask first which community structure is inferred from a single snapshot of the dynamic network at a time t, using the static BP algorithm for the inference of the marginals. This might occur, for example, if one is unaware that $\boldsymbol{A}^t$ is one observation of a dynamic process.

Thus we need to compute the posterior $\mathbb{P}(\boldsymbol{g}^t|\boldsymbol{A}^t)$ giving the probability of community structure when only the information on the network at time t is used.[2] It holds that the posterior $\mathbb{P}(\boldsymbol{g}^t|\boldsymbol{A}^t)$ is that of a static SBM with an effective assortativity depending on the dynamics, *i.e.*

$$\mathbb{P}(\boldsymbol{g}^t|\boldsymbol{A}^t) \propto \prod_{i,j>i} p_{g_i^t g_j^t}(a^t_{\alpha,\eta})^{A^t_{ij}}(1-p_{g_i^t g_j^t}(a^t_{\alpha,\eta}))^{1-A^t_{ij}},$$

with

$$a^t_{\alpha,\eta} = a\,\varepsilon^t_{\alpha,\eta} = a\left(\frac{1-\alpha}{1-\alpha\eta^2} + (\alpha\eta^2)^t\frac{\alpha(1-\eta^2)}{1-\alpha\eta^2}\right). \tag{4.7}$$

In fact it is sufficient to note that, from Bayes' rule, $P(\boldsymbol{g}^t|\boldsymbol{A}^t) \propto P(\boldsymbol{A}^t|\boldsymbol{g}^t)$, that can be always be written as

$$\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{g}^t) = \prod_{i,j>i}(p^t_{g_i^t g_j^t})^{A^t_{ij}}(1-p^t_{g_i^t g_j^t})^{1-A^t_{ij}}, \tag{4.8}$$

with $p^t_{ab} \equiv \mathbb{P}(A^t_{ij}=1|g^t_i=a,g^t_j=b)$. Marginalising over previous network instances we get the recursive equation

$$\begin{aligned} p^t_{ab} &= \alpha\,\mathbb{P}(A^{t-1}_{ij}=1|g^t_i=a,g^t_j=b)+(1-\alpha)p_{ab}.\\ &= \alpha\,\left(\eta^2 p^{t-1}_{ab}+(1-\eta^2)\bar{p}\right)+(1-\alpha)p_{ab}, \end{aligned} \tag{4.9}$$

where in the first equality we have conditioned and summed over A^{t-1}_{ij}, while in the second over g^{t-1}_i, g^{t-1}_j and where we used that $\mathbb{P}(A^t_{ij}=1)=\bar{p}$ and $\mathbb{P}(g^t_i=a)=1/k$ $\forall i,j,t,a$, that can be proved recursively. Since p^0_{ab} is simply p_{ab} we get

$$p^t_{ab} = \left(\alpha(1-\eta^2)\bar{p}+(1-\alpha)p_{ab}\right)\sum_{\ell=0}^{t-1}\left(\alpha\eta^2\right)^\ell + \left(\alpha\eta^2\right)^t p_{ab},$$

that gives (4.7) once used the representation (4.3).

Equation (4.7) states that the posterior of a single snapshot of a DSBM is equal to the posterior of a static SBM with modified assortativity parameter. It is important to point out this does not imply that a single snapshot inference gives the planted assignments with modified assortativity parameter. Instead, it states that, *if the inferred assignments are the planted ones*, then the estimated assortativity is the one

[2] Here we neglect the dependence on model parameters $\boldsymbol{\pi}$ for notational simplicity.

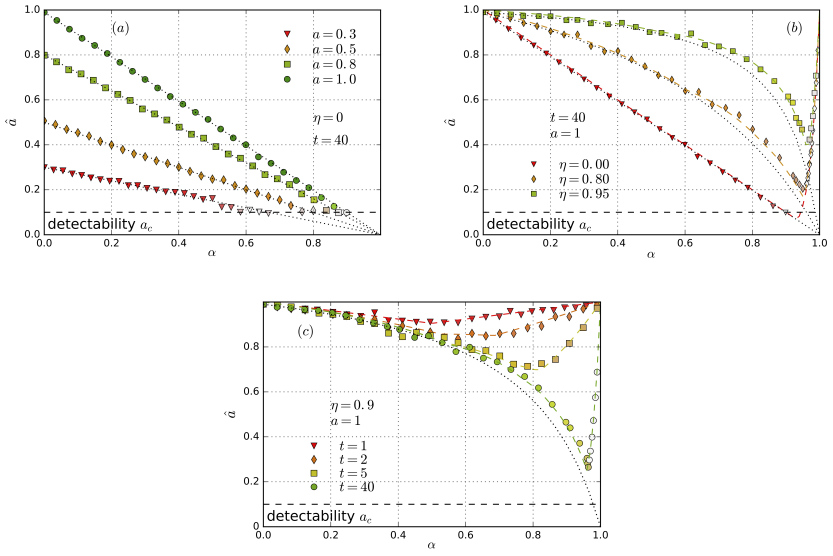


Fig. 4.3 Effective assortativity $\hat{a}$ inferred using static BP from single snapshot observations of a DSBM with link persistence, for different values of α, η and t. Black dotted lines represent $a^t_{\alpha,\eta}$ while colored dashed lines are the theoretical optimum $a^\star_t(\alpha,\eta)$ of Eq. (4.13) . Each point is the result of the inference on a dynamic network with $N = 300$, $T = 40$, $\bar{c} = 10$ and $k = 2$ evolving communities. Vividness of colors is proportional to the overlap $q(\boldsymbol{g}^t, \hat{\boldsymbol{g}}^t)$ between the planted and the inferred communities.

in (4.7), *i.e.* $\hat{a} = a^t_{\alpha,\eta}$, which is smaller than the value a in the generative model. This is because the link persistence α decreases the effective assortative structure of the network, increasing the number of links assigned randomly with respect to those assigned on the base of their group labels. This effect is partially mitigated by the persistence of communities η since it increases the probability that a link copied from a previous time is not actually random but was in turn assigned through the same community structure.

One of the consequences of (4.7) is that the signal provided by the observation of $\boldsymbol{A}^t$ to the community structure at the same time decreases by the effect of the dynamics as $a_c \to a_c/\varepsilon^t_{\alpha,\eta}$, reducing to the static one in absence of link persistence, *i.e.* $\alpha = 0$.[3]. For $t \to \infty$, it is $a^\infty_{\alpha,\eta} = a(1-\alpha)/(1-\alpha\eta^2)$. Figure 4.2 shows the asymptotic phase space as a function of α, where we have defined, in the same spirit of the static case, the detectability threshold as $a(1-\alpha)/(1-\alpha\eta^2) = \bar{c}^{-1/2}$.

Figure 4.3 compares the theoretical predictions of $\hat{a}$ with numerical simulations and the (static) BP inference of a DSBM, by varying planted parameters (a, α, η, t) .

[3] Note that the detectability threshold from single snapshot is however higher than the threshold of the dynamic problem, *i.e.* the inference of all the assignments given the observation of the entire network series. For example [Ghasemian et al., 2016] considers a DSBM without link persistence and shows that the detectability threshold a_c is in general lowered by the group persistence.

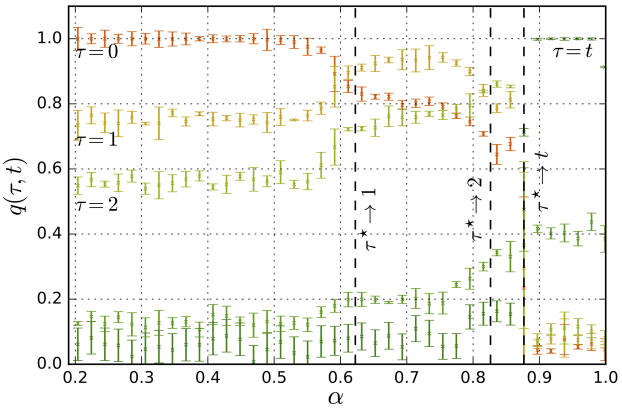


Fig. 4.4 Overlap $q(\hat{\boldsymbol{g}}^t, \boldsymbol{g}^{t-\tau})$ between the communities inferred at time t ($t = 10$) and the planted assignment at time $t-\tau$ of a dynamic network with $N = 300$, $T = 40$, $\bar{c} = 10$ and $k = 2$ evolving communities according to $\eta = 0.75$. Dashed lines are computed by solving the problem in Eq. (4.13).

In the top left panel, the agreement is very good and this holds also in the other panels in the regions when α is small. However the top right panel and the bottom one show that when both α and η are large, some discrepancies between the theoretical curve and the simulations appear. This does not contradict necessarily (4.7), which gives the assortativity parameter if the inferred assignments are the planted ones (or at least close to them). We now show that indeed the observed discrepancies can be explained by the fact that, for large community and link persistence, the inferred assignments are closer to a *past* planted assignment than to those at the time when the single snapshot inference is performed.

Definition 4.1. Given a network sequence of length T generated by the planted partition DSBM with parameters (α, η, a), we define *time lagged inference* the problem of inferring the communities at time $t-\tau$ given the observation of the network snapshot at time t.

It holds that the posterior $P(\boldsymbol{g}^{t-\tau}|\boldsymbol{A}^t) \propto \prod_{i,j>i} P(A^t_{ij}|\boldsymbol{g}^{t-\tau})$ is that of a static SBM with an effective assortativity

$$a^{(t,\tau)}_{\alpha,\eta} = \alpha^\tau a^{t-\tau}_{\alpha,\eta} + (1-\alpha)\eta^2 \frac{\eta^{2\tau} - \alpha^\tau}{\eta^2 - \alpha} a, \tag{4.10}$$

where $a^t_{\alpha,\eta}$ is given by (4.7).

In fact, similarly to (4.7), it sufficient to compute the quantity

$$\mathscr{L}^n = P(A^n_{ij} = 1 | \boldsymbol{g}^{t-\tau} = \boldsymbol{g}),$$

evaluated at $n = t$. For $n \geqslant t-\tau$, keepinig fixed i, j and t , it is

$$\mathscr{L}^n = \sum_{A_{ij}^{n-1}, \boldsymbol{g}^n} P(A_{ij}^n = 1|\boldsymbol{g}^n, A_{ij}^{n-1}) P(\boldsymbol{g}^n, A_{ij}^{n-1}|\boldsymbol{g}^{t-\tau} = \boldsymbol{g}) =$$
$$= \alpha \mathscr{L}^{n-1} + (1-\alpha) \sum_{\boldsymbol{g}^n} p_{g_i^n g_j^n} P(\boldsymbol{g}^n|\boldsymbol{g}^{t-\tau} = \boldsymbol{g}). \tag{4.11}$$

Moreover, defining $\mathscr{T}^n = \sum_{\boldsymbol{g}^n} p_{g_i^n g_j^n} P(\boldsymbol{g}^n|\boldsymbol{g}^{t-\tau} = \boldsymbol{g})$, for $n \geqslant t-\tau$ it is

$$\mathscr{T}^n = \sum_{\boldsymbol{g}^n, \boldsymbol{g}^{n-1}} p_{g_i^n g_j^n} P(\boldsymbol{g}^n|\boldsymbol{g}^{n-1}) P(\boldsymbol{g}^{n-1}|\boldsymbol{g}^{t-\tau} = \boldsymbol{g})$$
$$= \eta^2 \mathscr{T}^{n-1} + (1-\eta^2)\bar{p} \tag{4.12}$$

Solving (4.12) and then (5.11), *i.e.* the recursive equation

$$\mathscr{L}^n = \alpha \mathscr{L}^{n-1} + (1-\alpha)\mathscr{T}^n,$$

we get

$$\mathscr{L}^t = \alpha^\tau \mathscr{L}^{t-\tau} + (1-\alpha) \sum_{\ell=0}^{\tau-1} \alpha^\ell \mathscr{T}^{t-l}$$
$$= \alpha^\tau \mathscr{L}^{t-\tau} + (1-\alpha) \sum_{\ell=0}^{\tau-1} \alpha^\ell \left(\eta^{2(\tau-\ell)} p_{g_i g_j} + (1-\eta^{2(\tau-\ell)})\bar{p} \right).$$

Since $\mathscr{L}^{t-\tau}$ corresponds to the non-lagged $p_{g_i g_j}^{t-\tau}$ in (4.7), we obtain (4.10) simply using the representation (4.3).

The meaning of Equation (4.10) is that every lagged inference problem has the posterior of a static SBM with effective assortativity $a_{\alpha,\eta}^{(t,\tau)}$. Thus fixing t and varying τ we have a sequence of inference problems with the same posterior, same input data $\boldsymbol{A}^t$, and only different effective assortativity, thus detectability threshold.

Fig. 4.4 shows the overlap of (1.27) between the inferred communities $\hat{\boldsymbol{g}}^t$ and the planted ones at $t-\tau$. For small α the maximum overlap is with $\boldsymbol{g}^t$, while for larger α we observe a series of transitions where the largest overlap is with a $\boldsymbol{g}^{t-\tau}$ with $\tau > 0$.

We now show that the τ that maximizes the overlap $q(\hat{\boldsymbol{g}}^t, \boldsymbol{g}^{t-\tau})$ is the one for which the effective assortativity $a_{\alpha,\eta}^{(t,\tau)}$ is maximal. To this end, let us define

$$a_t^\star(\alpha,\eta) = \max_{\tau \leqslant t} a_{\alpha,\eta}^{(t,\tau)}, \qquad \tau_t^\star(\alpha,\eta) = \operatorname*{argmax}_{\tau \leqslant t} a_{\alpha,\eta}^{(t,\tau)}. \tag{4.13}$$

The top left panel of Figure 4.5 shows that for small link persistence α, $\tau_t^\star(\alpha,\eta) = 0$, *i.e.* a single snapshot inference solves the problem at the time of the observed snapshot t. At a critical α, depending on η and t, it is $\tau_t^\star(\alpha,\eta) > 0$, suggesting that the inference procedure converges to the assignments at time $t - \tau_t^\star$. In fact the dashed lines in Figure 4.4 are computed by solving the problem in (4.13) and it is clear that they correspond to the transitions in the overlap. Moreover the theoretical $a_t^\star(\alpha,\eta)$ is shown in Figure 4.3 to be in perfect agreement with the inferred assortativity $\hat{a}$.

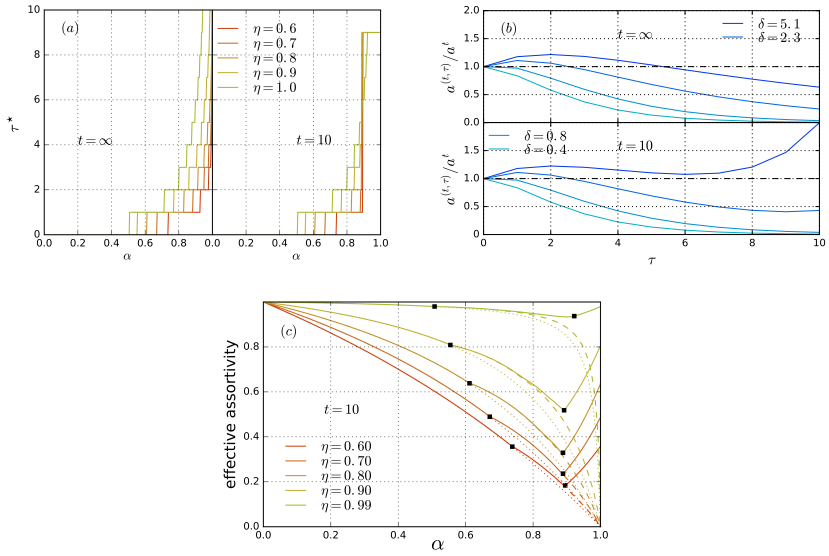


Fig. 4.5 Optimal lag $\tau_t^*(\alpha,\eta)$ (top left panel) as function of α and η, and $a_{\alpha,\eta}^{(t,\tau)}/a_{\alpha,\eta}^t$ (top right panel) as function of the lag τ for different $\delta = \eta^2\alpha/(1-\alpha)$, both in the asymptotic limit $t \to \infty$ and for finite t. Bottom panel: optimal effective assortativity $a_t^\star(\alpha,\eta)$ (full lines) compared with the non lagged $a_{\alpha,\eta}^t$ (dotted line) and the asymptotic $a_\infty^\star(\alpha,\eta)$ (dashed lines). Black squares indicate the two transitions (see text).

To gather a better intuition, note that for large t

$$a_{\alpha,\eta}^{(\tau)} := \lim_{t\to\infty} a_{\alpha,\eta}^{(t,\tau)} = a\left(\alpha^\tau \frac{1-\alpha}{1-\alpha\eta^2} + (1-\alpha)\eta^2 \frac{\eta^{2\tau}-\alpha^\tau}{\eta^2-\alpha}\right). \tag{4.14}$$

Since $a_{\alpha,\eta}^{(\tau)} \to 0$ as $\tau \to \infty$, when $a_{\alpha,\eta}^{(1)} > a_{\alpha,\eta}^{(0)}$, *i.e.*

$$\delta \equiv \eta^2 \frac{\alpha}{1-\alpha} > 1, \tag{4.15}$$

the maximum of $a_{\alpha,\eta}^{(\tau)}$ is not anymore at $\tau = 0$ (see the top right panel of Figure 4.5).

For finite t, there is a finite size effect since the range of τ is bounded by t. In this situation, for large α and η the maximum of $a_{\alpha,\eta}^{(t,\tau)}$ is achieved at the extremum $\tau = t$ (top panels of Figure 4.5).

Finally, the bottom panel of Figure 4.5 compares $a_t^\star(\alpha,\eta)$, $a_{\alpha,\eta}^t$, and $a_\infty^\star(\alpha,\eta)$. The black squares indicate the two transitions, the first one from zero to positive τ^* (computed with Equation (4.15)) and the second when $\tau^* = t$ due to the finite size effect. These correspond to the transitions observed in the empirical analysis of Figure 4.3.

4.3 Lagged Snapshot Dynamic (LSD) algorithm

In this section, we propose a single snapshot algorithm for the inference of the optimal assignments together with a set of learned model parameters from the observation of a dynamic network. In Section 4.2, we showed how a naive single snapshot inference procedure, applied to a dynamic network with group and link persistence, introduces a systematic bias in the result. This bias takes the shape of a temporal lag: communities inferred at time t share a maximum overlap with planted communities at time $t - \tau^\star$. This can affect also the goodness of the optimal parameters learned from data, for example the measured effective assortativity parameter is systematically overestimated at high link persistency. For this reason we now propose a single snapshot algorithm able to detect and thus correct the possible presence of a temporal lag. Using only observations of the time series $\boldsymbol{A}^t$, we look for a set of inferred parameters $\hat{\eta}$, $\hat{\alpha}$, $\hat{a}$ and group assignments $\hat{\boldsymbol{g}}^t$ using the following scheme, whose details are presented below:

1. for each snapshot we estimate the assortativity and the assignments using a static method (*e.g.* BP on SBM);
2. we estimate the link and group persistence $\hat{\alpha}$ and $\hat{\eta}$ from the sequence of inferred assignments;
3. we compute the optimal lag $\hat{\tau}^*$ in order to obtain an unbiased estimation of the assortativity parameter and the correct assignments at time t by considering the inferred assignments at time $t - \hat{\tau}^*$.

We now detail the three phases of the LSD algorithm.

4.3.1 Single snapshot BP inference

For each snapshot observation $\boldsymbol{A}^t$ we perform the inference from a static SBM, similarly to [Decelle et al., 2011a]. The result is a set of assignment $\boldsymbol{y}^t$ and an effective assortativity $\hat{a^\star}$. As shown in Section 4.2, the use of a static procedure introduces a bias in the result: $\hat{a^\star}$ is a downward biased estimation of the assortativity parameter and $\boldsymbol{y}^t$ is an estimate of the planted assignment sequence but shifted by a lag $\tau^\star$, *i.e.* $\boldsymbol{y}^t = \hat{\boldsymbol{g}}^{t-\tau^\star}$. Clearly at this point $\tau^\star$ is still unknown.

4.3.2 Learning the persistence parameters

The inference of the persistence parameters α and η is performed by maximizing the posterior distribution (4.5). Deriving $\log \mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\mathcal{A}})$ w.r.t. η, we hav

$$\frac{\partial_\eta \log Z(\boldsymbol{\pi})}{NT} = \frac{1}{NT}\mathbb{E}\left(\sum_{i,t}^{T}\frac{\delta_{g_i^t,g_i^{t-1}}-q_{g_i^t}}{\eta\mathbb{I}_{g_i^t,g_i^{t-1}}+(1-\eta)q_{g_i^t}}\right) \tag{4.16}$$

$$= \mathbb{E}\left(\frac{1}{NT}\sum_{i,t}^{T}\sum_{a,b}\delta_{g_i^t a}\delta_{g_i^{t-1}b}\frac{\delta_{a,b}-q_a}{\eta\delta_{a,b}+(1-\eta)q_a}\right) \tag{4.17}$$

$$= \mathbb{E}\left(\sum_{a=1}^{k}\frac{1-q_a}{\eta+(1-\eta)q_a}f_{g_i^t=g_i^{t-1}=a}-\frac{1}{1-\eta}f_{g_i^t\neq g_i^{t-1}}\right)=0 \tag{4.18}$$

where f_E means the empirical frequency of an event E over space and time. The quantity inside the expectation in (4.16) is exactly what we would obtain by fitting a given observed assignment $\vec{\boldsymbol{g}}$ with a Markov chain. The difference is that now it is averaged over the posterior. As a first approximation, assuming the posterior to be peaked around $\hat{\boldsymbol{y}}$, *i.e.* the assignments inferred from the single snapshot procedure, we can simply find the solution $\eta=\hat{\eta}$ of the polynomial equation

$$\sum_{a=1}^{k}\frac{1-q_a}{\eta+(1-\eta)q_a}f_{\hat{y}_i^t=\hat{y}_i^{t-1}=a}-\frac{1}{1-\eta}f_{\hat{y}_i^t\neq\hat{y}_i^{t-1}}=0. \tag{4.19}$$

Similarly, deriving $\log\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\mathcal{A}})$ with respect to α, we have

$$\frac{2\partial_\alpha \log Z(\boldsymbol{\pi})}{N(N-1)T} = \mathbb{E}\left(\frac{2}{N(N-1)T}\sum_{i,j>i}^{}\sum_{t=1}^{T}\frac{\delta_{A_{ij}^t,A_{ij}^{t-1}}-p_{g_i^tg_j^t}^{A_{ij}^t}(1-p_{g_i^tg_j^t})^{1-A_{ij}^t}}{\alpha\delta_{A_{ij}^t,A_{ij}^{t-1}}+(1-\alpha)p_{g_i^tg_j^t}^{A_{ij}^t}(1-p_{g_i^tg_j^t})^{1-A_{ij}^t}}\right) \tag{4.20}$$

$$= \mathbb{E}\left(\frac{2}{N(N-1)T}\sum_{i,j>i}^{}\sum_{t=1}^{T}\sum_{a,b}\sum_{\varepsilon,\varepsilon'}^{0,1}\delta_{g_i^ta}\delta_{g_j^tb}\delta_{A_{ij}^t\varepsilon}\delta_{A_{ij}^{t-1}\varepsilon'}\frac{\delta_{\varepsilon,\varepsilon'}-p_{ab}^{\varepsilon}(1-p_{ab})^{1-\varepsilon}}{\alpha\delta_{\varepsilon,\varepsilon'}+(1-\alpha)p_{ab}^{\varepsilon}(1-p_{ab})^{1-\varepsilon}}\right)$$

$$= \mathbb{E}\left(\sum_{a,b}\sum_{\varepsilon,\varepsilon'}^{0,1}\frac{\delta_{\varepsilon,\varepsilon'}-p_{ab}^{\varepsilon}(1-p_{ab})^{(1-\varepsilon)}}{\alpha\delta_{\varepsilon,\varepsilon'}+(1-\alpha)p_{ab}^{\varepsilon}(1-p_{ab})^{(1-\varepsilon)}}m_{\varepsilon'\to\varepsilon}^{ab}(\boldsymbol{g})\right)=0 \tag{4.21}$$

having the same structure of (4.16), averaged over the posterior (4.4) and where we have introduced the quantities

$$m_{\varepsilon'\to\varepsilon}^{ab}(\boldsymbol{g})=\frac{2}{TN(N-1)}\sum_{t=1}^{T}\sum_{(i,j)}\delta_{g_i^t,a}\delta_{g_j^t,b}\delta_{A_{ij}^{t-1},\varepsilon'}\delta_{A_{ij}^t,\varepsilon}. \tag{4.22}$$

Again, as soon as the posterior is concentrated around a set of inferred assignments $\hat{y}$, we can simply find the solution $\alpha=\hat{\alpha}$ of the equation

$$\sum_{a,b=1}^{k}\sum_{\varepsilon,\varepsilon'}^{0,1}\frac{\delta_{\varepsilon,\varepsilon'}-p_{ab}^{\varepsilon}(1-p_{ab})^{(1-\varepsilon)}}{\alpha\delta_{\varepsilon,\varepsilon'}+(1-\alpha)p_{ab}^{\varepsilon}(1-p_{ab})^{(1-\varepsilon)}}m_{\varepsilon'\to\varepsilon}^{ab}(\hat{\boldsymbol{g}})=0 \tag{4.23}$$

Note that as soon as we use the inferred assignment instead of the full posterior distribution, Equations (4.19) and (4.23) are not coupled, thus $\hat{\alpha}$ and $\hat{\eta}$ can be obtained independently. It is worth noticing that the presence of a temporal lag does not affect the result of learning link and group persistences even if we use $\boldsymbol{y}^t=\hat{\boldsymbol{g}}^{t-\tau^\star}$ instead

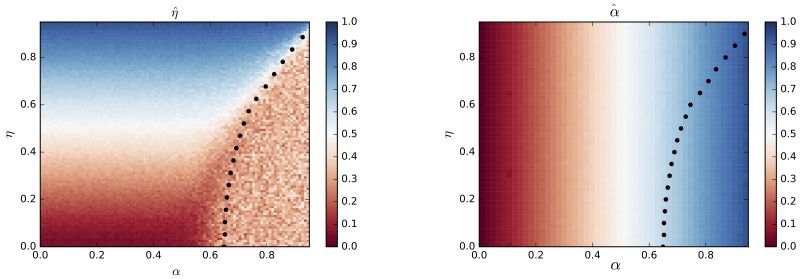


Fig. 4.6 Learning η and α from synthetic dynamic networks generated according to the DSBM with group and link persistence. We use $T = 50$ snapshots of networks with $N = 300$ nodes, $k = 2$ equally sized evolving communities and planted parameters η, α, $a = 1.0$, $\bar{c} = 10$. The panels show the learned $\hat{\eta}$ and $\hat{\alpha}$ as function of the planted η, α. They coincide at least until the detectability transition line (black dots), where the overlap $q(\hat{\boldsymbol{g}}, \boldsymbol{g})$ between inferred and planted assignments vanishes (see top left panel of Fig. 4.7.

of $\hat{\boldsymbol{g}}^t$. This is because asymptotically, at large t, the lag is constant, thus preserving the ordering, and the procedure bias can be considered as just a uniform shift over the inferred communities. At the same time, Equations (4.19-4.23) work as soon as a sequence of consecutive assignments is considered. In the next subsection we numerically test this procedure to infer the persistence parameters.

4.3.3 Lagged inference

Starting from the estimates $\hat{\alpha}$ and $\hat{\eta}$, we can obtain an estimate of the asymptotic optimal lag as

$$\hat{\tau}^\star = \text{argmax}_\tau \left(\hat{\alpha}^\tau \frac{1-\hat{\alpha}}{1-\hat{\alpha}\hat{\eta}^2} + (1-\hat{\alpha})\hat{\eta}^2 \frac{\hat{\eta}^{2\tau} - \hat{\alpha}^\tau}{\hat{\eta}^2 - \hat{\alpha}} \right), \tag{4.24}$$

from which we can shift back the inferred assignments $\hat{\boldsymbol{g}}^{t-\hat{\tau}^\star} = \hat{\boldsymbol{y}}^t$ and correct the effective learned assortativity $\hat{a^\star}$ to

$$\hat{a} = \hat{a^\star} \left(\hat{\alpha}^{\hat{\tau}^\star} \frac{1-\hat{\alpha}}{1-\hat{\alpha}\hat{\eta}^2} + (1-\hat{\alpha})\hat{\eta}^2 \frac{\hat{\eta}^{2\hat{\tau}^\star} - \hat{\alpha}^{\hat{\tau}^\star}}{\hat{\eta}^2 - \hat{\alpha}} \right)^{-1}. \tag{4.25}$$

We perform extensive numerical simulations to test the effectiveness of the LSD algorithm. Before showing the results of the full LSD, we first test step 2 of the algorithm, which estimates the persistence parameters from the (biased) estimation of the assignments. Figure 4.6 shows the result of learning η and α from (4.19) and (4.23) using the assignment $\hat{\boldsymbol{y}}$ from the single snapshot procedure. The learned

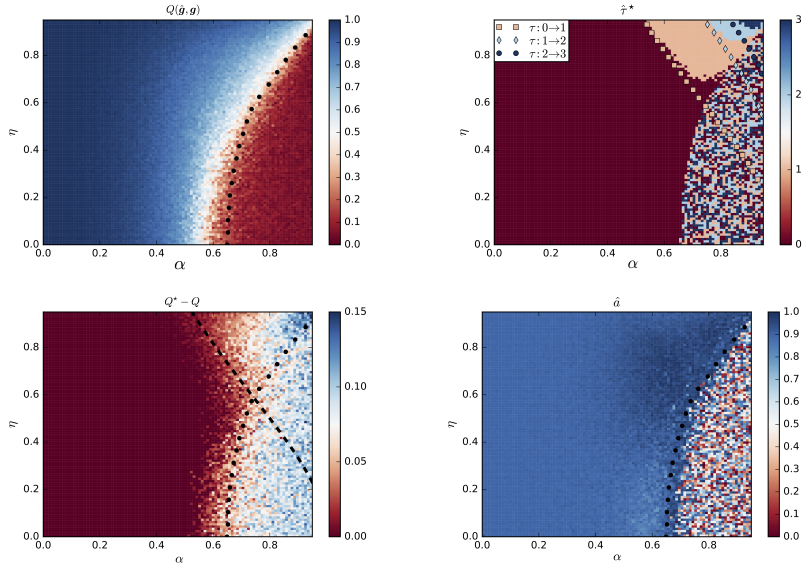


Fig. 4.7 Performances of the LSD algorithm on synthetic networks. We use $T = 50$ snapshots of networks with $N = 300$ nodes, $k = 2$ equally sized evolving communities and planted parameters η, α, $a = 0.9$, $\bar{c} = 10$. *Top left panel:* overlap between the inferred and planted assignments. *Top right panel:* optimal inferred lag $\hat{\tau}^\star$ from Equation 4.24 and comparison with the analytical transition lines (diamonds) signing lag jumps. *Bottom left panel:* difference between $Q^\star = Q(\hat{\boldsymbol{g}}^t, \boldsymbol{g}^t)$ and $Q = Q(\boldsymbol{y}^t, \boldsymbol{g}^t)$, *i.e.* after and before the time lag correction. There is a net positive gain to the right of the dashed line, indicating the first jump from zero to nonzero optimal lag at $\eta^2\alpha/(1-\alpha) = 1$. *Bottom right panel:* learned assortativity $\hat{a}$ as function of η and α. In all panels black dotted line is the detectability transition line $a^\star_\infty(\alpha, \eta) = \bar{c}^{-1/2}$.

parameters $\hat{\eta}$ and $\hat{\alpha}$ are in agreement with the planted ones, at least as soon as the overlap between the planted and inferred communities is far from zero.

We then test the performances of the LSD procedure against synthetic dynamic networks generated according to the DSBM with group and link persistence. We use $T = 50$ snapshots of networks with $N = 300$ nodes, mean degree $\bar{c} = 10$, $k = 2$ equally sized evolving communities and a wide range of planted parameters η, α, a. In the top left panel of Figure 4.7 we show the the overlap $Q(\hat{\boldsymbol{g}}, \boldsymbol{g})$ between planted and inferred assignments as a function of η, α. For a large region of the parameter space the overlap is very high, showing that the LSD algorithm is able to recover the planted assignment. The black dots indicate the detectability transition line of equation $a^\star_\infty(\alpha, \eta) = \bar{c}^{-1/2}$. As expected in the region to the right of this line the overlap is very small. The top right panel shows the estimated value of $\hat{\tau}^\star$ as a function of the persistence parameters. Notice that here the top right corner is the region where lagged inference is fundamental for the correct inference of labels. In fact, the bottom left panel shows $Q(\hat{\boldsymbol{g}}^t, \boldsymbol{g}^t) - Q(\boldsymbol{y}^t, \boldsymbol{g}^t)$ to highlight the role of time shift in assignment inference. As expected, the region where time shift is critical is the one where $\hat{\tau}^\star$ is different from zero. The transition line between these two region

is described by $\eta^2\alpha/(1-\alpha)=1$ (dashed line). Finally the bottom right panel shows the inferred $\hat{a}$, which in the detectability region is always very close to the planted value $a=0.9$.

4.4 Comparison with a full dynamic inference

In this section we compare the LSD algorithm, that is a single-snapshot based algorithm, with a fully dynamic algorithm, *i.e.* that uses at once the information of the whole time series of network snapshots and the dynamic rules of the generating process. It is a suitable modification of the dynamic algorithm by [Ghasemian et al., 2016], described also in Subsection 1.3.3. This method is based on the observation that if one was able to detect and remove links that have been copied from the past and not generated according to the communities at the present, then it would be as the resulting temporal network was generated according to DSBM but with zero link persistence, for which the dynamic BP algorithm introduced in [Ghasemian et al., 2016] is thought to be optimal. Thus the second method is based on the manual detection and subsequent remotion of links in the snapshot at time t that appears both in the snapshots at time t and $t-1$ (that in the sparse regime are with high probability links that have been copied from the past) followed by the model inference with the dynamic BP algorithm.

The comparison in this section concerns the detectability regions of the two procedures that can be found as suitable modification of the detectability region of a model without link persistence, see [Ghasemian et al., 2016] and Subsection 1.3.3, being in this case the detectability threshold equal to

$$a \geqslant \frac{1}{\sqrt{c}}\sqrt{\frac{1-\eta^2}{1+\eta^2}}, \tag{4.26}$$

where $a=(1-\varepsilon)/(1+\varepsilon)$ is the assortativity parameter, $\varepsilon = c_{out}/c_{in}$, and c is the mean degree. In absence of group persistence, we retrieve the static threshold $a \geqslant 1/\sqrt{c}$.

Case $\eta = 0$

Let's now start by considering the detectability performance of the two methods in the case of a dynamic network with $\alpha > 0$ and $\eta = 0$. According to the LSD algorithm, the role of the link persistence is to decrease the effective assortativity of the network, from a to $a(1-\alpha)$. Thus

$$a(1-\alpha) \geqslant \frac{1}{\sqrt{c}} \implies a \geqslant \frac{1}{\sqrt{c}(1-\alpha)}. \tag{4.27}$$

On the other hand, according to the modified method by Ghasemian et al., the network resulting from the cleaning of persistent links is statistically equivalent to a network generated through the same SBM, but with an effective degree that is smaller, from c to $c(1-\alpha)$. Thus the detectability region of this method is expected to be

$$a \geqslant \frac{1}{\sqrt{c(1-\alpha)}} = \frac{\sqrt{1-\alpha}}{\sqrt{c}(1-\alpha)} \tag{4.28}$$

that is larger than the one of the LSD algorithm. The conclusion is that, in both cases, link persistence affects the detectability, reducing the effective assortativity in the first case and the effective degree in the second case. Both outcomes make the inference harder. Nevertheless detecting and removing the source of the noise, at the cost of reducing the effective degree of the network, is better than leaving random uninformative links reducing the effective assortativity.

General case $\eta > 0$

In this case the presence of the group persistence mitigates the negative effect of the link persistence. In the LSD algorithm again we need just to consider the effective assortativity that in this case will be $a(\alpha,\eta)$: for example, outside from the lagged region $a(\alpha,\eta) = a(1-\alpha)/(1-\alpha\eta^2)$. Thus the detectable threshold associated with the LSD algorithm becomes

$$a\frac{1-\alpha}{1-\alpha\eta^2} \geqslant \frac{1}{\sqrt{c}} \implies a \geqslant \frac{1-\alpha\eta^2}{(1-\alpha)\sqrt{c}} \tag{4.29}$$

that is increasing with η. In the region of lagged inference, the bias correction allows to extend the region of detectability. This is captured by $a(\alpha,\eta)$ (in this region, the expression of $a(\alpha,\eta)$ is a bit more involved).

In the case of the full dynamic method, again we have a decrease in the effective degree but now compensated by a net gain in the detectability because of the group persistence, which is better exploited by the dynamic BP algorithm. Using (4.26), the detectability threshold associated with this method is thus

$$a \geqslant \frac{1}{\sqrt{c(1-\alpha)}}\sqrt{\frac{1-\eta^2}{1+\eta^2}}. \tag{4.30}$$

Since $1-\alpha\eta^2 \geqslant \sqrt{(1-\eta^2)(1-\alpha)/(1+\eta^2)}$, the detectable region of the LSD is always smaller. To visualize the difference it is possible to plot the critical line in the phase diagram (α,η), see Figure 4.8, keeping fixed c and a. Inverting the previous relations we get respectively for LSD and the full dynamic method

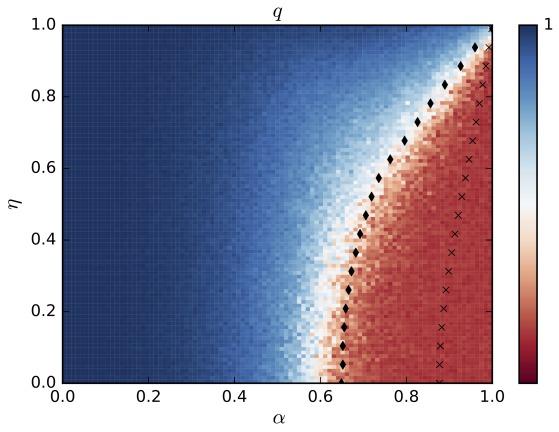


Fig. 4.8 Overlap after the LSD inference in terms of the two persistencies ($T = 50$, $N = 300$, $a = 1.0$, $\bar{c} = 10$ and $k = 2$ equally sized evolving communities). Squares and crosses represent the detectability lines respectively of the LSD algorithm and the full dynamic algorithm based on [Ghasemian et al., 2016].

$$\alpha \leqslant \frac{\sqrt{c}a-1}{\sqrt{c}a-\eta^2} \quad \text{and} \quad \alpha \leqslant 1-\frac{1}{a^2c}\frac{1-\eta^2}{1+\eta^2}. \tag{4.31}$$

In particular, as soon as $a \geqslant 1/\sqrt{c}$, both the critical link persistencies tend to 1 when $\eta \to 1$, being maximally different at $\eta = 0$. On the contrary, when $a \leqslant 1/\sqrt{c}$, the LSD algorithm is always in the undetectable region while the other one is always in the detectable region for η close enough to 1. The same comparison can be expressed in terms of the minimal average degree for detectability: again inverting (4.29) and (4.30), we get

$$c \geqslant \frac{1}{(1-\alpha)a^2}\frac{(1-\alpha\eta^2)^2}{(1-\alpha)} \quad \text{and} \quad c \geqslant \frac{1}{(1-\alpha)a^2}\frac{1-\eta^2}{1+\eta^2}, \tag{4.32}$$

for both LSD and the full dynamic method, respectively,

We can conclude that both procedures are able to leverage the group persistence to increase the detectability performances even in the presence of link persistence (for any α the communities are detectable for η large enough): the full dynamic algorithm, obtained by modifying the dynamic BP by Ghasemian et al. , makes use of the whole temporal network, the LSD procedure uses only the information of a single snapshot enriched by the information on past communities codified in the copied links. As expected the LSD procedure cannot be optimal but can be considered as an interesting tradeoff between a full dynamic inference that uses the whole time series history, accurate but expensive, and a naive single snapshot inference, blind to the dynamics but simpler. Moreover it can be also considered faster and more flexible in the perspective of online inference, in which, given all

the information up to t, the analysis of a single new snapshot at time $t+1$ is required without examining the entire past. It is faster because in this context the complexity of a full dynamic inference would be $O(TN)$ for each new snapshot, instead of $O(N)$ for a single snapshot inference and $O(\tau N)$ for LSD with time lag correction. It is more flexible because real temporal networks are typically not stationary, and a localized-in-time procedure can be desirable for a better tracking of the dynamics.

Conclusions

In this chapter, we have studied the problem of community detectability in a temporal network model where both communities and links are time varying. We have focused on static algorithms for temporal networks, where inference is performed on each snapshot network and found that link persistence is the driver of a new kind of detectability transition bringing to time lagged inference, *i.e.* a bias towards the detection of past communities. By analyzing static detection of dynamic communities, we have been able to introduce a time-lag corrected procedure, the lagged snapshot dynamic (LSD) algorithm, able to correct the bias thus outperforming naive single snapshot inference. This algorithm leaves room for improvement from new algorithms that, using the information given by the full temporal network, might reach optimality still maintaining the flexibility of a single-snapshot based approach. Suitable generalization of a single snapshot based procedure is necessary for online learning techniques where only few snapshots in the past (τ in the LSD case), instead of the whole time series, are necessary to get the assignments of a new network snapshot. This is particularly relevant especially if we aim at studying real networks whose dynamics is usually far from being stationary.

Part II
Feedback dynamics in financial networks

Here I say nothing more than that: people with no memory of the past are always in trouble.

Chapter 5
Networks in finance

Abstract In this chapter, we review some network models which focus on the study of systemic risk in financial networked systems. In particular, we show that the contagion of risk in financial networks displays two channels of propagations: direct contagion because direct exposures between counterparties and indirect contagion mediated by fire sale spillovers and overlapping portfolios. We point out that institutions operating in financial markets react actively when exposed to risk and portfolio decisions governs the dynamics of the financial system. Hence, we highlight what are the drivers of portfolio decisions, such as investment strategies, capital requirements, and, last but not least, expectations by financial institutions. Finally, we review the model of [Corsi et al., 2016], that is an analytical model of indirect contagion of risk.
This represents an introductory chapter to the research presented in Chapters 6 where we study the relation between expectation feedbacks and systemic risk from the point of view of dynamical stability of financial systems.

Introduction

The financial crisis of 2007-2008 made evident that any financial system exhibits today a high degree of interdependence among the subparts which compose the whole.

When looking at the many financial institutions operating in a financial market, it is evident that the set of financial instruments which are bought and sold between them tends to create an highly complex structure of interrelations, giving rise to a complex network of interactions. There are several possible sources of connections between financial institutions, which can be classified according to the side of the balance sheet they belong to, *i.e.* the asset or the liability side. For instance, banks are directly connected through mutual exposures and credit relations, formed in the interbank market. Moreover, holding similar portfolios or sharing the same invest-

ments creates indirect linkages between financial institutions. Networks can be a useful representation of financial systems.

One can usually consider two kinds of financial networks. First, a *unipartite* financial network describes any collection of financial institutions (*i.e.* the *nodes* of the network), such as traders, firms, banks and financial exchanges, and the *links* between them, representing all possible financial instruments that they buy or sell one from each other. Second, a *bipartite* financial network accounts for two types of nodes, financial institutions and financial instruments, and weighted links describe the composition of portfolios, each link representing the weight of that specific instrument in the portfolio of the institution. In effect, any financial network is a weighted network, *i.e.* each link is associated with a weight, that is the value of the financial investment. For instance, the interbank market is the network where banks borrow and lend money for liquidity management, each credit representing a directed link whose weight is the portion of the asset side in the giver's balance sheet, thus corresponding to a same portion of the liability side in the taker's balance sheet. Instead, an example of bipartite financial system is the network of security holdings, *e.g.* stock of publicly traded companies, where a firm's ownership of stock would represent a link between the firm and the stock.

As pointed out in the recent literature, *e.g.* see [Allen and Babus, 2009], network analysis may help to address two types of research questions: (i) network effects and (ii) network formation and evolution.

The first type of questions studies processes that take place on fixed networks. For instance, the resilience of a banking system to contagion can be evaluated according to the network structure that connects financial institutions. Recently, different authors have pointed out the micro-foundations of financial contagion [Allen and Gale, 2000], how to quantify the probability of systemic contagion and its extension given an initial shock [Gai and Kapadia, 2010], how different network structures respond differently to the propagation of a shock [Hurd et al., 2017], and how the fragility of the financial system depends on the location in the network of the institution that was initially affected [Caccioli et al., 2012, Amini et al., 2016], to name but a few. Network science also gives some tools to quantify how much a financial institution is subject to financial contagion, thus resulting more or less central in the propagation of financial distresses [Battiston et al., 2012]. All mechanisms that can lead to the breakdown of the financial network create financial *systemic risk*, and the the understanding of them is mainly based on network approaches, such as model of default cascades due to bilateral exposures or to overlapping portfolios.

Then, financial institutions are active players in the market, *i.e.* they manage actively their portfolios, especially in periods of financial distress. In fact, there are many empirical results, *e.g.* [Mendoza and Terrones, 2008, Schularick and Taylor, 2012], confirming that the balance sheet dynamics of financial institutions, far from being passive and exogenous, is instead the "endogenous engine" that drives the boom-bust cycles of funding and liquidity and hence the dynamics of financial networked systems. In particular, [Adrian and Shin, 2010] firstly recognized that financial institutions adjust their asset side rather than raising or redistributing equity capital in or-

der to target the value of financial leverage: buying or selling assets in financial markets, thus expanding or contracting the asset side rather than the level of equity, has a positive feedback on the price of the investments and hence on the risk itself. This effect is further amplified by the current financial policy which imposes to financial institutions Value-at-Risk constraints [Adrian and Shin, 2013]. This mechanism together with common accounting practice and risk management rules determine how financial networks evolve in time [Choi and Douady, 2012, Corsi et al., 2016]. Last but not least, any financial system is an expectation feedback system, *i.e.* the current decisions of players depends on what is expected to happen in the future [Hommes, 2013]. For instance, in periods of turmoil banks may have no confidence on the creditworthiness of some counterparties. When the *perceived* risk associated with the lending of funds to those counterparties is too high, links become too 'costly' relative to the benefits they bring, thus resulting in freezes of the interbank market. In cases like this, the dynamics is driven by the expectations of the banks. Notice, however, that the expectation does not always correspond to the future realization when the dynamics is endogenously determined by the aggregate behavior of the whole (except that all players are assumed to be rational). As a consequence, the process of expectations' formation becomes one of the most important drivers for the evolution of the system.

Hence, the second type of questions studies how financial institutions form and modify connections. New insights in the issue of systemic risk can be gained by understanding how financial institutions react when exposed to the risk of contagion. Moreover, understanding the dynamics of financial networks during risky periods may help explaining phenomena such as flight-to-quality or flight-to-liquidity [Brunnermeier and Pedersen, 2008]. Studying the drivers of such events could lead to preventing new occurrences in the future, thus preserving the stability of the financial system.

This is organized as follows: in Section 5.1 we review some milestone works on the contagion of risk in financial networks and in Section 5.2 we study what the drivers of the dynamics are, when a financial system is exposed to systemic risk. Then, in Section 5.3 we review the model of [Corsi et al., 2016], which is the baseline model of systemic risk in financial bipartite networks for the research presented in the next chapter.

5.1 Network model approach to systemic risk

When studying the process of propagation of risk in a financial system, many scientific works recognize two kinds of mechanisms: *direct* and *indirect* contagion of risk. The first one occurs when the insolvency of an institution directly affects the counterparties because of the presence of bilateral exposures. This mechanism can be studied by focusing on the unipartite network of exposures between financial institutions. However, the contagion of risk could be also mediated by common portfolio investments, *i.e.* in the presence of investments' illiquidity, distressed selling by an

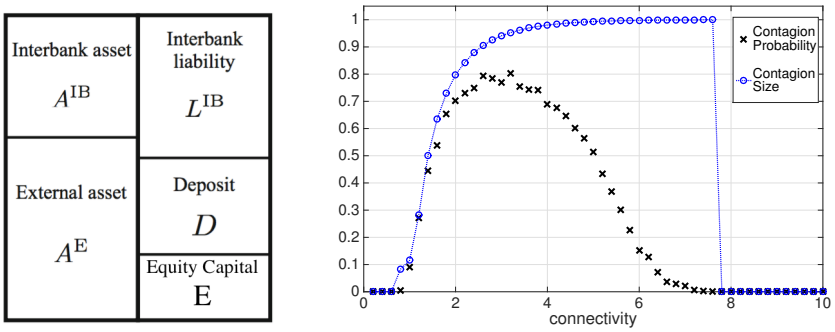


Fig. 5.1 Pictorial representation of the balance sheet of a financial institution (left) and the phase diagram associated with the Gai-Kapadia model representing both the probability of a global cascade and its size as a function of the connectivity of the network (right).

institution causes some depreciation which affects all the other institutions having those assets in their portfolios.

Regarding the direct contagion of risk, the milestone work of [Gai and Kapadia, 2010] has shown that systemic risk can be evaluated in terms of default cascades in financial networks, especially interbank networks in which banks lend to and borrow from each other. The basic structure of the Gai-Kapadia cascade model is based on the Watts model of global cascades [Watts, 2002]. This model describes the mechanism of how a node becomes 'infected' (*e.g.* 'defaulted' in the case of financial networks) depending on the fraction of its neighbors which are already 'infected'.[1] The main finding of the Watts model is that, on randomly connected networks, even a vanishingly small fraction of initial infected nodes may lead a significant fraction of the total nodes to become infected as long as the network is not too sparse or too dense.

In the case of financial contagion spreading in the interbank market, nodes represent banks and direct weighted links describe lending-borrowing relationships, respectively. In the model of [Gai and Kapadia, 2010], the network of connections is assumed to be a Erdös-Rényi random graph fully characterized by the parameter of connectivity. Then, an infection is interpreted as a bank default. Hence, the default condition is described in terms of positive capital of a bank. In fact, a simple stylization of the balance sheet (see the left plot of Figure 5.1) considers two types of assets: interbank assets, A^{IB}, and external assets A^{E} (such as stocks, bonds, etc.). Gai and Kapadia assume that interbank assets are evenly distributed over the number of links (*i.e.* the degree or the connectivity) of each bank-node. On the liability side, there can be interbank liabilities, L^{IB}, and deposits from customers, D. Then, the solvency condition for bank i is given by

$$A_i^{IB} + A_i^{E} - L_i^{IB} - D_i > 0,$$

[1] This model is in effect a *(linear) threshold model* for random networks.

which is equivalent to say that the equity capital of the bank should be positive. Since an asset investment of a bank is a liability for another bank, the initial default of a node (or a small set of nodes) represents a loss in the asset size for the counterparties, that may result in the insolvency and the subsequent default. Under the (simplifying) assumptions of the model, the default condition for bank i is

$$\phi_i > \frac{E_i}{A_i^{IB}}$$

where E_i is the equity and ϕ_i is the fraction of bank i's counterparties that have defaulted. The contagion of risk can thus propagate.

Then, within the model the probability of a systemic contagion and its size can be quantified by looking at the phase diagram as a function of the connectivity, see the right plot of Figure 5.1. There exists two thresholds for the connectivity: (i) above the lower one, both the probability of a systemic event and its global impact start to increase, both reaching some maxima at given values; (ii) then, there exists a second threshold which triggers the disappearing of systemic risk because the asset investments are so much diversified such that the default of a bank does not affect crucially the balance sheet of the counterparties. It is interesting to notice that below this second threshold there is a region characterized by a small probability of systemic risk, but with the size of contagion close to one, *i.e.* the whole network. This is the so-called *robust-yet-fragile* nature of the financial system, *i.e.* it is highly likely that an initial shock is absorbed, but if it is not, then the contagion spreads through the whole system. The phase diagram can be also analytically characterized by exploiting a general treelike approximation method for solving models of contagion in sparse static graphs [Gleeson and Cahalane, 2007], thus generalizing, *e.g.* , the study of cascades on networks with community structure or degree-degree correlation [Gleeson, 2008]. Regarding financial contagion, many extensions have been proposed to relax some crucial approximations in the Gai-Kapadia model: (i) loans are uniformly distributed (see [Hurd and Gleeson, 2013] for the generalization), (ii) credit relations create a Erdös-Rényi random graph (*e.g.* see [Dodds and Payne, 2009, Hurd et al., 2017] for possible extensions), (iii) default is based on a linear threshold mechanism (see [Petrone and Latora, 2018] for a model combining the contagion mechanism with a more realistic process for the bank default), and (iv) the risk of external assets is not considered.

In fact, the only contagion channel considered in the Gai-Kapadia model is the cascade of losses due to direct exposures between banks (or, more generally, financial institutions). However, this mechanism represents only one contribution to the systemic risk of a financial network. From the empirical point of view, contagion analysis of real interbank systems have shown domino effects triggered by the failure of a small number of banks are unlikely to occur in practice [Upper, 2011]. On the other hand, [Caccioli et al., 2015] have shown that financial networks can significantly amplify distress propagation in the presence of other contagion channels, such as for instance fire sales and overlapping portfolios. Indeed, the contagion may be propagated through overlapping portfolios, in which the falling of the price of

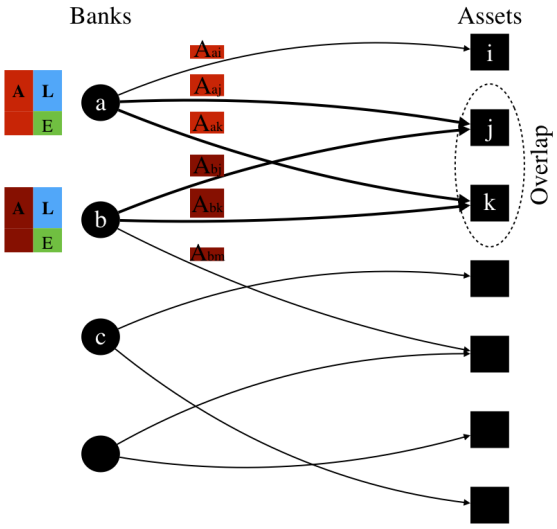


Fig. 5.2 Stylization of a bipartite financial network where one set of nodes (banks) connect to the other set of nodes (asset investements), and the link's weight represents the value of the asset investment.

an 'external' asset will affect simultaneously all institutions having that investment in their portfolios [Cifuentes et al., 2005, Huang et al., 2013, Caccioli et al., 2014]. This is supported also empirically by [Gualdi et al., 2016], which have introduced a method to assess the statistical significance of the overlap between heterogeneously diversified portfolios and have pointed out that the 'validated' overlap among portfolios of security holdings by US financial institutions has displayed a steady increase before the 2007-2008 financial crisis and reached a maximum when the crisis occurred. Finally, when assets are correlated (or there exists some mechanism turning originally uncorrelated assets into correlated ones, especially during periods of turmoil) fire sale spillover effects mediated by overlapping portfolios may be even more amplified by the diversification of portfolios [Beale et al., 2011].

The 'microscopic' mechanism capturing the indirect contagion of risk has been described by [Cont and Wagalath, 2013] which consider the (linear) price impact of distressed selling of two assets in an illiquid financial market, *i.e.* a market where buying or selling assets moves endogenously their price. Cont and Wagalath show that distressed selling has a positive feedback effect on both assets' volatility and correlation, resulting in the increase of the systematic component of risk which can not be diversified, thus affecting the overlapped portfolios. This 'microscopic' behavior turns into systemic risk when aggregated at the network level.

At the systemic level, the process of indirect contagion has been studied in [Huang et al., 2013] which describe the financial system as a bipartite network of banks and asset investments, see Figure 5.2 for a stylized representation. Similarly to the threshold model of Gai-Kapadia, Huang *et al.* study the indirect propagation

of risk conditioned to the default of one bank, followed by the liquidation of its asset position. Because of the linear price impact of distressed selling, contagion may propagate to other banks having some investments in common with the defaulted counterparty. Huang *et al.* characterize the stability of the system in relation to the illiquidity of the market and the entity of the initial shock. A similar model of overlapping portfolios has been introduced by [Caccioli et al., 2014], which focus more on the role of both leverage and diversification of portfolios: the financial leverage is defined as the ratio between the asset and the equity value, while the portfolio diversification, in this setting, can be measured as the number of links incident to the node. Caccioli *et al.* find there exists a region of instability for the financial system where it is likely to observe a global cascade. This region is characterized by a given threshold value for the financial leverage, below which no global cascades are observed whatever the value of diversification. In the region of instability, the probability of a global contagion and its size follows a non-monotonic behavior as a function of the diversification (*i.e.* the connectivity of the bipartite network), similarly to what observed in the Gai-Kapadia model (see the right plot of Figure 5.1), thus revealing again the robust-yet-fragile nature of the financial system.

5.2 Drivers of financial system dynamics

Financial institutions are active players in financial markets: they manage their portfolios according to their strategies and react when exposed to risk. As discussed before, default cascades are unlikely to be observed, but financial distresses may propagate through 'external' assets because of distressed selling and overlapping portfolios. Hence, when studying how a financial network evolves in time, it is important focusing on its bipartite structure, *i.e.* financial institutions (such as banks) investing in some risky assets (such as bonds, stocks, etc.), and identifying what the main drivers of the dynamics are.

As found empirically [Mendoza and Terrones, 2008, Schularick and Taylor, 2012], balance sheet aggregates drive endogenously the cycles of funding and liquidity in financial markets. [Danielsson et al., 2012] show further that while the seeds of the volatility are *exogenous*, a large part of the risk in periods of distress is due to amplifying effects within the system. That is, systemic risk is *endogenous*.

The balance sheet of a financial institution has two sides: assets on the left, and financing, which itself has two parts, liabilities and equity, on the right (*e.g.* see the left plot of Figure 5.1). Borrowing is an essential aspect of the business of financial institutions operating in financial markets because of the possibility of leveraging the investment returns by buying on margin. In the simplest description, the financial leverage is defined as the ratio between the asset value and the equity. It is a measure of the debt: in fact, given the equity value, the asset size is determined by the level of debts. Leveraging enables gains to be amplified by a factor equal to the value of the financial leverage. On the other hand losses are amplified too. That is, when the

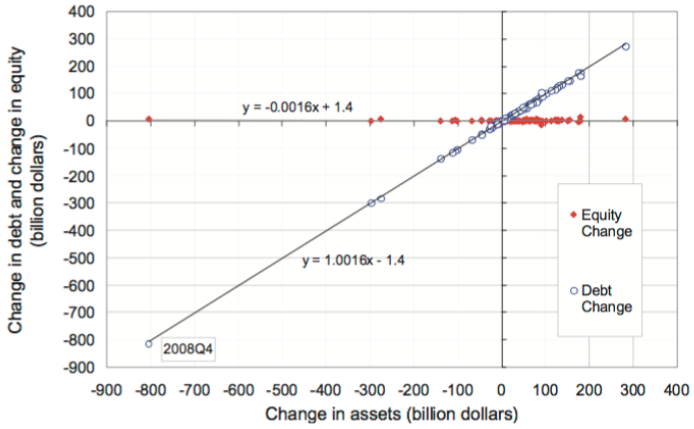


Fig. 5.3 Scatter plot of changes in debt and equity to changes in assets of the US broker dealer sector (1990Q1-2012Q2) (Source: [Adrian and Shin, 2013]).

system is interconnected, financial leverage makes it more susceptible to shocks, thus contributing to systemic risk.

The recent empirical literature [Adrian and Shin, 2010, Adrian and Shin, 2013] perfectly captured the leverage management of the largest investors operating in the market. The main players of financial markets tend to adopt the strategy of *leverage targeting*, *i.e.* they prefer increasing or decreasing the debt level rather than raising or redistributing equity capital, thus buying and selling assets. For instance, Figure 5.3 shows the relationship between changes in asset size and changes in both debt and equity for the US broker dealer sector in the period 1990Q1-2012Q2, thus confirming empirically the active management of the liability side. However, in the presence of investments' illiquidity, buying and selling assets to target the financial leverage may create a stronger interdependence among investments, thus contributing to increase the correlation endogenously , similarly to what was shown by [Cont and Wagalath, 2013] in the case of distressed selling.

Regulators try to limit the exacerbated use of leverage by imposing several constraints on the level of debts an institution can adopt, in order to make the financial system more robust and resilient to shocks. Value-at-Risk (VaR) constraint is probably the most popular one, but other more sophisticated ones have been proposed in recent years by the successive Basel regulations. In its simplest form, the VaR constraint determines the maximum value of the asset side (or equivalently, the liability side) given the equity and the risk of the portfolio (*i.e.* the *VaR*),

$$VaR_\alpha \times A \leqslant E, \tag{5.1}$$

with A and E the total asset value and the equity value, respectively. The Value-at-Risk VaR_α is defined as the α-quantile associated with the distribution of the portfolio return for a given time horizon, *i.e.* the worst expected loss with probability

α.[2] In other terms, an institution need to cover a loss VaR_α with its own equity capital. Hence, the larger is the portfolio risk (quantified by the Value-at-Risk), the smaller has to be the value of the investment, thus the level of debts.

However, VaR capital requirements, as other risk constraints, can induce a perverse demand function: in order to target leverage, a financial institution will sell more assets if their price drops and viceversa when their price rises. Thus, a marked-to-market and VaR constrained financial institution will have a positive feedback effect on the prices of the assets in its portfolio. This is studied, *e.g.* , in [Aymanns and Farmer, 2015] where the authors present a simple agent-based model of a financial system composed of leveraged investors such as banks that invest in stocks and manage their risk using a Value-at-Risk constraint, based on historical observations of asset prices. The Value-at-Risk constraint implies that when perceived risk is low, leverage is high and vice versa: a fact that is known as *procyclicality* of the Value-at-Risk constraint, which gives rise to the so-called *leverage cycles* [Geanakoplos, 2010], *i.e.* periods in which gradual increases in prices and leverage are followed by drastic market collapses.[3] Since the appearing or leverage cycles is related to the adopted regulation, *countercyclical* policies have been proposed to mitigate this feedback effect [Aymanns et al., 2016].

Furthermore, in a market where many financial institutions are forced to follow similar risk management rules, the coordinated rebalancing of portfolios composed by illiquid assets reinforces the feedback effect of the leverage targeting depending on the degree of the diversification of portfolios. Thus, while diversification of investments should reduce portfolio risk, a significant overlap (*i.e.* similarity) of portfolios of many financial institutions can instead destabilize the market and increase its susceptibility to price shocks because of fire sale spillovers, *i.e.* when asset sales depress prices, in which case one institution's sales impact other institutions with common exposures. Several models of fire sale spillovers have been proposed to evaluate the vulnerability of institutions [Greenwood et al., 2015] as well as to construct indexes of systemic risk in financial markets [Duarte and Eisenbach, 2018, Di Gangi et al., 2018]. The model of [Corsi et al., 2016] represents an analytical framework accounting for all these effects.

Finally, the implementation of the Value-at-Risk capital requirement depends on the expectations that financial institutions have on the risk of the assets in their portfolio and on their statistical dependence. For this reason, there is a vast literature on the estimation of risk and dependencies [Tsay, 2005, Hommes et al., 2007, Heemeijer et al., 2009, Bao et al., 2013], many of them based on the recent history of prices in a time window of the recent past. Since the endogenous risk may change as time goes on because of feedback effects, financial institutions adapt their portfo-

[2] This definition assumes mark-to-market pricing. Hence, it is important to notice that the risk is not an observed variable, but we need to estimate it. Methods based on historical observations of prices are the most commonly used for risk estimation. In this case, the Value-at-Risk depends on how institutions use these methods to form risk expectations, for instance what is the past estimation window for the measure of risk.

[3] The work by [Aymanns and Farmer, 2015] is close to the research presented in Chapter 6. Thus, we compare the output of two models at the end of the next chapter.

lios according to their risk perception. Hence, the process of expectation formation drives itself the dynamics of the financial system.

In the following, first we give some insights on the role of expectation feedbacks for a simple and well known model in Economics, *i.e.* the cobweb model. Then, in Subsection 5.3 we review the model of [Corsi et al., 2016], that is the baseline model for our research in Chapter 6 on the role of expectation feedbacks in financial networks.

Expectation feedbacks

In this paragraph, we discuss the role of expectations in the cobweb model in order to illustrate how expectation feedbacks give rise to a dynamical system describing the evolution of a financial market [Hommes, 2013]. Any financial system is an expectation feedback system because the current decisions of players depend on what they expect will occur in the future. The long-established paradigm for equilibrium models in finance refers to rationality and perfect knowledge of the underlying market dynamics by financial agents, which translates in the correct forecasting with no systematic mistakes. However, in the real-world, financial agents use models based on historical observations, especially in forming expectations of risk. Furthermore, when we consider the aggregate behavior of a financial system, it is also unrealistic to assume that the agents know the 'laws of motion' exactly.

There exist many econometric models to analyze financial data, each one capturing specific patterns of the dynamics of prices, returns, and other data. But most importantly, a common aspect for any model refers to how much memory of the past history is considered in forming the expectation. Indeed, this is critical, since there is a tradeoff between choosing a long estimation window to improve statistical significance and preferring a short window in order to obtain a more timely estimation.

The cobweb model [Nicholson and Snyder, 2012] is a partial equilibrium model which describes the commodity price p_t of a non-storable good that takes one time period to be produced. Hence, it describes cyclical supply S and demand D in a market where the produced amount q_t of the good must be chosen before the price is observed. This amount depends then on the price expectation p_t^e by producers, which derive their optimal production decision by maximizing the expected profit.

As explanatory example, we assume as supply curve the following function of the expected price,

$$S(p_t^e) = b + \tanh\left[\lambda\left(p_t^e - c\right)\right], \lambda > 0, b \geqslant 1, c \geqslant 0. \tag{5.2}$$

Consumer demand D depends on the current market price p_t and can be derived by making some assumptions on the consumer utility. However, in the simplest case, we can describe D as a linearly decreasing function,

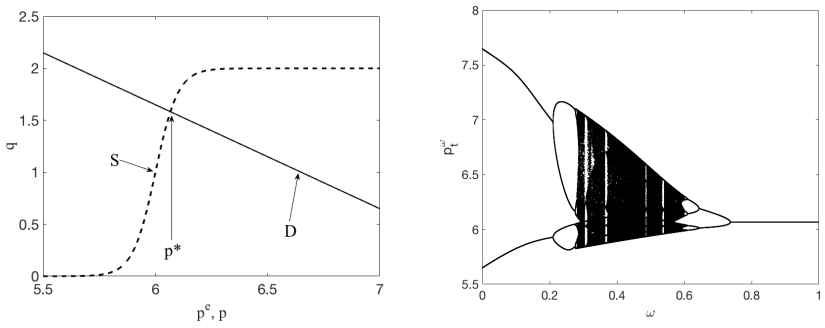


Fig. 5.4 Left: supply S and demand D curve in the cobweb model, p^* represents the equilibrium price. Right: bifurcation diagram of the cobweb model with adaptive expectations characterized by the memory parameter ω.

$$D(p_t) = a - dp_t, \tag{5.3}$$

where d is the slope parameter and a determines the demand level. Supply and demand curves, (5.2) and (5.3) respectively, are shown in the left plot of Figure 5.4.

When expectations are *homogeneous*, *i.e.* all producers have the same expectations, market clearing implies

$$D(p_t) = S(p_t^e) \;\rightarrow\; p_t = D^{-1}(S(p_t^e)) = \left[a - S(p_t^e)\right]/d. \tag{5.4}$$

Hence, the dynamics of the price p_t depends on how the expectation p_t^e is formed.

In the context of equilibrium models in finance, *i.e.* when we regard producers as rational and as having perfect knowledge of the system, the price expectation corresponds to the optimal solution which matches the supply by producers with the demand of consumers, *i.e.*

$$p_t^e = \mathbb{E}(p_t) = p^*, \tag{5.5}$$

where p^* is the price corresponding to the intersection point of demand and supply curves, see the left plot of Figure 5.4. Hence, by solving (5.4) together with (5.5), we obtain the equilibrium price.

In fact, price expectations are formed by using econometric models of past observations.[4] In order to capture the memory of the process of expectation formation, we describe here the so-called *adaptive expectations* [Nerlove, 1958], which are defined as

$$p_t^\omega = \omega p_{t-1}^\omega + (1-\omega)p_{t-1}, \tag{5.6}$$

with parameter ω. The (adaptive) expectation p_t^ω is the weighted average of the most recent expectation p_{t-1}^ω and the most recent observation p_{t-1}.

[4] This holds even more for the estimation of risk and dependence structure of assets in a portfolio.

By iterating (5.6), it is trivial to show that adaptive expectations are equivalent to an Exponentially Weighted Moving Average (EWMA) of past observations,

$$p_t^{\omega} = (1-\omega)\sum_{k=0}^{\infty}\omega^k p_{t-k-1} = (1-\omega)\sum_{k=0}^{\infty}e^{-\frac{k}{\tau}}p_{t-k-1}, \tag{5.7}$$

where $\tau = (\ln\frac{1}{\omega})^{-1}$ determines the 'effective' estimation window, *i.e.* the past observations selected by the exponentially decreasing kernel associated with EWMA. The limit $\omega \to 0^+$ ($\tau \to 0$) corresponds to zero memory of the past before p_{t-1}. By increasing ω, the estimation window τ increases. In effects, ω is the memory parameter of adaptive expectations.

The market clearing mechanism (5.4) together with expectation feedbacks described by (5.6) give rise to a discrete time dynamical system, *i.e.* a map, which relates the current state (p_t^{ω}, p_t) of the financial system with its past state $(p_{t-1}^{\omega}, p_{t-1})$.

A comprehensive insight of the dynamics is obtained by looking at the bifurcation diagram, *i.e.* the diagram of the long run dynamics after a transient time has passed as a function of a parameter. The bifurcation diagram of the cobweb model with adaptive expectations is shown in the right plot of Figure 5.4. Notice that when the memory is large enough, expectations converge to the equilibrium price (5.5), *i.e.* the fixed point of the map. However, decreasing memory gives rise to both periodic and chaotic cycles where expectations are systematically different from realizations. If systematic expectation errors are easily identifiable during periodic evolution and thus the expectation scheme can be improved, chaotic dynamics makes this task more challenging.

Indeed, a dynamical system is defined as chaotic when three conditions hold [Devaney, 2008]: (i) sensitivity to initial conditions; (ii) (topological) mixing; (iii) dense periodic orbits. One often also requires positivity of Kolmogorov-Sinai entropy (or, almost equivalently, of Lyapunov exponents). Roughly speaking, it is equivalent to say that small changes in initial conditions produce large changes in long-term outcomes and any autocorrelation function decays to zero, resulting in unpredictability of the system dynamics in the long-run. In particular, autocorrelation functions decaying to zero implies that expectation errors are highly unpredictable and the learning of the underlying dynamics of the system may be highly challenging. This may prevent the improvement of the expectation scheme.

In conclusion, expectation feedbacks give rise to some Markovian dynamics for the financial system and, most importantly, have the potential of destabilizing the system with periodic or chaotic cycles. The memory of the expectations plays a crucial role in determining systemic stability.

5.3 A model of systemic risk for bipartite financial networks

The model of [Corsi et al., 2016] is a model of indirect contagion of risk because of the combined effect of overlapping portfolios and target leverage in an illiquid finan-

cial market. It is in effect an analytical study of risk contagion mediated by 'external' assets in the portfolio of institutions. In fact, the authors propose an analytical dynamical model to investigate how micro-prudential changes shape macro-prudential outcomes, with a particular attention on the endogenous risk of investments.

The authors model a bipartite financial network (see Figure 5.2) of financial institutions and asset investments, where the institution have capital requirements in the form of VaR constraint and follow standard mark-to-market and risk management rules. Hence, they assess analytically the multivariate feedback effects between investment prices and institutions' behavior induced by the portfolio rebalancing for targeting the leverage, in a market characterized by investments' illiquidity.

The bipartite financial system is composed by N financial institutions (here called also banks) creating portfolios by investing in some of the M available risky assets. In the following, we adopt the convention of labeling risky investments as $i, j, k, \ldots$ and banks as $a, b, c, \ldots$. In the description of the model, we assume to observe the dynamics in a time window $[t, t+1]$.[5] At time t, financial institutions take portfolio decisions about leverage and diversification, then in the subsequent period $(t, t+1]$ they rebalance periodically their portfolios in order to target the financial leverage (consistently with the empirical study of [Adrian and Shin, 2013]). We assume that the portfolio rebalancing occurs n times within $(t, t+1]$.

All the financial institutions are assumed to be equivalent, that is they have the same initial equity capital and the same capital requirements, they solve the same portfolio problem and they have the same expectations of risk. For simplicity, we describe a stylized balance sheet where the left side A is totally composed by 'external' assets, while the right side is simply the sum of liability L and equity E, without distinguishing among different kinds of liabilities. Finally, risky investments are assumed to be ex-ante statistically equivalent. These simplifying assumptions are made for analytical tractability.[6]

The two building blocks of the model of [Corsi et al., 2016] (see below) are: (i) how financial institutions take their decisions about their portfolios, thus determining the optimal number of investment assets m (which is a measure of *diversification*) and the target value for the financial leverage λ, together with the impact of the financial regulation on the portfolio decisions; (ii) given the portfolio decisions, what is the impact of leverage targeting on both the price dynamics and the endogenous risk of investments.

Portfolio decisions

Let us consider the problem of portfolio optimization for a generic bank at time t. The equity at time t is given by $E_t = A_t - L_t$, where L_t (A_t) represents the liabilities (assets) of the bank at time t. For simplicity banks do not face funding restrictions

[5] This is done for notational simplicity and consistency with the research presented in the next Chapter.

[6] However, in [Corsi et al., 2016], the authors show how heterogeneity does not affect crucially the results of the model.

and they can decide to increase or to reduce their liabilities according to their needs as long as they fulfil the VaR constraint. Finally r_L is the per dollar average interest expense on the liability side and the financial leverage is $\lambda_t = \frac{A_t}{E_t}$.

[Corsi et al., 2016] consider a simplified setting where each bank determines the optimal value for leverage and diversification. Since all the risky investments are ex-ante statistically equivalent, financial institutions adopt a simple investment strategy consisting in forming an equally weighted portfolio by randomly selecting m risky investments from the whole collection of M available investment assets. Hence, banks have to find the optimal number of investment assets in the presence of diversification costs $\bar{c}$, see [Constantinides, 1986]. The costs of diversification represent all the informational and infrastructural costs preventing each institution to achieve full diversification of its portfolio.

Banks select also the optimal value for leverage λ_t, being bounded by the VaR constraint $VaR \times A_t \leqslant E_t$. By assuming a functional shape of the return distribution, VaR is the α-multiple of the expected holding period volatility of the portfolio σ^p, *i.e.* $\alpha\sigma^p$. By assuming normally distributed portfolio returns, the VaR quantile associated with a $P_{VaR} = 5\%$ is given by $\alpha = 1.64$.

Since investments are equivalent, μ is the expected return in the holding period and then the net interest margin of the financial institutions is $\mu - r_L$. Finally, financial institutions correctly perceive that each risky investment entails both an idiosyncratic (diversifiable) risk component and a systematic (undiversifiable) risk component, *i.e.* the expected variance of an investment i is $\sigma^2_{i,t} = \Sigma_{u,t} + \Sigma_{d,t}$, where the first (second) term represents the expected variance of the systematic factor (idiosynchratic noise). For example, $\Sigma_{u,t}$ is the expectation at time t of the systematic risk in the period $(t, t+1]$.

The expected variance of a portfolio of m_t assets is

$$\sigma^2_{p,t} = \frac{\Sigma_{d,t}}{m_t} + \Sigma_{u,t}. \tag{5.8}$$

As shown in [Corsi et al., 2016], the portfolio optimization problem is

$$\max_{\lambda_t, m_t} \lambda_t(\mu - r_L) - cm_t \quad \text{s.t.} \quad \alpha\sigma_{p,t}\lambda_t \leqslant 1, \tag{5.9}$$

where c measures the diversification cost for each investment. The solution in implicit form is

$$m_t = \lambda_t \sqrt{\Sigma_{d,t}} \sqrt{\frac{\alpha}{2c} \frac{\mu - r_L}{\sigma_{p,t}}}, \tag{5.10}$$

$$\lambda_t = \frac{1}{\alpha\sigma_{p,t}}. \tag{5.11}$$

Given $\Sigma_{d,t}$ and $\Sigma_{u,t}$, there is only one independent variable in the system of Equations (5.10) and (5.11). Indeed, it is

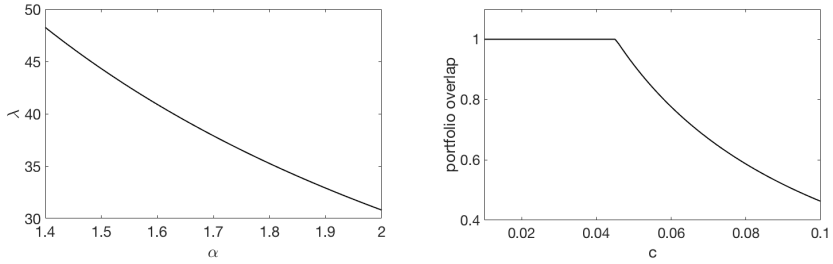


Fig. 5.5 Optimal financial leverage as a function of the parameter α of the Value-at-Risk (left) and (fractional) portfolio overlap as a function of the cost of diversification c (right). Model parameters are: $\mu - r_L = 0.08$, $M = 20$, $\sqrt{\Sigma_{d,t}} = 5\%$, $\sqrt{\Sigma_{u,t}} = 1\%$, $c = 0.05$ (left), and $\alpha = 1.64$ (right).

$$m_t \equiv m_t(\lambda_t; \Sigma_{d,t}, \Sigma_{u,t}) = \frac{\Sigma_{d,t}}{\frac{1}{\alpha^2 \lambda_t^2} - \Sigma_{u,t}}, \tag{5.12}$$

where λ_t is the only positive real solution[7] of the following quartic equation,

$$\Sigma_{d,t}^{\frac{1}{3}} \left(\alpha^{\frac{2}{3}} \left(\frac{\mu - r_L}{2c} \right)^{\frac{1}{3}} \lambda_t \right)^{-1} - \left(\frac{1}{\alpha^2 \lambda_t^2} - \Sigma_{u,t} \right)^{\frac{2}{3}} = 0. \tag{5.13}$$

The left panel of Figure 5.5 shows the solution of (5.13) as a function of the parameter α of the Value-at -Risk constraint. As expected, by decreasing α, *i.e.* less stringent capital requirements, the optimal value for the financial leverage increases.

Bipartite financial network

Once m_t and λ_t are determined by solving the portfolio problem, the financial network is represented by a bipartite graph where the first set of nodes representing the N banks is connected to the second set of nodes describing the M risky investments. Each bank selects randomly m_t of the M available investments. This can be represented by m_t links connecting the bank with the randomly chosen investment-nodes. Finally, each link is associated with a weight $(\lambda_t E_t)/m_t$ because the portfolio is equally weighted. Hence, a realization of the portfolio choice by all banks lead to a specific instance of the bipartite graph, see Figure 5.2.

Within the model, we can compute what is the expected overlap between two portfolios. In fact, the number of banks ν having a specific risky investment in their portfolio is a random variable described by the binomial distribution

$$\mathbb{P}(\nu | N, M, m) = \binom{N}{\nu} \left(\frac{m_t}{M} \right)^{\nu} \left(1 - \frac{m_t}{M} \right)^{N-\nu}$$

[7] It can be shown that Eq. 5.13 has only one positive real solution for $\lambda_t \in \mathbb{R}^+$ in the space of feasible parameters, *i.e.* $\alpha > 0$, $c \in [0,1]$, $\mu - r_L > 0$, $\Sigma_{d,t}, \Sigma_{u,t} > 0$.

whose mean value is $\mathbb{E}(\nu) = m_t N/M$. Given two banks, we can define the overlap o between the two portfolios as the number of investments in common. The overlap is a random variable itself which is described by the hypergeometric distribution

$$\mathbb{P}(o|M,m_t) = \frac{\binom{m_t}{o}\binom{M-m_t}{m_t-o}}{\binom{M}{m_t}} \quad 0 \leqslant o \leqslant m_t$$

whose mean value is $E(o) = m_t^2/M$, that is the expected overlap between two portfolios in the bipartite network of N banks and M investments. In order to obtain a rescaled measure of the overlap which is between 0 and 1, let us define the fractional overlap as $\bar{o} \equiv o/m_t$, thus resulting in $\mathbb{E}(\bar{o}) = m_t/M$. Therefore, the number of risky investments in the portfolio m_t, which is a measure of diversification, represents also the level of overlapping among the portfolios of financial institutions.

In the right panel of Figure 5.5, we show the relation between the portfolio overlap and the cost of diversification c. Optimal portfolio diversification m_t is inversely related to diversification costs c, see Equation (5.9). Then, a decrease of c has the consequence of increasing the overlapping of portfolios, thus resulting in a more interconnected financial systems.

Target leverage and endogenous dynamics

As empirically shown by [Adrian and Shin, 2010], financial institutions adopting target leverage adjust their assets and liabilities rather than raising or redistributing equity capital. This means that equity changes over time as a consequence of the bank profits and losses, but it is not actively managed.

[Corsi et al., 2016] model the dynamics of the return of the risky investment i in the n time steps of length $1/n$ during the interval $(t,t+1]$ as the sum of two components,[8]

$$r_{i,t+k/n} = \eta_{i,t+k/n} + e_{i,t+(k-1)/n}, \qquad k = 1,2,...,n. \tag{5.14}$$

The exogenous component $\eta_{i,t+k/n} = \mu_1 + f_{t+k/n} + \varepsilon_{i,t+k/n}$ is the sum of a drift term representing the expected return plus a systematic market factor $f_{t+k/n}$ common to all risky investments, plus a noise term $\varepsilon_{i,t+k/n}$ representing the idiosyncratic innovation. Without loss of generality, both the noise term and the systematic factor are Gaussian distributed, *i.e.* $\varepsilon_{i,t+k/n} \sim \mathcal{N}(0,\sigma_\varepsilon^2)$ and $f_{t+k/n} \sim \mathcal{N}(0,\sigma_f^2)$ $\forall k = 1,2,...,n$.

The endogenous component $e_{i,t+(k-1)/n}$ depends on the price impact of the demand for the risky investment i arising from the portfolio rebalancing. Given a target leverage, when asset prices increase[9], both asset value and equity increase, since li-

[8] Note that in (5.14) we use a fractional time labeling to indicate the times between t and $t+1$ when portfolio rebalancing occurs. This is done for consistency with the generalization of the model which is presented in the next chapter.

[9] When the asset prices decrease, the sign is simply reverted.

abilities remain constant, and as a consequence the leverage decreases[10]. Thus, in order to keep leverage equal to the target value, banks manage their balance sheet by increasing the liabilities and using the borrowed money for purchasing new assets[11]. In the presence of asset illiquidity, buying or selling assets for leverage targeting will move their prices.

In mathematical terms, given at the generic fractional time $s = t + k/n$ with $k = 1, 2, ..., n$, the desired asset size for a generic bank a, $A^*_{a,s} = \lambda_t E_{a,s}$, the bank a rebalances the portfolio by trading a quantity which is the difference between the desired asset size and the current one, $A^*_{a,s} - A_{a,s}$, which is given by

$$\begin{aligned} A^*_{a,s} - A_{a,s} &= \lambda_t E_{a,s} - A^*_{a,s-1/n}(1 + r^p_{a,s}) = \\ &= \lambda_t (E_{a,s-1/n} + r^p_{a,s} A^*_{a,s-1/n}) - A^*_{a,s-1/n}(1 + r^p_{a,s}) = \\ &= (\lambda_t - 1) A^*_{a,s-1/n} r^p_{a,s}, \end{aligned} \tag{5.15}$$

where $r^p_{a,s}$ is the portfolio return at time s for the bank a. Equation (5.15) shows that any profit or loss from investments in the portfolio ($A^*_{a,s-1/n} r^p_{a,s}$) will result in a change in the asset size, amplified by the value of the financial leverage.

The impact of the portfolio rebalancing by all the banks at time s will affect asset prices at time $s + 1/n$. The total demand of the risky investment i at time s will be the sum of the individual demand of the banks who picked investment i in their portfolio,

$$D_{i,s} = \sum_{a=1}^{N} I_{i \in a} \frac{1}{m_t} (A^*_{a,s} - A_{a,s}), \tag{5.16}$$

where $I_{i \in a}$ is the indicator function which takes value one when investment i is in the portfolio of institution a and zero otherwise. [Corsi et al., 2016] show that this quantity can be written as

$$D_{i,s} = (\lambda_t - 1) \frac{A^*_{s-1/n}}{m_t} \frac{N}{M} \left(r_{i,s} + \frac{m_t - 1}{M - 1} \sum_{k \neq i} r_{k,s} \right).$$

By assuming that price impact is linear in the traded volumes, [Corsi et al., 2016] model the endogenous component as

$$e_{i,s} = \frac{1}{\gamma} \frac{D_{i,s}}{C_{i,s}}, \tag{5.17}$$

where γ is a parameter measuring the investment liquidity, $C_{i,s} \equiv \sum_{a=1}^{N} I_{i \in a} \frac{A^*_{a,s-1/n}}{m_t}$ is a proxy for capitalization of investment i at time s. Coherently with the assumption of statistically equivalence for risky investments, we assume that each invest-

[10] E.g. if the profit is equal to δA, we have $\frac{A + \delta A}{E + \delta A} < \frac{A}{E}$.

[11] According to the assumption of statistically equivalence of risky investments, the increment of asset size is equally distributed over assets in the portfolio [Greenwood et al., 2015].

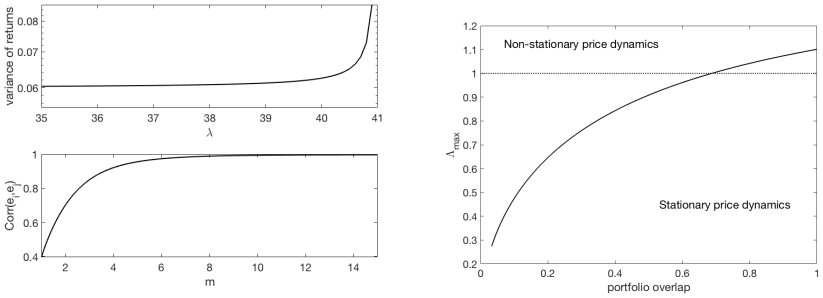


Fig. 5.6 Variance of returns (top left) and correlation of endogenous components (bottom left) as a function of the leverage λ and diversification m, respectively. Largest eigenvalue (right) of the VAR(1) process (5.18) as a function of the (fractional) portfolio overlap m/M. Model parameters are: $\mu - r_L = 0.08$, $M = 20$, $\sqrt{\Sigma_{d,t}} = 5\%$, $\sigma_\varepsilon = 1\%$, $\sigma_f = 1\%$, $c = 0.05$, and $\alpha = 1.64$.

ment is characterized by the same liquidity parameter γ. Then, the authors describe the following Vector Autoregressive (VAR) dynamics for the vector of endogenous components,

$$\boldsymbol{e}_s = \boldsymbol{\Phi}_t \boldsymbol{r}_s = \boldsymbol{\Phi}_t (\boldsymbol{e}_{s-1/n} + \boldsymbol{\eta}_s), \;\; s = t + k/n, \; k = 1, 2, ..., n\,, \tag{5.18}$$

where

$$\boldsymbol{\Phi}_t = \frac{(\lambda_t - 1)}{\gamma} \begin{bmatrix} \frac{1}{m_t} & \frac{1}{m_t}\frac{m_t-1}{M-1} & \cdots & \frac{1}{m_t}\frac{m_t-1}{M-1} \\ \frac{1}{m_t}\frac{m_t-1}{M-1} & \frac{1}{m_t} & \cdots & \frac{1}{m_t}\frac{m_t-1}{M-1} \\ \vdots & & \ddots & \vdots \\ \frac{1}{m_t}\frac{m_t-1}{M-1} & \frac{1}{m_t}\frac{m_t-1}{M-1} & \cdots & \frac{1}{m_t} \end{bmatrix}. \tag{5.19}$$

Notice that $\boldsymbol{\Phi}_t \equiv \boldsymbol{\Phi}(\lambda_t, \Sigma_{d,t}, \Sigma_{u,t})$ because $m_t \equiv m_t(\lambda_t, \Sigma_{d,t}, \Sigma_{u,t})$ according to (5.12).

The VAR(1) process (5.18) determines the return process (5.14) and completely defines the price dynamics of risky investments in the time window $(t, t+1]$.

Within the analytical framework introduced by [Corsi et al., 2016], we can assess the impact on risk of targeting leverage in the presence of portfolios' overlapping by evaluating the endogenous variance of investment returns as well as the dependence structure of the endogenous component of the price dynamics. In particular, it increases both the variance of returns (see the top left panel of Figure 5.6) and the correlation (see the bottom left panel of Figure 5.6) because of the endogenous component in (5.14). Furthermore, the impact on the variance is diverging by approach the threshold value which determines a transition from stationarity to non-stationarity for the return dynamics. In fact, the dynamics of the VAR(1) process (5.18) is dictated by the largest eigenvalue of the matrix (5.19), that is

$$\Lambda_{max} = \frac{\lambda_t - 1}{\gamma}.$$

When Λ_{max} becomes larger than one, the return dynamics is non-stationary. This may occur because an increase of the financial leverage (*e.g.* induced by increased diversification, see the right panel of Figure 5.6) or even for an exogenous reduction of liquidity.

5.3.1 Endogenous systemic risk

[Corsi et al., 2016] consider the portfolio decisions as completely exogenous, in the sense that expected diversifiable and systematic risks in Equation (5.8) are exogenous parameters not depending on the past price dynamics. Hence, they study the properties of the financial system in relation to the return process (5.14) for the asset investments.

In particular, the authors show how changes in the constraints of the bank portfolio optimization (such as changes in the cost of diversification or changes in the micro-prudential policies) endogenously drive the dynamics of bank balance sheets, asset prices, and systemic risk. Specifically, a reduction of diversification costs, by increasing the level of diversification and hence relaxing the VaR constraint, allows the financial institutions to increase the optimal leverage and it also increases the degree of overlap, and thereby correlation, between the portfolios of financial institutions. This increases the feedback effects in the return dynamics that, in turn, increase the risk of investments.

The VAR(1) process (5.18) captures the multivariate feedback effects between investment prices and balance sheet dynamics induced by portfolio rebalancing in the presence of asset illiquidity and show how the results crucially depend on its largest eigenvalue. From a mathematical point of view, a transition from stationarity to non stationarity of the price dynamics occurs when the largest eigenvalue of $\boldsymbol{\Phi}$ becomes larger than one. In other words, the strength (due to higher leverage) and coordination (due to similarity of bank portfolios) of feedback effects trigger a transition from a stationary dynamics of price returns to a non-stationary one characterized by steep growths (bubbles) and plunges (bursts) of market prices.

Finally, [Corsi et al., 2016] show that the bank size heterogeneity does not affect crucially the analytical results obtained in the homogenous case. On the contrary, bank size heterogeneity makes the financial system more unstable as compared to the homogeneous case because the maximum eigenvalue is larger for the heterogenous than for the homogeneous case, making the system closer to the transition between the stationary and the non-stationary dynamics.

Conclusion

In this chapter, we have discussed the role of network theory in assessing the systemic risk of a financial system. In particular, the recent literature have highlighted

how network theory can be successfully applied to study the process of contagion of risk in a networked financial system. In particular, we have pointed out that risk contagion can be mediated by direct exposures between counterparties or, indirectly, by common investments in the portfolios. Recent empirical evidences have shown how financial distress propagates mainly because of overlapping portfolios and that balance sheet aggregates drive endogenously the cycles of funding and liquidity in financial systems, in particular the financial leverage is the endogenous engine governing the systemic stability.

The model of [Corsi et al., 2016] describes analytically how the combined effect of both leverage targeting and overlapping portfolios feed on the dynamics of prices in presence of investments' illiquidity.

In the next chapter, we generalize the model of [Corsi et al., 2016] to study how the endogenous risk feeds back to the portfolios of financial institutions. For instance, VaR-constrained investors are forced to reduce their leverage when the risk has increased. In fact, the definition itself of the Value-at-Risk assumes mark-to-market pricing and then the past observations of prices feed back on the current expectations of risk, thus determining the current portfolio decisions. Hence, we study the role of expectation feedbacks on systemic stability of a bipartite financial system in the presence of risk contagion mediated by overlapping portfolios.

Chapter 6
Expectation feedbacks and systemic risk

Abstract In this chapter, we present an analytical model to study the role of expectation feedbacks and overlapping portfolios on systemic stability of financial systems. Building on [Corsi et al., 2016], we model a set of financial institutions having Value-at-Risk capital requirements and investing in a portfolio of risky assets, whose prices evolve stochastically in time and are endogenously driven by the trading decisions of financial institutions. Assuming that they use adaptive expectations of risk, we show that the evolution of the networked system is described by a slow-fast random dynamical system, which can be studied analytically in some regimes. The model shows how the risk expectations play a central role in determining the systemic stability of the financial system and how wrong risk expectations may create panic-induced reduction or over-optimistic expansion of balance sheets. Specifically, when investors are myopic in estimating the risk, the fixed point equilibrium of the system breaks into leverage cycles and financial variables display a bifurcation cascade eventually leading to chaos. We discuss the role of financial policy and the effects of some market frictions, as the cost of diversification and financial transaction taxes, in determining the stability of the system in the presence of adaptive expectations of risk.
Almost all results in this chapter previously appeared in [Mazzarisi et al., 2019b].

Introduction

In this chapter we present an analytical model of the financial financial bipartite described firstly in [Corsi et al., 2016] where two feedbacks mechanisms are present: on the one hand, as documented by many papers (for a review see the Section 5.2), VaR capital requirements together with the strategy of leverage targeting have a positive feedback effect on the prices of the assets in the portfolios of financial institutions, *i.e.* a financial institution will sell more assets if their prices drop and vice versa when their prices rise; on the other hand, the implementation of any capital requirement depends on the expectations financial institutions have on the risk of

the assets in the portfolio and on their statistical dependence, thus the past history of the price dynamics feeds back on the current expectations of risk and hence on the current portfolio decisions.

The first feedback mechanism, known as *procyclicality* of the Value-at-Risk, drives the return dynamics of risky investments together with the dynamics of the balance sheet itself. Moreover, when financial institutions follows similar risk management rules and investment strategies, its impact may become more important because of the effect of portfolio overlap.

The second feedback mechanism relies on the common practice of risk estimation based on the past observations of prices. In fact, the definition itself of the Value-at-Risk capital constraint, *e.g.* , assumes mark-to-market pricing. There exists a countless number of models on the estimation of risk and dependencies among investments' returns. Many of them are based on the recent history of prices in a time window of the recent past. The choice of the length of the estimation window is critical, since there is a tradeoff between choosing a long estimation window to improve statistical significance and preferring a short window in order to capture a more timely measure of risk. In period of financial turbulence, when non stationary effects are more likely, investors might prefer to use short estimation windows. Since trading decisions drive endogenously the market in the presence of illiquid assets and depend on expectations, the length of the risk estimation window can impact the dynamical properties of prices. This is the second feedback effect in addition to the one, described above, due to target leveraging.

Building on the model of [Corsi et al., 2016], we model a set of financial institutions (banks) investing in a portfolio of risky and illiquid assets follow a target leveraging strategy to satisfy Value-at-Risk capital requirements. The estimations of risk of the investment assets, and as a consequence the leverage, are periodically updated and banks use a backward-looking expectation scheme which considers price returns in a past time window to build estimates. The two feedback mechanisms are coupled by the price dynamics, which on one side is used to mark-to-market the portfolio and to estimate risk and correlations, and on the other one is endogenously affected by the trading activity of financial institutions.

Interestingly, the two feedback mechanisms described above act on different time scales. In our model the time scale of leverage targeting is shorter than the time scale over which financial institutions update their risk expectations. This separation of time scales is crucial in our modeling. Since the slow variables, associated with updates of risk expectations, evolve in time as a function of averages over the fast variables, associated with leverage targeting, our model can be casted as a discrete time slow-fast dynamical system[1]. The ratio between the two time scales is the key parameter determining the type of mathematical modeling. We show that when this ratio tends to infinity, *i.e.* financial institutions are continuously marked-to-market, the dynamics is described by a deterministic map. The window used to form expectations of risk plays a central role in determining systemic stability and leverage cycles appear when investors become more myopic relative to past history

[1] The mathematical framework for slow-fast systems in continuous time is studied in [Kuehn, 2011]. A review about random dynamical systems is [Bhattacharya and Majumdar, 2003].

of asset prices, *i.e.* the memory becomes smaller than a given threshold. Our model predicts that the deterministic dynamics of the financial system becomes chaotic when the memory decreases further and goes below a second smaller threshold. When the ratio between the two time scales is finite a random slow-fast dynamical system describes the system. Even if mathematically this is harder to study, because of the joint chaotic and stochastic dynamics, we show by analytical arguments and numerical simulations that the main dynamical characteristics remain unchanged.

We are therefore able to characterize the possible dynamical outcomes for the considered financial system as a function of the memory window used to form expectations, the tail parameter of the Value-at-Risk, the number of asset investments, the ratio between the two time scales (related to the presence of market frictions), and a parameter determining the level of financial innovation. We show how the breaking of the fixed point equilibrium for the financial system occurs via a period-doubling bifurcation when any of these parameters are varied and how the dynamics of the financial system may be intrinsically chaotic in certain parameter regions. Finally, we aim to stress the fact that each of these parameters can at least in part be controlled by regulators, thus our model is able to provide policy recommendation for enhancing financial stability, as discussed at the end of the chapter.

This work aims to combine several streams of literature: (i) the analysis of portfolio rebalancing induced by the mark-to-market accounting rules and VaR constraint [Adrian and Shin, 2010, Adrian and Shin, 2013]; (ii) the investigations on the impact of the imposition of capital requirements on the behavior of financial institutions and their possible procyclical effects [Danıelsson et al., 2004, Danielsson et al., 2012, Tasca and Battiston, 2016, Corsi et al., 2016]; in particular, we generalize the model of Corsi et al. by studying how the procyclical effects on asset prices and risk expectations influence the portfolio decisions about leverage and diversification; (iii) the literature on distressed selling and its impact on the balance sheets of the financial institutions because of the overlapping among portfolios [Cont and Wagalath, 2013, Caccioli et al., 2014]; (iv) the research on the role of expectation feedbacks in financial systems [Hommes, 1994, Farmer et al., 2012, Hommes, 2013]; in particular, this work focuses on the role of risk expectations which are formed by using statistical models of past observations of investment prices; (v) finally this work contributes to literature on the application of dynamical systems theory to the problem of systemic risk in financial markets [Choi and Douady, 2012, Castellacci and Choi, 2015] and, specifically, to the study of the dynamics of leverage cycles and its relation with the financial regulation [Brunnermeier and Pedersen, 2008, Geanakoplos, 2010, Poledna et al., 2014, Aymanns and Farmer, 2015, Aymanns et al., 2016, Halling et al., 2016].

The remainder of this chapter is organized as follows. In Section 6.1 we describe our model of expectation feedbacks in financial bipartite systems. In Section 6.2 we analyze the dynamical properties of the model and its policy implications for financial markets in the limit in which the model is fully analytical. This limit corresponds to study the deterministic skeleton of the slow-fast random dynamical system. In Section 6.3, we present an analytical argument to extend the obtained results also in the random framework, we give more insights into the model via nu-

merical simulations, and we discuss the role of taxation for financial systemic risk. Then, the last section contains some conclusions.

6.1 Financial system with expectation feedbacks

We model a financial market consisting of leveraged financial institutions investing in some risky assets. Financial institutions, here called banks, have capital requirements in the form of VaR constraint and face costs of diversification in forming a portfolio. We focus on the study of systemic risk in financial markets from the point of view of indirect contagion of risk. In presence of illiquid assets, losses and distressed selling propagate indirectly among financial institutions by common investments in their portfolios.

The adopted framework is the one introduced in [Corsi et al., 2016] and reviewed in Subsection 5.3, where authors model a bipartite network of investment assets and banks. To form a portfolio each bank solves an optimization problem to determine simultaneously the optimal values of leverage and diversification. Determining the optimal level for the financial leverage is related to the fulfilment of the VaR constraint. Hence, portfolio decisions about leverage and diversification depend on the expectations of risk by banks. More specifically, portfolio decisions at time t are taken by solving Equations (5.10-5.11), or equivalently solving (5.13) for the leverage λ_t (within our framework, the diversification m_t is a dependent variable of λ_t and risk expectations $\Sigma_{d,t}$ and $\Sigma_{u,t}$, see Equation (5.12)).

However, [Corsi et al., 2016] are agnostic on the process of expectation formation, since expectations of returns (μ in Equation (5.9)) and risks (diversifiable $\Sigma_{d,t}$ and undicversifiable $\Sigma_{u,t}$ components in Equation (5.8)) are described by exogenous parameters. As a consequence, there is no dynamics associated with the evolving banks' expectations.

Here we explicitly introduce a process for the formation of risk expectations[2] where banks estimate the risk of their portfolio in the holding period from past price movements [Hommes and Wagener, 2009]. In our model, we introduce a simple process for the backward-looking expectations which are characterized by one parameter related to the memory of expectations. The mechanism of expectation feedbacks induces a dynamical component in the portfolio decisions: depending on the evolution of market prices, the perception of risk by banks changes and accordingly also the portfolio decisions about leverage and diversification. Since asset prices evolve stochastically in time, the market value of the portfolio changes. By assuming that the liability size remains unchanged in the meanwhile, the exogenous variation of the leverage occurs in the inverse direction of the portfolio, *i.e.* if the prices increase, the financial leverage decreases. As empirically found [Adrian and Shin, 2010, Adrian and Shin, 2013], banks manage actively their balance sheet by rebalancing periodically the portfolio in order to keep the leverage

[2] Expected asset returns remain exogenous to the price dynamics.

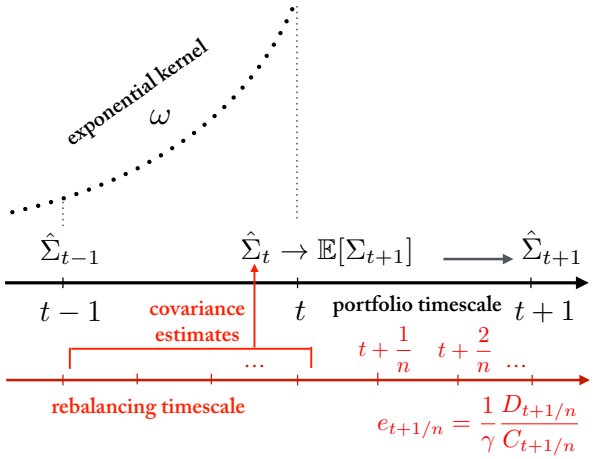


Fig. 6.1 Timing of the slow-fast dynamical model describing the financial system.

equal to its optimal level or target value. The time scale associated with the portfolio rebalancing is the time spent by banks to bring the leverage back to its target value. The frequency of balance sheet adjustments determines how well the bank is marked-to-market in the portfolio holding period.

In our model, we define the time scale of the dynamics for portfolio decisions as the unit time scale (*slow* timescale) and n is the number of times a bank rebalances its portfolio for leverage targeting in the unitary holding period. Hence, the time scale associated with the portfolio rebalancing is $1/n$ (*fast* timescale). In the presence of finite liquidity, the trading of assets drives endogenously the dynamics of the price. Since the price history is used to form expectations of risk, a feedback coupling portfolio decisions and price dynamics is created.

A pictorial representation of the coupled feedback dynamics is in Figure 6.1.

6.1.1 Formation of risk expectations

[Corsi et al., 2016] describe the impact of portfolio decisions on the price dynamics in a time window $(t,t+1]$ for given exogenous expectations of risk, see Subsection 5.3. Here, we assume that the banks update their expectations of risk at time intervals of unitary length and accordingly that they take decisions about leverage and portfolio diversification (the black axis in Figure 6.1). In the time interval $(t,t+1]$ with $t \in \mathbb{Z}$, banks rebalance their portfolio to target the leverage, but without changing the risk expectations. The rebalancing takes place in n time subintervals within $(t,t+1]$, *i.e.* at $\{t+\frac{1}{n},t+\frac{2}{n},t+\frac{3}{n},...,t+\frac{n}{n} \equiv t+1\}$. $1/n$ with $n \in \mathbb{N}^+$ is the time scale associated with the *portfolio rebalancing* (the red axis in Figure 6.1), whereas

the unit value is the time scale associated with the *portfolio decisions* (the black axis in Figure 6.1).

We generalize the model of [Corsi et al., 2016] by describing the endogenous process of formation of risk expectations. Several empirical and experimental studies, for example [Hommes and Wagener, 2009, Bao et al., 2013], have shown that financial agents use statistical models of past observations of the price to forecast the future, the so-called backward-looking expectations. In particular, we assume *adaptive* expectations of risk (see Section 5.2), defined as follows.

Risk estimation

At each portfolio decision time t, banks estimate the covariance matrix of risky investments between $t-1$ and t, that is

$$\widehat{\boldsymbol{\Sigma}}_t = \begin{pmatrix} \widehat{Var}[R_{i,t}] & \widehat{Cov}[R_{i,t},R_{j,t}] & \cdots \\ \widehat{Cov}[R_{i,t},R_{j,t}] & \ddots & \vdots \\ \vdots & \cdots & \cdots \end{pmatrix}, \tag{6.1}$$

where $\widehat{Var}[R_{i,t}]$ and $\widehat{Cov}[R_{i,t},R_{j,t}]$, with $i \neq j$, are the maximum likelihood estimators of the variance and covariance of assets' returns aggregated at the time scale of the portfolio decisions, *i.e.* $R_{i,t} \equiv \sum_{k=1}^{n} r_{i,t-1+k/n}$[3]. We use the hat, *e.g.* $\widehat{\Sigma}_{u,t}$, to indicate an *estimation* of the variance formed by using past observations. On the contrary, unlabeled symbols indicate *expectations*.

We assume that banks correctly perceive the return dynamics as evolving according to the vector autoregressive process VAR(1), see Equatiosn (5.14) and (5.18-5.19), as well as the statistical equivalence of asset investments. As a consequence, symmetric conditions can be imposed on $\widehat{\boldsymbol{\Sigma}}_t$, namely the diagonal components of the covariance matrix are equal to each other and the same for the off-diagonal components. It is

$$\widehat{\boldsymbol{\Sigma}}_t = \widehat{\Sigma}_{d,t}\mathbb{1} + \widehat{\Sigma}_{u,t}\mathbf{1}, \tag{6.2}$$

where $\mathbb{1}$ is the identity matrix, $\mathbf{1}$ is the $M \times M$ matrix whose entries are equal to one and $\widehat{\Sigma}_{d,t} = \widehat{Var}[R_{i,t}] - \widehat{Cov}[R_{i,t},R_{j,t}]$ and $\widehat{\Sigma}_{u,t} = \widehat{Cov}[R_{i,t},R_{j,t}]\ \forall i,j = 1,...,M,\ i \neq j$.

We do not exploit explicitly the maximum likelihood estimators of the variance and covariance of returns to obtain the results in the next Section. However, we show in the appendix section to this chapter how to obtain the explicit formulas for completeness. Here, let us stress that $\widehat{\Sigma}_{d,t}$ and $\widehat{\Sigma}_{u,t}$ are only functions of the fast variables $\boldsymbol{r}_{t-1+k/n},\ k = 0,1,2,...,n$, *i.e.*

[3] For notational simplicity, we are considering that the asset returns are centered around the mean.

$$\begin{cases} \widehat{\Sigma}_{d,t} \equiv \widehat{\Sigma}_{d,t}(\{r_{i,t-1+k/n}\}_{k=0,1,\ldots,n}^{i=1,\ldots,M}) \\ \widehat{\Sigma}_{u,t} \equiv \widehat{\Sigma}_{u,t}(\{r_{i,t-1+k/n}\}_{k=0,1,\ldots,n}^{i=1,\ldots,M}). \end{cases}$$

Risk expectations

Once banks have estimated the covariance matrix of risky investments, they form risk expectations of the two independent quantities of the covariance matrix, *i.e.* $\Sigma_{d,t}$ and $\Sigma_{u,t}$. We assume the banks expectation at time t of an element of the covariance matrix is a weighted sum of the previously adopted expectation for the same element and the current estimation, that is

$$\begin{cases} \Sigma_{d,t}^{\omega} = \omega \Sigma_{d,t-1}^{\omega} + (1-\omega)\widehat{\Sigma}_{d,t} \\ \Sigma_{u,t}^{\omega} = \omega \Sigma_{u,t-1}^{\omega} + (1-\omega)\widehat{\Sigma}_{u,t}, \end{cases} \tag{6.3}$$

where $\omega \in [0,1]$ is the memory parameter of the expectation scheme. Equation (6.3) defines the adaptive expectation of risk.

6.1.2 Expectation feedback system

By summarizing the building blocks of our model, the dynamics of the financial system is described by a slow-fast random dynamical system at discrete time which is specified by the following equations,

$$\begin{cases} \left(\Sigma_{d,t}^{\omega}\right)^{\frac{1}{3}} \left(\alpha^{\frac{2}{3}} \left(\frac{\mu - r_L}{2c}\right)^{\frac{1}{3}} \lambda_t\right)^{-1} - \left(\frac{1}{\alpha^2 \lambda_t^2} - \Sigma_{u,t}^{\omega}\right)^{\frac{2}{3}} = 0 \\ \Sigma_{d,t}^{\omega} = \omega \Sigma_{d,t-1}^{\omega} + (1-\omega)\widehat{\Sigma}_{d,t} \\ \Sigma_{u,t}^{\omega} = \omega \Sigma_{u,t-1}^{\omega} + (1-\omega)\widehat{\Sigma}_{u,t} \end{cases} \tag{6.4}$$

$$\boldsymbol{r}_s = \boldsymbol{\eta}_s + \boldsymbol{\Phi}(\lambda_{t-1}, \Sigma_{d,t-1}^{\omega}, \Sigma_{u,t-1}^{\omega})\, \boldsymbol{r}_{s-1/n} \;\; s = t-1+k/n, \; k = 1,2,\ldots,n\,, \tag{6.5}$$

where $\widehat{\Sigma}_{d,t}$ and $\widehat{\Sigma}_{u,t}$ are the maximum likelihood estimators of the (diversifiable) variance and covariance of returns aggregated at the time scale of portfolio decisions and can be obtained as explained in the appendix section, see (6.31).

Equation (6.4) describes the dynamics of banks' portfolio decisions, *i.e.* the slow component of the dynamics described by the slow variables λ_t, $\Sigma_{d,t}^{\omega}$, and $\Sigma_{u,t}^{\omega}$. Equation (6.5) describes the price evolution of the risky investments, *i.e.* the fast component of the dynamics described by the fast variables $r_{i,s}$.

The model must satisfy a stationarity condition, namely covariance stationarity of the autoregressive process, which is $\lambda < \gamma + 1$.[4] It is also $\lambda \geqslant 1$ and $0 \leqslant m \leqslant M$, by construction.

Summarizing the main steps of the dynamics: (i) portfolio decisions at time $t-1$ determine the value of leverage depending on risk expectations, see (6.4), and (ii) affect the price dynamics between $t-1$ and t because of the autoregressive coefficient $\boldsymbol{\Phi}(\lambda_{t-1}, \Sigma^{\omega}_{d,t-1}, \Sigma^{\omega}_{u,t-1})$, of the return process (6.5); (iii) at time t, banks estimate asset risks by using past observations of prices, $\widehat{\Sigma}_{d,t}(\{\boldsymbol{r}_{t-1+k/n}\}_{k=1,2,...,n})$ and $\widehat{\Sigma}_{u,t}(\{\boldsymbol{r}_{t-1+k/n}\}_{k=1,2,...,n})$. Then, (iv) banks form new risk expectations $\Sigma^{\omega}_{d,t}$ and $\Sigma^{\omega}_{u,t}$, and (v) make new portfolio decisions at time t.

One dimensional setting

In the following, we study also a reduced version of the model obtained by considering one bank and one risky investment. In this setting, the price dynamics is governed by the autoregressive process (5.14) with $N = M = 1$. We refer to the investment return at time s as r_s and to its variance estimated by using observations $\{r_{t-1+k/n}\}_{k=1,2,...,n}$ as $\hat{\sigma}^2_t$. In this setting we lose the aspect related to the diversification and the portfolio problem reduces to find the optimal value of the leverage subject to the Value-at-Risk constraint, *i.e.*

$$\max_{\lambda_t} \lambda_t(\mu - r_L) \ \text{ s.t. } \ \alpha \sigma^{\omega}_t \lambda_t \leqslant 1,$$

where σ^{ω}_t is the expected volatility of the investment. Then, the equation governing the portfolio decisions is $\lambda_t = (\alpha \sigma^{\omega}_t)^{-1}$ with $(\sigma^{\omega}_t)^2 = \omega(\sigma^{\omega}_{t-1})^2 + (1-\omega)\hat{\sigma}^2_t$. We can reduce the slow component of the model dynamics to a single equation representing a one-dimensional map for the dynamic variable λ_t. The reduced model is specified as follows,

$$\begin{cases} \lambda_t = \left(\omega \frac{1}{\lambda^2_{t-1}} + (1-\omega)\, \alpha^2\, \widehat{Var}\left[\sum_{k=1}^{n} r_{t-1+k/n}\right] \right)^{-\frac{1}{2}} \\ r_s = \varepsilon_s + \frac{\lambda_{t-1}-1}{\gamma} r_{s-1/n} \qquad s = t-1+k/n,\ k = 1,2,...,n \end{cases}, \tag{6.6}$$

where

$$\widehat{Var}\left[\sum_{k=1}^{n} r_{t-1+k/n}\right] = \left(1 + 2\frac{\hat{\phi}_{t-1}(1-\hat{\phi}^n_{t-1})}{1-\hat{\phi}_{t-1}} - 2\frac{(n\hat{\phi}_{t-1} - n - 1)\hat{\phi}^{n+1}_{t-1} + \hat{\phi}_{t-1}}{n(1-\hat{\phi}_{t-1})^2} \right) \frac{n\hat{\sigma}^2_{\varepsilon}}{1-\hat{\phi}^2_{t-1}}$$

is the maximum likelihood estimator of the variance of the return aggregated at the time scale of the portfolio decisions for generic $n > 1$, with

[4] The VAR(1) process (6.5) is covariance stationary when the largest eigenvalue of the matrix $\boldsymbol{\Phi}$ is smaller than one. This is equivalent to assume $\lambda_t < \gamma + 1$, see Subsection 5.3.

$$\hat{\phi}_{t-1} = \frac{\sum_{k=1}^{n} r_{t-1+k/n} r_{t-1+(k-1)/n}}{\sum_{k=1}^{n} r^2_{t-1+(k-1)/n}}$$

$$\hat{\sigma}^2_{\varepsilon} = \frac{\sum_{k=1}^{n} \left(r_{t-1+k/n} - \hat{\phi}_{t-1}\, r_{t-1+(k-1)/n}\right)^2}{n}$$

being the maximum likelihood estimators of the autoregressive coefficient $\phi_{t-1} \equiv \frac{\lambda_{t-1}-1}{\gamma}$ and the variance of the idiosyncratic noise σ^2_{ε} of the AR(1) process.

In Equation (6.6) covariance stationarity for the AR(1) process is equivalent to the condition $\lambda \in [1, \gamma+1)$.

Since the slow variable evolves in time depending on averages over the fast variables, namely the variance of the random autoregressive process AR(1), this is a slow-fast random dynamical system.

In this setting, we can also consider the case $n = 1$, *i.e.* the time scale of portfolio decisions coincides with the one of portfolio rebalancing. In appendix to this chapter, we show that this case is similar to the one presented in [Aymanns and Farmer, 2015] and, with a further assumption for the asset price dynamics, the two models coincide.

In the following we will consider two cases.

- **Asymptotic deterministic limit of the slow-fast random dynamics.** The first case is the limit $n \to \infty$, *i.e.* portfolio are rebalanced very actively between $t-1$ and t. In this limit the model becomes analytically tractable.
- **Two generic time scales.** The second is the case of finite n and, for simplicity, we consider the reduced version of the slow-fast random dynamical system. We will present some numerical results and some analytical approximation.

6.2 Asymptotic deterministic limit of the slow-fast random dynamics

The limit $n \to \infty$ allows to approach the problem in a fully analytical way. Since n represents the number of times financial institutions rebalance their portfolios, it is a measure of how closely they are marked-to-market in the capital structure. The asymptotic limit $n \to \infty$ is equivalent to consider all financial institutions continuously marked-to-market in their capital structures. For a VaR constrained financial institution this means that the asset size is equal to the desired position with respect to the market value of the investments in such a way that the financial leverage is equal to its target during the portfolio holding period. Very active management of portfolio means frequent trading in the financial market and thus can be accomplished only in absence of or with very low transaction costs and other trading frictions. For this reason, we can interpret the limit $n \to \infty$ as an ideal market without trading frictions.

In this limit, the estimator of the covariance matrix, $\widehat{\boldsymbol{\Sigma}}_t$, converges in probability to the covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ for the returns $R_{i,t}$,

$$\widehat{\boldsymbol{\Sigma}}_t \xrightarrow{\mathbb{P}} \widetilde{\boldsymbol{\Sigma}} = \widetilde{\Sigma}_d(\lambda_{t-1}, \Sigma^{\omega}_{d,t-1}, \Sigma^{\omega}_{u,t-1}, \Sigma_\varepsilon, \Sigma_f)\mathbb{1} + \widetilde{\Sigma}_u(\lambda_{t-1}, \Sigma^{\omega}_{d,t-1}, \Sigma^{\omega}_{u,t-1}, \Sigma_\varepsilon, \Sigma_f)\mathbf{1}, \tag{6.7}$$

where Σ_ε and Σ_f are the variance of the idiosyncratic noise $\varepsilon_{i,s}$ and of the factor f_s, respectively, aggregated at the time scale of portfolio decisions[5], *i.e.* $\Sigma_\varepsilon = \lim_{n\to\infty} n\sigma^2_\varepsilon$ and $\Sigma_f = \lim_{n\to\infty} n\sigma^2_f$. $\widetilde{\Sigma}_d$ and $\widetilde{\Sigma}_u$ can be computed analytically, exploiting a result in [Corsi et al., 2016], and their expression is in Equation (6.34) in the appendix section.

Substituting the estimator with the covariance matrix (6.7), the portfolio dynamics is described by a three-dimensional deterministic map, which in implicit form is:

$$\begin{cases} \left(\Sigma^{\omega}_{d,t-1}\right)^{\frac{1}{3}} \left(\alpha^{\frac{2}{3}} \left(\frac{\mu - r_L}{2c}\right)^{\frac{1}{3}} \lambda_t\right)^{-1} - \left(\frac{1}{\alpha^2 \lambda_t^2} - \Sigma^{\omega}_{u,t-1}\right)^{\frac{2}{3}} = 0 \\ \Sigma^{\omega}_{d,t} = \omega \Sigma^{\omega}_{d,t-1} + (1-\omega)\widetilde{\Sigma}_d(\lambda_{t-1}, \Sigma^{\omega}_{d,t-1}, \Sigma^{\omega}_{u,t-1}, \Sigma_\varepsilon, \Sigma_f) \\ \Sigma^{\omega}_{u,t} = \omega \Sigma^{\omega}_{u,t-1} + (1-\omega)\widetilde{\Sigma}_u(\lambda_{t-1}, \Sigma^{\omega}_{d,t-1}, \Sigma^{\omega}_{u,t-1}, \Sigma_\varepsilon, \Sigma_f). \end{cases} \tag{6.8}$$

The map (6.8) is the so-called *deterministic skeleton* of the financial model obtained by removing the sources of stochasticity.

6.2.1 Deterministic skeleton of the financial system

We analyze the model in the asymptotic limit via a dynamical systems approach by investigating the map (6.8). Table 6.1 shows the values of the parameters used to illustrate numerically the main properties of the model.

A comprehensive insight of the skeleton dynamics is obtained by looking at the bifurcation diagram in the top panel of Figure 6.2. For our model we can use as the bifurcation parameter either the memory parameter ω, or the VaR parameter α, or the diversification cost c. Later, we will show the model dynamics depending on α and c to have more insights regarding policy implications.

The top panel of Figure 6.2 shows the bifurcation diagram as a function of ω. Decreasing the memory parameter, a period-doubling cascade to chaos occurs. In other words the deterministic skeleton of the financial system shows cycles of increasing complexity with more and more periods until chaos occurs[6]. The signature

[5] In the limit $n \to \infty$, the time scale of the portfolio rebalancing, *i.e.* $1/n$, goes to zero as well as σ^2_ε and σ^2_f, but $\Sigma_\varepsilon = \lim_{n\to\infty} n\,\sigma^2_\varepsilon$ and $\Sigma_f = \lim_{n\to\infty} n\,\sigma^2_f$ remain finite.

[6] In this analytical approach to the study of dynamics of the financial system, the domain for leverage and risk expectations is $\{\lambda, \Sigma^{\omega}_d, \Sigma^{\omega}_u\} \in [1, \gamma+1) \times \mathbb{R}^+ \times \mathbb{R}^+$. In particular, bounds for leverage are equivalent to the conditions of stationarity for the process (6.5).

Notation	Description	Value
M	total number of investments	60
N	number of banks	30
$\mu - r_L$	Net Interest Margin of a bank	0.08
γ	asset liquidity	100
$\sqrt{\Sigma_\varepsilon}$	exogenous idiosyncratic volatility at the time scale of portfolio decisions	0.03
$\sqrt{\frac{\Sigma_f}{\Sigma_\varepsilon}}$	ratio between exogenous volatility of systematic factor at the time scale of portfolio decisions, *i.e.* Σ_f, and Σ_ε	0.1
A_0	Initial asset size for each bank	100
E_0	Initial equity for each bank	$\frac{A_0}{\lambda_0}$
c	cost of diversification w.r.t. the initial equity E_0	0.1
α	quantile of P_{VaR}	1.64
$\lambda_0, \Sigma^\omega_{d,0}, \Sigma^\omega_{u,0}$	chosen randomly in the domain	

Table 6.1 Simulation parameters for the model. The most important parameters in our analysis are ω, c and α. We will specify ω below, while α and c are set as in the Table if it is not specified differently. Since we study the long-run dynamics, the initial conditions are not relevant. We set the transient time in the simulations of the model as equal to 100 time steps.

of the chaotic behavior is the positive Lyapunov exponent associated with it. The bottom panel of Figure 6.2 shows the estimated Lyapunov exponent Λ as a function of ω. Positive values of Λ, observed for small ω, signal the presence of deterministic chaos.

Note that the unit of ω in Figure 6.2 is the inverse of the time scale of the portfolio decisions (see Equation 5.7). By using as unit the time scale of portfolio rebalancing, the time scale becomes $\tilde{\omega} = \omega^{\frac{1}{n}}$. For example, if banks rebalance their portfolios at daily frequency but the portfolio decisions are taken once a month, a memory parameter $\tilde{\omega} = 0.97$ at the daily time scale corresponds to an effective memory of $\tau \approx 33$ days. The corresponding value of the memory parameter at the monthly time scale is $\omega \approx 0.4$. Thus small values of ω, which in the bifurcation diagram corresponds to cycles or chaos, are associated with values of $\tilde{\omega}$ close to 1.

One can understand the behavior in Figure 6.2 by considering that, when memory is large enough, banks learn from past history and, at least in the long-run, they make exact forecasts of future risks. Thus large memory stabilizes the financial system and leverage and diversification reach a fixed point equilibrium. When ω decreases below a given threshold, risk expectations with smaller memory break the fixed point equilibrium by inducing cycles of period two. Indeed, a shorter memory has the potential of creating a panic-induced fall in portfolio holdings reflecting sudden decrease in leverage and diversification. When the adopted financial leverage and the overlap between banks' portfolios are high, the price impact of the portfolio rebalancing increases significantly the observed risks. Small memory in risk expectations tends to overestimate future risks and as a consequence banks reduce suddenly their positions. Hence, leverage and diversification suddenly decrease. The oppo-

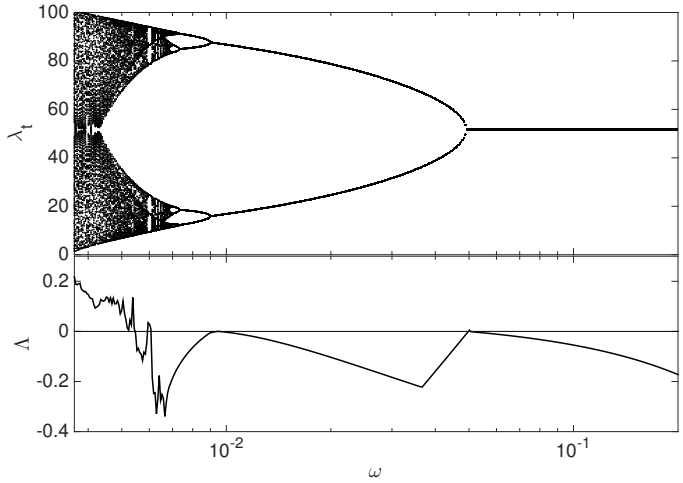


Fig. 6.2 Top panel: bifurcation diagram associated with the leverage λ_t for the analytical mapping of Eq. 6.8 as a function of the memory parameter ω. Bottom panel: the largest Lyapunov exponent Λ associated with the bifurcation diagram. Λ is obtained numerically via the method of the orbits [Wolf et al., 1985].

site situation occurs when leverage and diversification are small due to previously overestimated risk. The 2-period cycle is characterized by a mismatch between the expectation and the realization of portfolio volatility and leverage cycles reflect the mismatch between the banks' perceived risk and the 'true' risk.

When the memory decreases further, our model predicts cycles of larger periods and, eventually, a chaotic dynamics. Notice that the amplitude of both periodic and chaotic leverage cycles increases by decreasing ω. When $\omega > \omega^*$, λ oscillate within $[1, \gamma + 1)$ and the covariance stationarity of the return is guaranteed, while when $\omega < \omega^*$ the process becomes non stationary. Hence, ω^* corresponds to a transition from stationarity to non stationarity for the model dynamics. In the left panel of Figure 6.3, we show the contour map of ω^* as a function of the square root of the idiosyncratic variance $\sqrt{\Sigma_\varepsilon}$ and the total number of investments M. The larger is $\sqrt{\Sigma_\varepsilon}$, the smaller is the value of ω^*: this means that larger exogenous risk reduces the value of the financial leverage because of the Value-at-Risk constraint and, as a consequence, the feedback effects of leverage targeting. More surprisingly, the value of ω^* increases with M, that is introducing financial innovations and new instruments requires investors having risk expectations formed with larger memory to keep the system in stationary equilibrium.

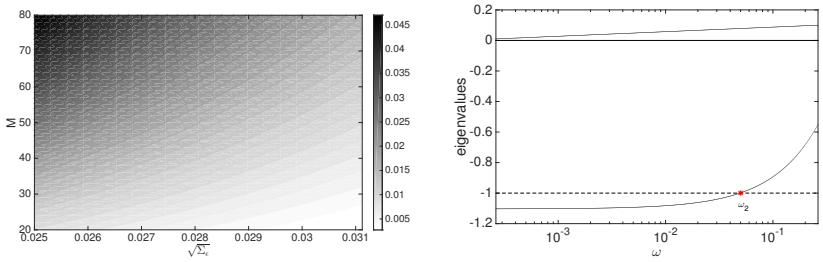


Fig. 6.3 Left panel: contour map of ω^* (*i.e.* the value of the memory parameter ω at which the transition to non stationarity occurs) as a function of idiosyncratic risk $\sqrt{\Sigma_\epsilon}$ and the number of investments M. Here the exogenous systematic variance $\Sigma_f = 0$ and the other parameters are specified as in Table 6.1. Right panel: eigenvalues of the Jacobian for the map (6.8) as a function of the memory parameter ω (the value of the other parameters are specified as in Table 6.1).

6.2.2 Bifurcation analysis

The breaking of fixed point equilibrium occurs by a period-doubling bifurcation or flip bifurcation (see for example [Crawford, 1991]). In the case of ω as bifurcation parameter, we refer to the value of the memory parameter for which the period-doubling bifurcation occurs as ω_2. The type of bifurcation can be analyzed by studying the Jacobian associated with the map (6.8). The map describes a three-dimensional dynamical system but one eigenvalue of the Jacobian is identically null, suggesting that the two equations associated with the process of expectation formation in (6.8) are not independent at the first order in the expansion. The positive eigenvalue describes the eigenspace associated with the independent component of these equations. The dynamics in this eigenspace is characterized by a fixed point, since the corresponding eigenvalue is inside the unit circle for all ω. The third eigenvalue is instead negative. The right panel of Figure 6.3 shows the non vanishing eigenvalues of the Jacobian as a function of ω. When the negative eigenvalue hits the critical value -1 for $\omega = \omega_2$, a period-doubling bifurcation occurs. After that, the attractor for the dynamics is not the fixed point anymore and a new solution appears describing the 2-period cycles. The new solution is the fixed point for the map of (6.8) iterated twice. In a self-similar way, the period-doubling bifurcation repeats again but for the iterated map giving rise to the period-doubling cascade to chaos. The phenomenon is very general, was noticed for the first time by studying the logistic map [May, 1976] and was described in [Feigenbaum, 1978].

We also study the stability properties of the financial system by analyzing the value of ω_2 as a function of the other model parameters. In the top left panel of Figure 6.4, we show the relation between the value ω_2 and the liquidity parameter γ. The larger is the market liquidity of risky investments, the less important are the effects related to the expectation feedbacks. Hence, the larger is the market liquidity, the more stable is the dynamics of the financial system. Intuitively, when the price impact of the portfolio rebalancing to target the financial leverage does not

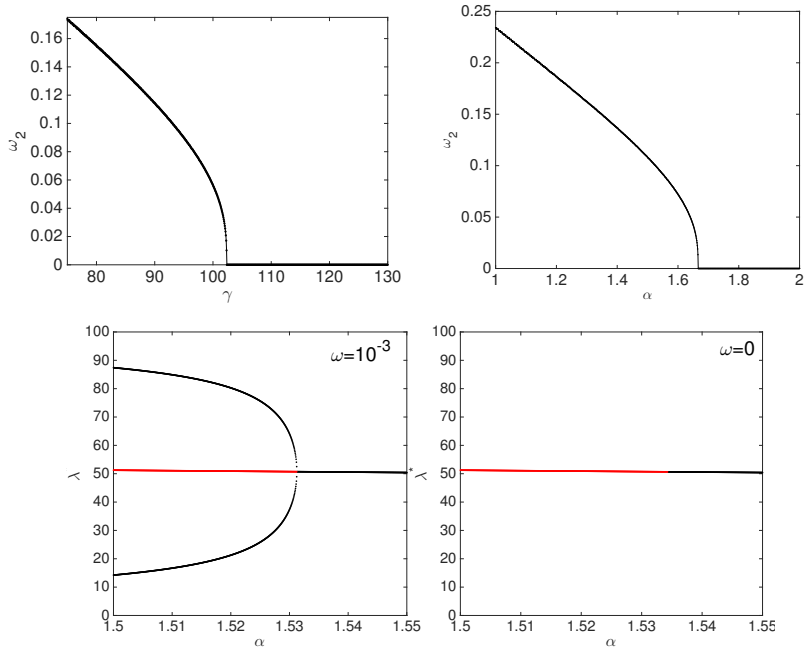


Fig. 6.4 Top panels: ω_2, *i.e.* the value of the memory parameter ω corresponding to the first period-doubling bifurcation, as a function of market liquidity of assets γ (left) and quantile α of the *VaR* (right). Bottom panels: fixed point analysis of the analytical mapping (6.8). Black dots represent attractive stable points (fixed points or stable cycles of period 2), while red dots represent unstable fixed points. Bottom left panel: adaptive expectations with non zero memory ($\omega = 10^{-3}$). Bottom right panel: *naive* expectations characterized by zero memory ($\omega = 0$).

affect significantly the investment prices, then volatility and portfolios do not fluctuate much. Liquid markets are more stable because endogenous impact of market players' strategy affects less the risk [Danielsson et al., 2012]. Finally notice that, since the transition values, such as ω_2, depend on liquidity, an abrupt change in liquidity can drive the market from a stable to an unstable regime. In the top right panel of Figure 6.4, we show the relation between α and ω_2. The decreasing behavior indicates that when Value-at-Risk is less stringent, the financial market becomes dynamically stable if the memory used by agents to build expectations is longer.

In the bottom panels of Figure 6.4 we show the fixed point analysis of the analytical mapping (6.8) in the neighbourhood of the first bifurcation point. We compare the case of adaptive expectations of risk (6.3) with naive expectations, *i.e.* the case of zero memory ($\omega = 0$ in Equation (6.3)). The bottom left panel shows the stable points of the map (black dots) for the variable λ when memory is different from zero, *i.e.* the fixed point of the map (6.8) or the periodic orbit that is the fixed point for the map iterated twice. When the bifurcation parameter α decreases below the threshold determining the period-doubling bifurcation, the new solution for the map

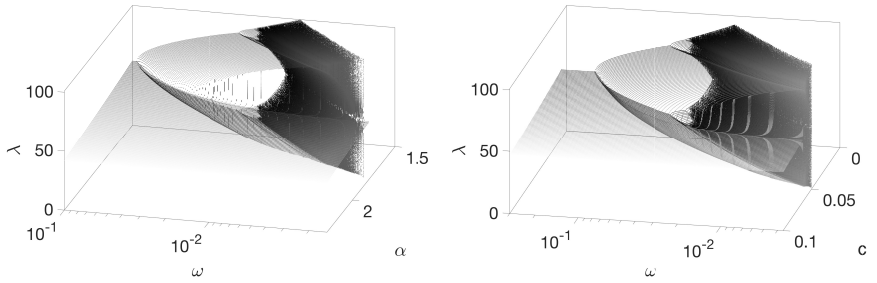


Fig. 6.5 2*D*-bifurcation analysis for the leverage λ in the (ω,α) (left) and (ω,c) (right) plane.

iterated twice appears and the fixed point of the map becomes unstable (red dots). This behavior holds for any ω different from zero. The bottom right panel of Figure 6.4 shows the case of naive expectations of risk. At the bifurcation point, no solutions for the map iterated twice exist. However, the fixed point is not attractive anymore. In this case, the dynamics of the system is cyclical but characterized by an amplitude increasing in time[7]. Hence, when financial agents behave naively in risk forecasting, the breaking of fixed point equilibrium corresponds to a transition from stationary to non stationary dynamics for the financial system.

Finally, we focus on the 2*D*-bifurcation analysis of the model, namely the bifurcation diagram when two parameters are varied. The 2*D*-bifurcation diagrams in the (ω,α) and (ω,c) planes are shown in Figure 6.5. Results confirm that the dynamics of the model is described by period-doubling bifurcations leading to chaos, whatever it is the bifurcation parameter. This is further confirmed in Figure 6.6. However, it is interesting to notice that numerical simulations suggest the existence of two separate regions in the parameter space according to a threshold for α (or c) above which the period-doubling cascade to chaos does not occur. That is, the system is asymptotically in the fixed-point equilibrium independently from the memory of risk expectations. We investigate analytically this behavior in the reduced version of the model in Subsection 6.2.4.

6.2.3 Policy implications

From the point of view of financial policy, our results can be summarized as follows.

- Decreasing the parameter of the VaR, thus allowing a less stringent capital constraint, makes financial system more unstable. For example, in the described financial system characterized by an average investment volatility of 3% in the holding period of the portfolio, diversification costs about 10% of the equity value and a memory in risk expectations equal to $\omega = 0.1$, relaxing the prob-

[7] Clearly, the map (6.8) makes sense as long as the variables are in the domain.

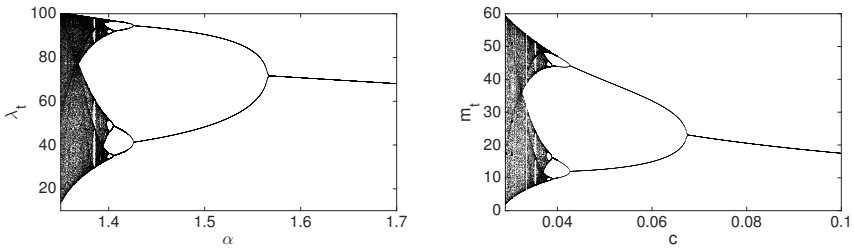


Fig. 6.6 Left panel: bifurcation diagram associated with the leverage λ_t for the analytical mapping (6.8) as a function of VaR parameter α. Right panel: bifurcation diagram for the diversification m_t as a function of the cost of diversification c. In both panels, the value of the memory parameter ω is set equal to 0.1..

ability of loss P_{VaR} from 5% to 6% ($\alpha = 1.64 \to 1.5$) breaks the fixed point equilibrium inducing leverage cycles. Furthermore, allowing probability of loss associated with Value-at-Risk of about $7-8\%$ leads the financial system towards chaotic evolution, see the left panel of Figure 6.6.

- Decreasing the cost of diversification tends to increase the dynamical instability of the financial system because of increased coordination of feedback effects due to similarity of bank portfolios. A similar effect is obtained by introducing new financial instruments which increase the number of available asset investments M. This behavior has been investigated in the recent literature, e.g. [Wagner, 2008, Brock et al., 2009, Ibragimov et al., 2011], and in the context of [Corsi et al., 2016] the authors have proved the existence of a region of instability. Specifically, there exists a stationarity condition for the dynamics of asset returns which defines the maximal value of the financial leverage, above which the return dynamics governed by portfolio decisions of banks becomes non stationary and characterized by steep growths (bubbles) and plunges (bursts) of market prices. However, in [Corsi et al., 2016] the focus is on how financial innovations (modeled by a reduction of c) feed on the price dynamics and, in particular, on the endogenous risk of assets, without considering how risk expectations change consequently. Here, we study how increased endogenous risks feed back on risk expectations of financial institutions, and therefore on portfolio decisions. In particular, we point out that financial innovations can destabilize the dynamics of portfolio decisions by breaking the fixed-point equilibrium into cycles of increasing complexity, eventually leading to a chaotic dynamics of leverage and diversification[8], see Figure 6.6.

[8] Comparing the outputs of the two models for the same set of parameters one finds that: (i) for large c both models converge to an equilibrium, (ii) for small values of c the model of [Corsi et al., 2016] is non stationary, while our model can exhibit periodic or chaotic behavior, in line with the idea that memory delays the transition to the non stationary phase. Interestingly, there is an intermediate range of c in which our model exhibits a 2-periodic orbit, whereas the model of [Corsi et al., 2016] has a stationary equilibrium (although at a higher leverage level).

- Imposing that financial institutions operate in the market to adopt larger memory in their expectation schemes for risk forecasting has always a stabilizing effect on market stability. Not surprisingly, increasing ω means to take into account more information about the system, converging asymptotically to the fixed-point equilibrium. On the contrary, short memory means overreactions to risk news, especially in the presence of positive feedbacks, which deviate the system from the equilibrium state. Thus, our analysis suggests that an appropriate financial policy is necessary to promote a systemic behavior which prefers smooth adjustments driven by more information to overreactions driven by news[9].
  Our finding regarding the role of expectations in systemic stability of a financial system shows similar features with the concepts of equilibrium introduced for nonlinear economic models [Brock and Hommes, 1997, Hommes and Sorger, 1998, Hommes et al., 2005] where particular attention is devoted to the problem of asset pricing equilibrium. One of the main findings of these works concerns the stability of the steady state depending on how much information lies in the expectations of agents. For instance, in [Brock and Hommes, 1997] the concept of *adaptively rational equilibrium* has been introduced and the authors have studied the stability of the steady state of a financial system where agents can choose among a finite set of expectations schemes which weight differently the past observations. In the context of asset pricing models, several experimental studies [Hommes et al., 2008, Heemeijer et al., 2009, Bao et al., 2012, Bao et al., 2013] have highlighted that expectation feedbacks can have a large impact on whether the market can reach a steady state and, in fact, trend following strategies may drive the system far away from equilibrium by amplifying random deviations of price (news) when recent observations are overweighted. The role of the memory in the formation of expectations has been investigated in other works [Chiarella and He, 2003, Hommes et al., 2012] where some conditions for convergence to the steady state have been established.
  These results are in line with what we observe in our model, namely adaptive expectations with large memory of past observations tend to stabilize the system because they have more information on system dynamics, resulting in lower prediction errors, at least in the long-run.

6.2.4 One dimensional analysis

The dynamics of the reduced model in the limit $n \to \infty$ is governed by the following one dimensional map for λ_t,

[9] Regarding the capital requirements for banks' exposure to market in terms of Value-at-Risk estimations, the Basel III regulatory framework [BCBS III, 2011] establishes banks are allowed to develop their own internal VaR models, subject to supervisory approval, but the observation period to be considered in the weighting scheme for historical data must be at least one year. In the asymptotic limit we consider, $\tau = 1\,year$ corresponds to $\omega = 1/e \approx 0.37$.

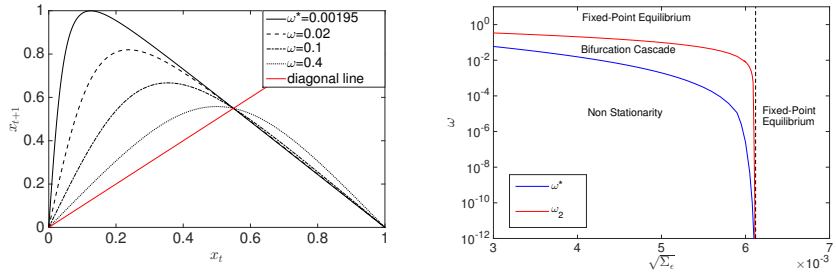


Fig. 6.7 Left panel: shape of the one dimensional map (6.9) with domain rescaled to unitary interval, *i.e.* we consider the linear transformation $x \equiv \frac{\lambda-1}{\gamma}$ where λ is the financial leverage and γ is the liquidity parameter. The other parameters in Equation (6.9) are: $\alpha = 1.64$ (VaR parameter), $\gamma = 100$, $\sqrt{\Sigma_\varepsilon} = 0.005$ (exogenous idiosyncratic risk). Right panel: the value of the memory parameter at which the first period-doubling bifurcation occurs, *i.e.* ω_2 (red line), and the value at which the transition to non stationarity occurs, *i.e.* ω^* (blue line), as a function of $\sqrt{\Sigma_\varepsilon}$. The other parameters are $\alpha = 1.64$ and $\gamma = 100$.

$$\lambda_t = f(\lambda_{t-1};\omega,\alpha,\gamma,\Sigma_\varepsilon) = \left(\omega \frac{1}{\lambda_{t-1}^2} + (1-\omega)\,\alpha^2 \left(1 + 2\frac{\lambda_{t-1}-1}{\gamma-\lambda_{t-1}+1}\right) \frac{\Sigma_\varepsilon}{1-\left(\frac{\lambda_{t-1}-1}{\gamma}\right)^2} \right)^{-\frac{1}{2}}, \tag{6.9}$$

where $\lambda \in [1, \gamma+1)$ and $\Sigma_\varepsilon = \lim_{n\to\infty} n\sigma_\varepsilon^2$ is the variance of the idiosyncratic noise at the slow time scale of portfolio decisions, see (6.6). In the appendix section, we show how to compute analytically the variance of the aggregated return $\sum_{q=1}^{n} r_{t-1+q/n}$ in the limit $n \to \infty$ to obtain (6.9). The map f in explicit form provides valuable analytical insights into the dynamical behavior of the system. The fixed point of the map, *i.e.* $\lambda^* = f(\lambda^*;\omega,\alpha,\gamma,\Sigma_\varepsilon)$, corresponds to the asymptotic equilibrium when $f'(\lambda^*;\omega,\alpha,\gamma,\Sigma_\varepsilon)$ lies inside the unit circle. When $f'(\lambda^*;\omega,\alpha,\gamma,\Sigma_\varepsilon)$ hits the value -1, the period-doubling bifurcation occurs. The left plot of Figure 6.7 shows the shape of the map, under the linear transformation $x \equiv \frac{\lambda-1}{\gamma}$, as a function of ω. Notice that the value of the fixed point does not depend on ω and the maximum of the map is a decreasing function of ω. Furthermore, the value ω^* at which the transition to non-stationarity occurs corresponds to the value of ω at which the maximum of f is equal to $\gamma+1$ (or the unit value when we rescale the invariant interval of the map as in Figure 6.7).

It is interesting to notice that the complex behavior of the period-doubling cascade to chaos may occur when the exogenous variance of the idiosyncratic noise is below a threshold value. In the right panel of Figure 6.7 we show ω_2 and ω^* as a function of $\sqrt{\Sigma_\varepsilon}$. When we approach the threshold (black dotted line), ω_2 and ω^* tend to coincide and both values become arbitrarily low. Above the threshold, no breaking of the fixed point equilibrium occurs. The curves describing ω_2, ω^*,

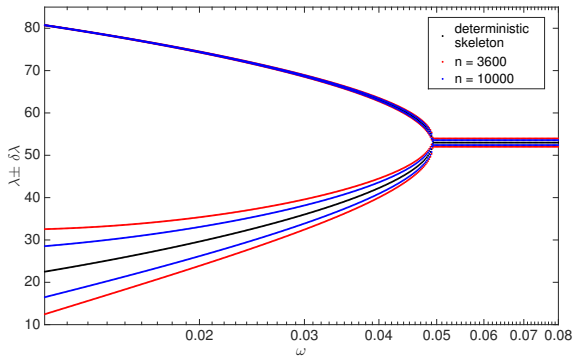


Fig. 6.8 Bifurcation diagram of the leverage as a function of the memory parameter ω and perturbations around the deterministic skeleton of the map 6.8 for different values of n (*i.e.* how many times the portfolio is rebalanced before a new optimization) obtained according to the perturbation analysis described in Subsection 6.3.1.

and the threshold partition the parameter space in regions characterized by specific dynamical properties.

6.3 Two generic time scales

The results of Section 6.2 are obtained in the asymptotic limit $n \to \infty$ corresponding to a situation where financial institutions are continuously marked-to-market in their capital structure. In reality, there are several types of frictions in trading operations which prevent continuous marking-to-market. It is interesting to investigate how the results obtained via the dynamical systems approach change when n is finite. From a policy perspective this corresponds to investigate the role of frictions (*e.g.* transaction taxes) on the stability of the financial system.

6.3.1 Perturbation analysis

For large but finite n, the estimator $\hat{\boldsymbol{\Sigma}}_t$ of the covariance matrix fluctuates around the covariance matrix $\widetilde{\boldsymbol{\Sigma}}$. These fluctuations in risk estimation affect portfolio decisions about leverage and diversification and, as a consequence, also the dynamics of the financial system. Here, we want investigate how much fluctuations in risk estimations affect the deterministic skeleton of financial system. In the following, we present an analytical argument to answer this question. Then, in Subsection 6.3.2 we justify it by comparing the analytical results with Monte Carlo simulations of the reduced version of the model.

The asset return dynamics is governed by the VAR process (6.5). By estimating the autoregressive process between $t-1$ and t when n is finite, financial institutions can obtain the exogenous components of the VAR (1) process, *i.e.* the residuals $\{\varepsilon_{i,s}, f_s\}_{s=t-1+k/n,\, k=1,2,...,n}^{i=1,...,M}$[10]. The estimator of the idiosyncratic and systematic risks at the time scale of portfolio rebalancing is the realized variance of the residuals,

$$\hat{\sigma}_{\varepsilon}^2 = \frac{1}{n-1}\sum_{k=1}^{n} \varepsilon_{i,t-1+k/n}^2 \;\; \forall i = 1,...,M\,, \qquad \hat{\sigma}_{f}^2 = \frac{1}{n-1}\sum_{k=1}^{n} f_{t-1+k/n}^2. \tag{6.10}$$

Since by assumption residuals are i.i.d. and normal, the quantities $(n-1)\hat{\sigma}_{\varepsilon}^2/\sigma_{\varepsilon}^2$ and $(n-1)\hat{\sigma}_{f}^2/\sigma_{f}^2$ follow a chi-squared distribution χ_{n-1}^2 with $n-1$ degrees of freedom, and the 90% confidence interval of $\hat{\sigma}_{\varepsilon,f}^2$ is

$$\delta\hat{\sigma}_{\varepsilon,f}^2 \equiv \frac{\sigma_{\varepsilon,f}^2}{n-1}[(\chi_{n-1}^2)^{-1}(0.95) - (\chi_{n-1}^2)^{-1}(0.05)]. \tag{6.11}$$

We use this confidence interval as a measure of the fluctuations in risk estimation. Let us notice that it goes to zero when n goes to infinity, as numerically confirmed in the simulation of the model. Fluctuations in the estimation of the covariance matrix (6.7) are obtained according to formulas in Equations (6.17), (6.18), and (6.28).[11]

Portfolio decisions are directly affected by fluctuations in risk estimation. Through the analytical mapping (6.8) and by using $\Sigma_{\varepsilon} \equiv n\sigma_{\varepsilon}^2$ and $\Sigma_{f} \equiv n\sigma_{f}^2$, we obtain the range of variability of the leverage as a function of $\delta\hat{\sigma}_{\varepsilon,f}^2$. This is the 90% confidence interval of leverage, $\delta\lambda$, with respect to the skeleton dynamics when n is finite[12].

Figure 6.8 shows the deterministic skeleton and the fluctuations around it when n is finite. When n is large, the dynamical evolution of the system is very close to its skeleton. By decreasing the value for n, fluctuations become more and more important. Empirically we observe that $\delta\lambda$ is inversely proportional to n. However, at least for the 2-period cycles, properties of the dynamical model are conserved when n is of order $\mathcal{O}(10^3)$ or larger.

Notice that the approach presented here cannot be applied to the chaotic dynamics. In dynamical systems theory, when chaos occurs, it is not possible to describe the system through the linearization around an equilibrium because there is not any-

[10] From a mathematical point of view, the process (6.5) is equivalent to the process for the endogenous components of returns, see Equation (5.18). As shown in [Corsi et al., 2016], it can be decomposed in $M+1$ independent AR(1) process. The residuals of the process (6.5) are simply linear combinations of the residuals associated with the $M+1$ AR(1) processes.

[11] From the point of view of VAR(1) process estimation, model variables λ_{t-1}, m_{t-1} in formulas in the appendix section correspond to parameters of the VAR(1) process, Φ_{t-1} in Equation (6.5). Similarly to the variance of residuals $\hat{\sigma}_{\varepsilon}^2$ and $\hat{\sigma}_{f}^2$, in the ideal process of estimating $\boldsymbol{\Phi}_{t-1}$, we obtain in turn the estimator $\widehat{\boldsymbol{\Phi}}_{t-1}$ within confidence intervals. For simplicity, we assume a picked distribution on the real values.

[12] $\delta\lambda$ is obtained by differentiating the map (6.8) with respect to λ_t and $\sigma_{\varepsilon,f}^2$.

more a generalized equilibrium point (such as a stable fixed point or a periodic orbit), see [Eckmann and Ruelle, 1985]. Hence, this analytical argument based on the perturbation of a generalized equilibrium cannot be applied when the deterministic skeleton is chaotic.

6.3.2 Numerical results

In this Subsection we study numerically the model for finite n, *i.e.* when the ratio between the slow time scale and the fast one is finite. For tractability, we focus on the reduced version of the original model by considering one financial institution investing in one risky asset. In the reduced model, we lose the aspects related to the diversification of the portfolio. In turn, the comparison of numerical simulations with the theoretical results of the previous Subsection is easily obtained, as it will be clear below. Finally, this approach via the reduced model focuses more clearly on the role of n in systemic stability of the financial system. In the numerical simulations, we assume that the bank correctly perceives the autoregressive dynamics of the asset price and the variance $\hat{\sigma}_t^2$ is obtained by estimating the AR(1) process on observed returns, see Equation (6.6). We consider for simplicity the case of large memory in risk expectations ($\omega = 0.4$) in such a way that the deterministic skeleton is a fixed point equilibrium. The case of 2-period cycles is equivalent from the point of view of the analytical approximations.

In Figure 6.9 we show the fluctuations of the financial leverage $\delta\lambda$ as a consequence of the fluctuations in the estimator of risk. We compare the analytical approximation (black dots) with the result from Monte Carlo simulations (blue dots). The analytical approximation is obtained via Equation (6.11) with $M = 1$ risky investment. Numerically, we obtain $\widehat{\delta\lambda}$ in the following way. We simulate the reduced version of the slow-fast random dynamical system. Then, we collect the data after the initial transient is passed. Hence we obtain the empirical probability distribution $\hat{F}_\lambda$ associated with a simulated path $\{\hat{\lambda}_t\}_{t=0,\dots,T}$. Coherently with $\delta\lambda$ obtained via the analytical approximation, we determine $\widehat{\delta\lambda}$ as the 90% confidence interval

$$\widehat{\delta\lambda} = \hat{F}_\lambda^{-1}(0.95) - \hat{F}_\lambda^{-1}(0.05). \tag{6.12}$$

The theoretical prediction and the numerical results agree when $n \geqslant 10^4$.

Discussion: trading costs

In the recent literature on leverage cycles, the possibility of two different time scales for portfolio decisions and leverage targeting has not been analyzed. We strongly believe that this aspect deserves more attention since it is directly related to the discussion about the role of taxation in financial systemic risk [Matheson, 2012, Masciandaro and Passarelli, 2013]. We can interpret n as an indirect measure of fi-

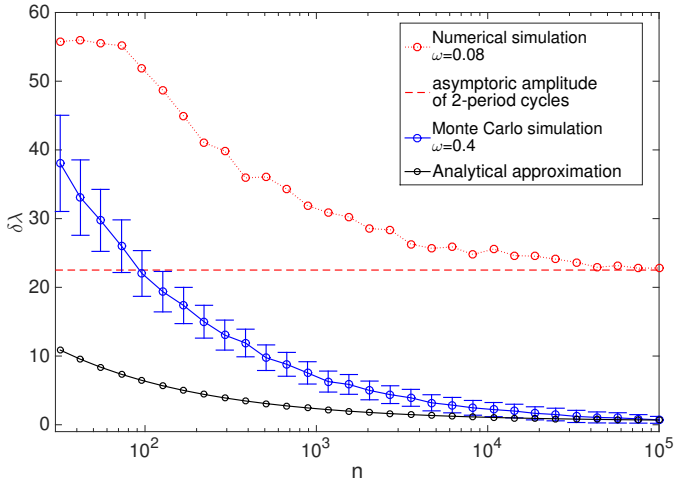


Fig. 6.9 90% confidence interval $\delta\lambda$ for the leverage λ , see Equation (6.12), as a function of the number of intervals n at which the portfolio is rebalanced, for two different values of the memory ω: $\omega = 0.4$ (blue dots) and $\omega = 0.08$ (red dots). Theoretical prevision for $\omega = 0.4$ obtained via Equation (6.11) in the case $N = M = 1$ (reduced model) is represented by the black dots. In the Monte Carlo simulations of the reduced model, for each value of n we average over 50 seeds to obtain the considered value for $\widehat{\delta\lambda}$ and the error bar represents the standard deviation of the data. The other model parameters are: $\gamma = 100$, $\alpha = 1.64$, $\Sigma_\varepsilon = 0.05$, $T = 2000$.

nancial transaction taxes (e.g. the Tobin tax) and other trading frictions. Indeed, n is related to the number of trading operations in the market, because it is equal to the number of rebalancing operations by a bank to have the balance sheet close to the desired capital structure. Clearly high transaction taxes decrease the number of operations for a financial institution.

In our analysis, we study how the amplitude of leverage cycles depends on n in the reduced model. In Figure 6.9, we show $\delta\lambda$ as a function of n. The plot shows the results of Monte Carlo simulations and each point is obtained by averaging over 50 realizations of the system dynamics with $T = 2000$. We show the obtained results for two different values of ω. The case $\omega = 0.4$ is associated with the fixed point equilibrium in the deterministic skeleton, while the case $\omega = 0.08$ corresponds to the 2-period cycles. When $n \gg 1$ we recover the deterministic skeleton as shown in the previous Section. In this limit, $\delta\lambda$ tends to zero in the case of fixed point equilibrium because oscillations tend to disappear. When the deterministic skeleton corresponds to the 2-period cycles, the asymptotic value of $\delta\lambda$ corresponds to the amplitude of the bifurcated 2-period orbit. When $n \gg 1$, both simulated and theoretical amplitude of the leverage cycles tend to coincide, that is a further proof of the consistency of the analytical approximation introduced previously.

In both cases, we can notice that the amplitude of leverage cycles is a decreasing function of n. The intuition is that when n is small, leverage oscillations tend to

correlate more strongly with the price movements. Hence, increments in target leverage reflects the increasing of asset size and at the same time perception of low risks. However, since the bank is not marked-to-market but it has increased considerably its target leverage, when an exogenous shock for the price occurs, a panic-induced fall in financial leverage follows. On the contrary, when n is large and the bank is marked-to-market in the capital structure, the financial system tends to be close to its deterministic skeleton. In the case of fixed point equilibrium, the leverage cycles tend to disappear because of exact risk previsions, at least in the long run.

Also when the adopted memory in risk expectations is small and the deterministic skeleton corresponds to the 2-period orbit, the amplitude of the leverage cycles is smaller for larger n. Furthermore, in this case the cyclical structure of the dynamics can be easily recognized, e.g. in the autocorrelation functions of the realized variance. This information can be usefully adopt to improve risk estimations.

Our results suggest that allowing mark-to-market in the capital structure by removing trading frictions tends to reduce the amplitude of leverage cycles and as a consequence the financial system is closer to the fixed point equilibrium.

Conclusions

In this chapter, we have studied the implications of backward-looking expectations of risk on portfolio decisions and as a consequence on the stability properties of the financial system. To this end we have developed a model which can be described as a slow-fast random dynamical system. The main characteristics of this stylized financial market are: (i) financial institutions having capital requirements in the form of the VaR constraint and following standard mark-to-market and risk management rules; (ii) asset illiquidity; (iii) indirect contagion of risk mediated by the overlapping of portfolios; (iv) backward-looking expectations via statistical models of past observations of prices.

In the asymptotic deterministic limit, we are able to study analytically the fixed point equilibrium of the financial system and how the breaking of systemic stability occurs. The main result is the analytical classification of the possible dynamical outcomes for the considered financial system and its relation with market parameters which are the memory of expectations ω, the probability of Value-at-Risk defined by α, and the cost of diversification c. We have shown how the breaking of the fixed point equilibrium for the financial system occurs via a period-doubling bifurcation which determines the appearing of leverage cycles. Furthermore, we have shown that the dynamics of the financial system is chaotic in certain conditions and, at the best of our knowledge, this work represents the first analytical proof of the appearing of chaotic dynamics in this context. Some recent literature, see for example [Aymanns and Farmer, 2015], has argued the presence of chaotic attractors in models of systemic risk in financial systems by means of numerical arguments. Then, [Choi and Douady, 2012] proposed a new instability indicator whose goal is capturing the chaotic dynamics of cash flows among fi-

nancial institutions during turmoil periods, and [Castellacci and Choi, 2015] have modeled financial contagion in the Eurozone crisis with a similar aim. Finally, [Raffaelli and Marsili, 2006, Marsili et al., 2009] have pointed out that the feedback dynamics coupling the portfolio and the price exhibits a critical point from a stable to an unstable evolution. Moreover, the authors have supported this conclusion by estimating the models on empirical data, thus suggesting that real markets operate close to a dynamically unstable point.

The main goal here is to characterize the policy implications in relation to the dynamical outcomes of our model. In particular, the results suggest how expectation feedbacks may give rise to leverage cycles, which can show periodic or chaotic behavior. The period doubling cascade to chaos is a signal of how much the leverage dynamics might become less and less predictable (and therefore the assessment of the stability of the financial system becomes harder and harder) when financial innovations are introduced and/or capital constraints become less stringent. Since the observed dynamics concerns the portfolio variables, our model suggests that the relevant time scale is the slow one, corresponding in the real world to the time horizon of portfolios. Hence, in order to capture empirical evidences of the dynamical effects described here, the ideal dataset should contain complete information on the evolution of portfolios of financial institutions, including also how many times the portfolio is rebalanced in the holding period. Having information on when institutions take new decisions depending on their expectations (slow time scale) and on how fast they rebalance the portfolio (fast time scale), the implementation of the slow-fast random dynamical model predicts the dynamics of the system and some characteristics that, in principle, we can observe, *e.g.* the Lyapunov exponent associated with the orbits of the financial leverage. However, this is beyond the scope of this work. Furthermore, it is important to notice that the detection of signals of chaotic dynamics is a very hard task when random sources are present, see [Guegan, 2009] for a discussion.

We have studied the dynamical outcomes of the model in order to answer the following questions, within this model:

1. *How do the system's stability properties depend on expectations of risk formed by banks?* Our answer is that, all else being equal, the larger is the memory ω in the process of expectations formation, the more stable is the financial system dynamics.
2. *How important is the constraint on financial leverage?* In our analysis we have shown that, whatever is the ratio between the slow time scale and the fast one, there exists a tipping point for the tail parameter of the Value-at-Risk (α) which defines a 'transition' from the fixed point equilibrium to a region of instability where the system shows periodic or chaotic behavior. Hence, a more stringent regulation for the financial leverage is always stabilizing for the market.
3. *What are the consequences of introducing new financial instruments?* We have found that a market with a larger number of asset investments requires larger memory in forming risk expectations to be in dynamical equilibrium. Moreover, financial innovations which tend to increase the portfolio diversification, lead to an increase of the financial leverage. The combined effects of the larger overlap

of banks' portfolios with the larger price impact of the leverage targeting make the system more unstable.

4. *What is the role of market frictions from a systemic risk point of view?* Decreasing the cost of diversification, here represented by c, may appear positive from a microscopic point of view but may lead to increase coordination of feedback effects due to similarity of banks portfolios, triggering a transition from a stable dynamics to the unstable one. On the contrary, our analysis suggests that decreasing transaction costs and removing all trading frictions may induce financial investors to adopt the strategy of being marked-to-market in their capital structure, *i.e.* large values for the control parameter n. Within our model, this represents the control strategy of the balance sheet which has the consequence of reducing the amplitude of the cycles of leverage.

Finally, we believe that several research questions can be addressed inside the framework we introduced and will lead to further developments of this work. Among other aspects, we believe that the following points deserve attention for future outlooks:

1. the experimental work of [Hommes, 2011] supports the view that heterogeneity is a crucial aspect of a theory of expectations, first of all because of empirical evidences about different behaviors of agents in the market. Here, we have assumed adaptive risk expectations with a given memory that is equal for all banks, *i.e.* the 'agents' of our model. The introduction of different time horizons in the process of risk expectations formation will represent a first step towards heterogeneity of expectations. Furthermore, banks might use different heuristics to form their expectations of risk, such as scenario generation based on historical data or stress-testing on macro factors. More research in this area is called for to define the conditions for systemic stability as macro-outcome of the interaction of heterogenous expectations at the micro-level;
2. another open question is to relax the homogeneity assumptions regarding bank sizes and investment assets. As already pointed out by [Corsi et al., 2016], bank size heterogeneity does not affect importantly the dynamics of asset returns, in particular bank size distribution is irrelevant in determining the largest eigenvalue governing the vector autoregressive process which describes the asset returns. As a consequence, bank size heterogeneity would not change significantly the results about the dynamics of portfolio decisions. What is more challenging concerns heterogeneity of investments, *i.e.* a financial system with different asset classes characterized by different risk and liquidity. Since portfolio decisions drive endogenously the market evolution, fire sales by an institution may increase the risk perception of investments which were previously characterized by lower levels of risk. Thus, expectation feedbacks amplify further this perception, contributing to decrease liquidity of these investments in favour of others. Hence, expectation feedbacks would give rise endogenously to flight-to-liquidity.
3. the Basel III regulation has established several risk measures to be adopted in risk management by financial institutions. Some measures, *e.g.* , have been proposed to overcome the problem of procyclicality of the Value-at-Risk. Hence, different

measures of risk can be implemented as constraints of the portfolio optimization problem and the role of expectation feedbacks in relation to each measure can be assessed.

Appendix

Covariance matrix of VAR(1) process at the slow time scale

In this Section we show how to find the maximum likelihood estimator of the variance and covariance of the VAR(1) process (6.4) and how to compute analytically the covariance matrix (6.7). However, before considering the multivariate VAR(1) process, let us focus on the univariate AR(1) process of the reduced model for which the computation of the variance represents a simpler problem which could shed light on the subsequent multivariate case.

Variance of the AR(1) process at the slow time scale

Let us consider the AR(1) process (6.6), *i.e.*

$$r_s = \varepsilon_s + \phi\, r_{s-1/n},\ \ s = t-1+k/n,\, k = 1,2,...,n\,, \tag{6.13}$$

where $\varepsilon_s \sim \mathcal{N}(0,\sigma_\varepsilon^2)$, $\forall s$, $|\phi|<1$ for the assumption of covariance stationarity and starting point r_{t-1}. The variance of the variable r_s is $Var[r_s] = \mathbb{E}[r_s^2] = \frac{\sigma_\varepsilon^2}{1-\phi^2}$, see [Tsay, 2005]. The variance of the process aggregated between $t-1$ and t, *i.e.* $Var[\sum_{k=1}^n r_{t-1+k/n}]$, is

$$\mathbb{E}\left[\left(\sum_{k=1}^{n} r_{t-1+k/n}\right)^2\right] = n\mathbb{E}[r_s^2] + 2\left(n\sum_{k=1}^{n}\mathbb{E}[r_s r_{s-k/n}] - \sum_{k=1}^{n} k\mathbb{E}[r_s r_{s-k/n}]\right), \tag{6.14}$$

where we use the assumption of covariance stationarity between $t-1$ and t. It is $\mathbb{E}[r_s r_{s-k/n}] = \phi^k \mathbb{E}[r_s^2]$, $\forall s = t-1+k/n$, $k = 1,2,...,n$, by applying recursively (6.13). By exploiting this result, it is

$$\sum_{k=1}^{n}\mathbb{E}[r_s r_{s-k/n}] = \mathbb{E}[r_s^2]\sum_{k=1}^{n}\phi^k = \mathbb{E}[r_s^2]\frac{\phi(1-\phi^n)}{1-\phi}, \tag{6.15}$$

$$\sum_{k=1}^{n} k\mathbb{E}[r_s r_{s-k/n}] = \mathbb{E}[r_s^2]\sum_{k=1}^{n} k\phi^k = \mathbb{E}[r_s^2]\frac{(n\phi-n-1)\phi^{n+1}+\phi}{(1-\phi)^2}. \tag{6.16}$$

By substituting (6.15) and (6.16) in (6.14), we obtain the expression for the variance in Equation (6.6). In the limit $n\to\infty$, it is $\phi^n \to 0$ and $n\phi^n \to 0$ because of $|\phi|<1$. Then, the formula for the variance in Equation (6.9).

Multivariate case: VAR(1)

Let us focus on the VAR(1) process (6.5) where the endogenous component follows the VAR(1) process (5.18). Let us assume $n \gg 1$ for analytical tractability. At the time scale of the portfolio rebalancing, *i.e.* $1/n$, the variance and the covariance associated with the endogenous component can be computed analytically (see [Corsi et al., 2016]). When $s = t-1+k/n,\ k = 1,2,...,n$, it is

$$\begin{aligned} Var[e_{i,s}] = &-\frac{\tilde{\lambda}^2}{(\gamma^2-\tilde{\lambda}^2)(m^2(\gamma^2(M-1)^2-\tilde{\lambda}^2)+2-\tilde{\lambda}^2 mM-\tilde{\lambda}^2M^2)}\times \\ &\times\Big(m^2(\sigma_\varepsilon^2(\tilde{\lambda}^2-\gamma^2(M-1))+\sigma_f^2(\tilde{\lambda}^2-\gamma^2(M-1)^2))+ \\ &+2m(M(\sigma_\varepsilon^2(\gamma^2-\tilde{\lambda}^2)-\tilde{\lambda}^2\sigma_f^2)-\gamma^2\sigma_\varepsilon^2)+ \\ &+M(M(\sigma_\varepsilon^2(\tilde{\lambda}^2-\gamma^2)+\tilde{\lambda}^2\sigma_f^2)+\gamma^2\sigma_\varepsilon^2)\Big), \end{aligned} \tag{6.17}$$

$$\begin{aligned} Cov[e_{i,s},e_{j,s}] = &-\frac{\tilde{\lambda}^2}{(\gamma^2-\tilde{\lambda}^2)(m^2(\gamma^2(M-1)^2-\tilde{\lambda}^2)+2-\tilde{\lambda}^2 mM-\tilde{\lambda}^2M^2)}\times \\ &\times\Big(m^2(\sigma_f^2(\tilde{\lambda}^2-\gamma^2(M-1)^2)-\gamma^2(M-2)\sigma_\varepsilon^2)+ \\ &-2m(\tilde{\lambda}^2M\sigma_f^2+\gamma^2\sigma_\varepsilon^2)+M(\tilde{\lambda}^2M\sigma_f^2+\gamma^2\sigma_\varepsilon^2)\Big), \end{aligned} \tag{6.18}$$

where we have defined the excess leverage as $\tilde{\lambda} \equiv \lambda - 1$ and for notational simplicity we do not label λ and m with the time index $t-1$. Hence, at the time scale of the portfolio rebalancing the variance and the covariance of the returns, $r_{i,s}$[13], are

$$Var[r_{i,s}] = \sigma_\varepsilon^2 + \sigma_f^2 + Var[e_{i,s}] \equiv \theta_0 \tag{6.19}$$

$$Cov[r_{i,s},r_{j,s}] = \sigma_f^2 + Cov[e_{i,s},e_{j,s}] \equiv \psi_0, \tag{6.20}$$

respectively. The variance and the covariance of the returns at the time scale of the portfolio decisions, *i.e.* $R_{i,t} \equiv \sum_{k=1}^{n} r_{i,t-1+k/n}$, are

$$Var[R_{i,t}] = \mathbb{E}[(R_{i,t})^2] = n\mathbb{E}[r_{i,s}^2] + 2(n-1)\mathbb{E}[r_{i,s}r_{i,s-1/n}] + 2(n-2)\mathbb{E}[r_{i,s}r_{i,s-2/n}] + \dots,$$

$$Cov[R_{i,t},R_{j,t}] = \mathbb{E}[R_{i,t}R_{j,t}] = n\mathbb{E}[r_{i,s}r_{j,s}] + 2(n-1)\mathbb{E}[r_{i,s}r_{j,s-1/n}] + 2(n-2)\mathbb{E}[r_{i,s}r_{j,s-2/n}] + ..., \ j \neq i,$$

because of the stationarity of the return dynamics between $t-1$ and t. By defining $\theta_k \equiv \mathbb{E}[r_{i,s}r_{i,s-k/n}]$ and $\psi_k \equiv \mathbb{E}[r_{i,s}r_{j,s-k/n}]$ with $j \neq i$, the previous formulas read as

[13] For simplicity, we are assuming that the asset returns are centered around the mean.

$$Var[R_{i,t}] = n\,\theta_0 + 2n\sum_{k=1}^{n}\theta_k - 2\sum_{k=1}^{n} k\,\theta_k, \tag{6.21}$$

$$Cov[R_{i,t}, R_{j,t}] = n\,\psi_0 + 2n\sum_{k=1}^{n}\psi_k - 2\sum_{k=1}^{n} k\,\psi_k. \tag{6.22}$$

In a similar fashion of the univariate case, by applying recursively the VAR(1) process (6.5) and by taking expectation, we obtain two systems of equations whose solutions are the analytical expressions of the terms $\sum_{k=1}^{n}\theta_k$, $\sum_{k=1}^{n} k\,\theta_k$, $\sum_{k=1}^{n}\psi_k$ and $\sum_{k=1}^{n} k\,\psi_k$ in Equations (6.21) and (6.22).

For notational simplicity, let us define

$$\varphi \equiv \frac{\lambda - 1}{\gamma m}, \tag{6.23}$$

$$\beta \equiv \frac{\lambda - 1}{\gamma m}, \frac{m-1}{M-1} \tag{6.24}$$

such that the matrix of autoregressive coefficients in Equation (6.5) reads as $\boldsymbol{\Phi} = (\varphi - \beta)\mathbb{1} + \beta\mathbf{1}$ where $\mathbb{1}$ is the identity matrix and $\mathbf{1}$ is the matrix whose entries are equal to one.

It is possible to verify that $\Theta_1 \equiv \sum_{k=1}^{n}\theta_k$ and $\Psi_1 \equiv \sum_{k=1}^{n}\psi_k$ are the solutions of the following linear system of equations,

$$\begin{cases} (1-\varphi)\Theta_1 - \beta(M-1)\Psi_1 & = \varphi\,\theta_0 + \beta(M-1)\,\psi_0 \\ -\beta\,\Theta_1 + (1-(\varphi+\beta(M-2)))\Psi_1 & = \beta\,\theta_0 + (\varphi+\beta(M-2))\,\psi_0, \end{cases} \tag{6.25}$$

where we have assumed that $\theta_n \ll 1$ and $\psi_n \ll 1$ which holds when $n \gg 1$, see [Tsay, 2005].

Similarly, $\Theta_2 \equiv \sum_{k=1}^{n} k\,\theta_k$ and $\Psi_2 \equiv \sum_{k=1}^{n} k\,\psi_q$ are the solutions of the following linear system of equations,

$$\begin{cases} (1-\varphi)\Theta_2 - \beta(M-1)\Psi_2 & = \varphi\,\Theta_1 + \beta(M-1)\,\Psi_1 \\ -\beta\,\Theta_2 + (1-(\varphi+\beta(M-2)))\Psi_2 & = \beta\,\Theta_1 + (\varphi+\beta(M-2))\,\Psi_1, \end{cases} \tag{6.26}$$

where we have assumed $\theta_n \ll 1$, $n\,\theta_n \ll 1$, $\psi_n \ll 1$ and $n\,\psi_n \ll 1$ when $n \gg 1$.

According to our assumptions about statistical equivalence of risky investments, the covariance matrix is a matrix whose diagonal entries are equal to each other, *i.e.* $\bar{\Sigma}_d + \bar{\Sigma}_u \equiv Var[R_{i,t}]$, $\forall i$, and the same for the off-diagonal ones, *i.e.* $\bar{\Sigma}_u \equiv Cov[R_{i,t}, R_{j,t}]$, $\forall i \neq j$. Hence, for $n \gg 1$ the covariance matrix of the returns in Equation (6.5) at the time scale of the portfolio decisions is

$$\bar{\boldsymbol{\Sigma}} = \bar{\Sigma}_d\mathbb{1} + \bar{\Sigma}_u\mathbf{1}, \tag{6.27}$$

with

$$\begin{cases} \bar{\Sigma}_d = n\left((\theta_0 - \psi_0) + 2(\Theta_1 - \Psi_1) - \frac{2}{n}(\Theta_2 - \Psi_2)\right) \\ \bar{\Sigma}_u = n\left(\psi_0 + 2\Psi_1 - \frac{2}{n}\Psi_2\right) \end{cases}, \tag{6.28}$$

and by substituting the formulas (6.17), (6.18), (6.19), and (6.20) in Equation (6.28), the covariance matrix is obtained explicitly.

Maximum likelihood estimation

In the ideal procedure of estimating the VAR(1) process (6.5)s, *i.e.*

$$\boldsymbol{r}_s = \boldsymbol{\Phi}\,\boldsymbol{r}_{s-1/n} + \boldsymbol{\eta}_s,\ s = t-1+k/n,\ k = 1,2,...,n, \tag{6.29}$$

we can obtain the maximum likelihood estimators of $\bar{\Sigma}_d$ and $\bar{\Sigma}_u$, *i.e.* $\widehat{\bar{\Sigma}}_d$ and $\widehat{\bar{\Sigma}}_u$.

In Equation (6.29), $\boldsymbol{\eta}_s \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\eta)$, $\forall s = t-1+k/n$, $k = 1,...,n$ where $\boldsymbol{\Sigma}_\eta = \sigma_\varepsilon^2 \mathbb{1} + \sigma_f^2 \mathbf{1}$ with $\sigma_\varepsilon, \sigma_f > 0$ and $\boldsymbol{\Phi} = (\varphi - \beta)\mathbb{1} + \beta \mathbf{1}$ with φ and β defined in Equations (6.23) and (6.24).

Given the observations between $t-1$ and t, the likelihood of the VAR(1) process is

$$\mathbb{P}[\boldsymbol{r}_{t-1+1/n}, \boldsymbol{r}_{t-1+2/n}, ..., \boldsymbol{r}_t | \boldsymbol{r}_{t-1}] = \prod_{k=1}^{n} \mathbb{P}[\boldsymbol{r}_{t-1+k/n} | \boldsymbol{r}_{t-1+(k-1)/n}] = \prod_{k=1}^{n} \mathcal{N}(\boldsymbol{\Phi}\boldsymbol{r}_{t-1+(k-1)/n}, \boldsymbol{\Sigma}_\eta),$$

because the conditional distribution of the returns is multivariate Gaussian. Hence, the log-likelihood is

$$L(\boldsymbol{\Phi}, \boldsymbol{\Sigma}_\eta) = -\frac{n}{2}\log|\boldsymbol{\Sigma}_\eta| - \frac{1}{2}\sum_{k=1}^{n} (\boldsymbol{r}_{t-1+k/n} - \boldsymbol{\Phi}\boldsymbol{r}_{t-1+(k-1)/n})^\intercal \boldsymbol{\Sigma}_\eta^{-1} (\boldsymbol{r}_{t-1+k/n} - \boldsymbol{\Phi}\boldsymbol{r}_{t-1+(k-1)/n}) \tag{6.30}$$

and the maximum likelihood estimators of parameters, *i.e.* $\hat{\varphi}$, $\hat{\beta}$, $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_f^2$, can be obtained as solution of the following system of equations

$$\begin{cases} \frac{\partial L(\varphi,\beta,\sigma_\varepsilon,\sigma_f)}{\partial \varphi} = 0 \\ \frac{\partial L(\varphi,\beta,\sigma_\varepsilon,\sigma_f)}{\partial \beta} = 0 \\ \frac{\partial L(\varphi,\beta,\sigma_\varepsilon,\sigma_f)}{\partial \sigma_\varepsilon} = 0 \\ \frac{\partial L(\varphi,\beta,\sigma_\varepsilon,\sigma_f)}{\partial \sigma_f} = 0. \end{cases}$$

Let us notice that the maximum likelihood estimators of parameters are only functions of the observed fast variables $\{\boldsymbol{r}_{t-1+k/n}\}_{k=0,1,2,\ldots,n}$, *i.e.* $\hat{\varphi} \equiv \hat{\varphi}(\{\boldsymbol{r}_{t-1+k/n}\}_{k=0,1,2,\ldots,n})$, $\hat{\beta} \equiv \hat{\beta}(\{\boldsymbol{r}_{t-1+k/n}\}_{k=0,1,2,\ldots,n})$, $\hat{\sigma}_\varepsilon \equiv \hat{\sigma}_\varepsilon(\{\boldsymbol{r}_{t-1+k/n}\}_{k=0,1,2,\ldots,n})$ and $\hat{\sigma}_f \equiv \hat{\sigma}_f(\{\boldsymbol{r}_{t-1+k/n}\}_{k=0,1,2,\ldots,n})$.

Then, according to the following standard result of time series analysis, see [Tsay, 2005],

$$(\hat{\theta}_0 - \hat{\psi}_0)\mathbb{1} + \hat{\psi}_0 \mathbf{1} \equiv \sum_{k=0}^{\infty} \widehat{\boldsymbol{\Phi}}^k \widehat{\boldsymbol{\Sigma}}_\eta (\widehat{\boldsymbol{\Phi}}^k)^\intercal,$$

where $\widehat{\boldsymbol{\Sigma}}_\eta = \hat{\sigma}_\varepsilon^2 \mathbb{1} + \hat{\sigma}_f^2 \mathbf{1}$ and $\widehat{\boldsymbol{\Phi}} = (\hat{\varphi} - \hat{\beta})\mathbb{1} + \hat{\beta}\mathbf{1}$, we are able to compute the maximum likelihood estimators of the variance and covariance at the fast time scale, *i.e.* $\hat{\theta}_0 \equiv \widehat{Var}[r_{i,s}]$ and $\hat{\psi}_0 \equiv \widehat{Cov}[r_{i,s}, r_{j,s}]$ with $j \neq i$. Hence, by computing $\hat{\Theta}_1 \equiv \Theta_1(\hat{\theta}_0, \hat{\psi}_0, \hat{\varphi}, \hat{\beta})$, $\hat{\Psi}_1 \equiv \Psi_1(\hat{\theta}_0, \hat{\psi}_0, \hat{\varphi}, \hat{\beta})$, $\hat{\Theta}_2 \equiv \Theta_2(\hat{\theta}_0, \hat{\psi}_0, \hat{\varphi}, \hat{\beta})$ and $\hat{\Psi}_2 \equiv \Psi_2(\hat{\theta}_0, \hat{\psi}_0, \hat{\varphi}, \hat{\beta})$, the maximum likelihood estimators of the (diversifiable) variance and covariance of the return dynamics (6.29) aggregated at the slow time scale are

$$\begin{cases} \widehat{\Sigma}_d = n\left((\hat{\theta}_0 - \hat{\psi}_0) + 2(\hat{\Theta}_1 - \hat{\Psi}_1) - \frac{2}{n}(\hat{\Theta}_2 - \hat{\Psi}_2)\right) \\ \widehat{\Sigma}_u = n\left(\hat{\psi}_0 + 2\hat{\Psi}_1 - \frac{2}{n}\hat{\Psi}_2\right). \end{cases} \tag{6.31}$$

Asymptotic limit $n \to \infty$

In the limit $n \to \infty$ the last terms in (6.28) are negligible. However, the terms $n((\theta_0 - \psi_0) + 2(\Theta_1 - \Psi_1))$ and $n(\psi_0 + 2\Psi_1)$ remain finite when n goes to infinity because of finite $\Sigma_\varepsilon = \lim_{n\to\infty} n\sigma_\varepsilon^2$ and $\Sigma_f = \lim_{n\to\infty} n\sigma_f^2$.

By defining

$$\Theta_0 \equiv \Sigma_\varepsilon + \Sigma_f - \frac{\tilde{\lambda}^2\left(m^2(\Sigma_\varepsilon(\tilde{\lambda}^2 - \gamma^2(M-1)) + \Sigma_f(\tilde{\lambda}^2 - \gamma^2(M-1)^2)) + 2m(M(\Sigma_\varepsilon(\gamma^2 - \tilde{\lambda}^2) - \tilde{\lambda}^2\Sigma_f) - \gamma^2\Sigma_\varepsilon) + M(M(\Sigma_\varepsilon(\tilde{\lambda}^2 - \gamma^2) + \tilde{\lambda}^2\Sigma_f) + \gamma^2\Sigma_\varepsilon)\right)}{(\gamma^2 - \tilde{\lambda}^2)(m^2(\gamma^2(M-1)^2 - \tilde{\lambda}^2) + 2 - \tilde{\lambda}^2 mM - \tilde{\lambda}^2 M^2)}, \tag{6.32}$$

$$\Psi_0 \equiv \Sigma_f - \frac{\tilde{\lambda}^2\left(m^2(\Sigma_f(\tilde{\lambda}^2 - \gamma^2(M-1)^2) - \gamma^2(M-2)\Sigma_\varepsilon) - 2m(\tilde{\lambda}^2 M\Sigma_f + \gamma^2\Sigma_\varepsilon) + M(\tilde{\lambda}^2 M\Sigma_f + \gamma^2\Sigma_\varepsilon)\right)}{(\gamma^2 - \tilde{\lambda}^2)(m^2(\gamma^2(M-1)^2 - \tilde{\lambda}^2) + 2 - \tilde{\lambda}^2 mM - \tilde{\lambda}^2 M^2)}, \tag{6.33}$$

in the asymptotic limit $n \to \infty$, it is

$$\begin{cases} \widetilde{\Sigma}_d = \Theta_0 - \Psi_0 + 2\frac{(\varphi\beta(M-2) + \varphi^2 - \varphi - \beta^2(M-1))(\Theta_0 - \Psi_0) + \beta\Theta_0 + \beta(M-1)\Psi_0 - (M-2)\Psi_0}{(1+\beta-\varphi)(\beta(M-1)+\varphi-1)} \\ \widetilde{\Sigma}_u = \Psi_0 + 2\frac{-\beta\Theta_0 + (\varphi\beta(M-2) + \varphi^2 - \varphi - \beta^2(M-1) + \varphi\beta(M-2))\Psi_0}{(1+\beta-\varphi)(\beta(M-1)+\varphi-1)}. \end{cases} \tag{6.34}$$

Hence, by substituting (6.32) and (6.33) in (6.34) and since $\tilde{\lambda} \equiv \tilde{\lambda}_{t-1}$, $m_{t-1} \equiv m_{t-1}(\lambda_{t-1}, \Sigma_{d,t-1}, \Sigma_{u,t-1})$ according to Equation (5.12), we obtain the explicit analytical expressions for $\widetilde{\Sigma}_d$ and $\widetilde{\Sigma}_u$ in Equation (6.7).

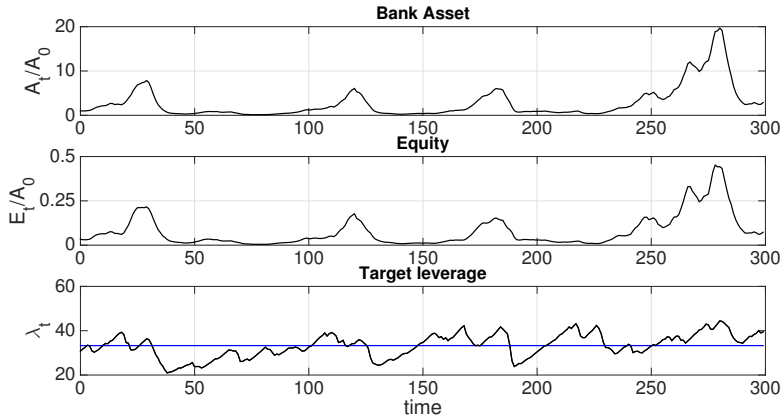


Fig. 6.10 Numerical simulation of reduced model with $n = 1$, *i.e.* both the fast and the slow time scales coincide. Top panel: evolution of asset size normalized by the initial value A_0. Middle panel: evolution of equity. Bottom panel: dynamics for the financial leverage. We simulate the model in a time window $T = 300$. The other model parameters are: $\gamma = 40$ (liquidity parameter), $\alpha = 1.64$ (VaR parameter), $\mu - r_L = 0.08$ (net interest margin), $\sigma_\varepsilon = 0.04$ (idiosyncratic volatility), $\omega = 0.9$ (memory).

Single time scale

Here we consider the case $n = 1$ for the reduced model, *i.e.* bank updates the risk expectation at the same frequency of portfolio rebalancing. When the time scales are the same, the process of expectation formation reduces to model the variance as in IGARCH-type models [Engle and Bollerslev, 1986],

$$\sigma_t^2 = \omega\sigma_{t-1}^2 + (1-\omega)r_t^2, \tag{6.35}$$

with memory parameter ω.

This approach is similar to the RiskMetrics one, see [Longerstaey and Spencer, 1996], and, as highlighted by [Bauwens et al., 2006], despite the simplicity, this kind of model is usually adopted by practitioners with a decay factor ω equal to 0.94 for daily data and 0.97 for monthly data.

The model specified by Equation (6.6) with (6.35) as the process of expectation formation is close to the one presented in [Aymanns and Farmer, 2015] which differs for the price dynamics. Aymanns and Farmer assume that the equity value is fixed and there is only one investor in the market, thus it is $A_t \equiv p_t = \lambda_t E$ with p_t the price of the traded risky investment[14]. Given the log return $r_t = \log \frac{p_t}{p_{t-1}}$ and by substituting the price with $\lambda_t E$, a two-dimensional dynamical system is obtained. With this further assumption, the two models coincide.

[14] Here, the number of shares is normalized to one.

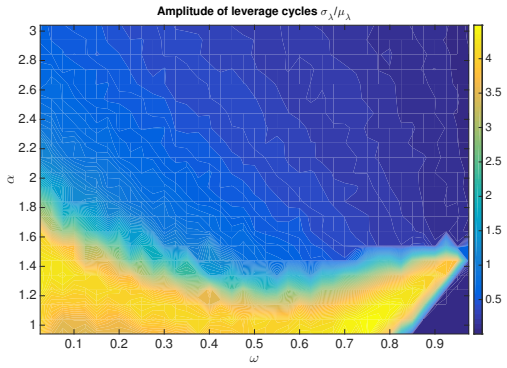


Fig. 6.11 The amplitude of leverage cycles as a function of the parameter of the Value-at-Risk α and of the memory parameter ω. As measure of the amplitude of leverage cycles we use the standard deviation of data, σ_λ, divided by the mean μ_λ. We run the reduced model for $T = 1000$ with idiosyncratic risk $\sigma_\varepsilon^2 = 0.05$ and liquidity parameter $\gamma = 40$ by varying $\alpha \in [1,3]$ and $\omega \in (0,1)$. For each point (ω, α) in the plot, we average over 50 seeds to obtain the considered value for $\frac{\sigma_\lambda}{\mu_\lambda}$.

Figure 6.10 shows the simulated dynamics of bank asset, equity and financial leverage of the reduced model with $n = 1$. We notice that fluctuations in the value of target leverage are strongly correlated with fluctuations of the equity value. The changes in the equity value reflects the price movements. Asset size evolves in such a way to have financial leverage equal to its targeting value. Hence, boom and bust of the price are driven by leverage cycles which in turn are determined by risk perceptions.

As indicator of dynamical instabilities of the financial system we consider an adimensional measure of the amplitude of leverage cycles, specifically the standard deviation σ_λ of target leverage in model simulations divided by its mean μ_λ.

In Figure 6.11 we show the normalized amplitude of leverage oscillations as a function of the VaR parameter α and of the memory parameter ω. As expected, stringent capital constraints (large α) and large memory in forecasting risk (large ω) stabilize the dynamics of the financial system. Therefore, increasing α and ω is followed by smaller leverage oscillations and as a consequence the dynamics of financial system is close to the fixed point equilibrium. The parameter space for α and ω can be approximately divided in two regions. A region of instability for small value of α where strong oscillations occur independently on the memory of risk expectations and a relatively stable region for large α in which the coefficient $\frac{\sigma_\lambda}{\mu_\lambda}$ is smaller. In this region, larger memory tends to stabilize the system. Finally, let us notice that in the unstable region oscillations in financial leverage disappear for $\omega \to 1^-$. We do not fully understand why we observe this but we suppose the reason being the very high inertia in risk expectations which turns off the leverage oscillations.

Conclusions

This thesis contributes to the network literature in several directions.

We have generalized the problem of community detectability to static networks with link direction, thus showing a new class of macroscopic structures of interest, and to dynamic networks with link persistence, thus pointing out time-lagged inference, *i.e.* the identification of past communities rather than present ones. Hence, we have solved analytically the detectability problem, in both cases. We have shown that asymmetry in link directions makes easier the detection of communities. Moreover, when the direction asymmetry is above a given threshold, the community structure is always detectable, independently from the level of assortativity of the network. In the dynamic case, we show how to quantify analytically the effect of introducing link persistence, in particular by assessing what is the delay in community inference. Accordingly, our novel algorithm, the Lagged Snapshot Dynamic (LSD) algorithm, is able to identify the estimation bias due to link persistence, thus leading to correct labeling.

Then, we have studied the coupling between the Markovian behavior of link persistence with the latent autoregressive dynamics for a suitable dynamic generalization of the fitness model. We have introduced a novel inference method, based on Maximum A Posteriori approach, to disentangle the role of the two linkage mechanisms in the network dynamics, thus leading to unbiased estimation of both the latent dynamics and the model parameters. The proposed model can be used to recognize preferential relations in dynamic networks, such as preferential lending in the interbank market. We expect, *e.g.* , similar results also for the detection of preferential relations in social networks. Moreover, the proposed inference method is in effect a statistical filtering algorithm and can be adopted for link prediction.

In the second part of the thesis, we have studied theoretically how expectation feedbacks in a model of systemic risk mediated by fire sale spillovers and overlapping portfolios may induce dynamical instabilities. We have characterized analytically the dynamical outcomes of the proposed model, thus showing that expectation feedbacks lead to a period-doubling cascade to chaos. To the best of our knowledge, this is the first analytical proof of chaotic behavior in this context. The obtained results have some policy implications, in particular we point out that the length of the

past time window of observations used in forming risk expectations is crucial, short time windows having the potential of destabilizing the systemic stability.

We believe our results deserve attention because they give new insights in the process of network evolution. In particular, preferential relations in real-world networks tend to display link persistence, because of frequently repeated connections between some sets of nodes. Then, link persistence gives rise to the time-lagged inference, a novel phenomenon in the statistical inference of models combining a Markovian behavior for links with a latent dynamics for node-specific characteristics. This is a crucial aspect to be taken into account when learning the network dynamics: this work is the first analytical study in this direction, to the best of our knowledge.

It is worth emphasizing that some key ideas developed for the proposed inference algorithms can help in creating a unified methodology for disentangling the two linkage behaviors associated with link (or more generally network metrics) persistence and latent dynamics of node-specific characteristics. Future outlooks of this work would consider building a class of nested network models, where describing either the explicit persistence of network metrics or the latent dynamics governing topologies, the full model containing both. Then, either MMAP and LSD approaches would be generalized for model estimation and the asymptotic properties be studied for assessing the general regularity conditions for identifiability and consistency of estimators. Finally, applications may be countless, any time we need to disentangle different linkage mechanisms in temporal networks. In fact, nested network models represent a powerful tool in this direction, because of the possibility of assessing time by time the importance of a given mechanism by quantifying changes in link probability, from the bottom to the top of the nested structure of models.

Bibliography

References

Abbe, 2017. Abbe, E. (2017). Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531.

Abbe et al., 2016. Abbe, E., Bandeira, A. S., and Hall, G. (2016). Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487.

Adrian and Shin, 2010. Adrian, T. and Shin, H. S. (2010). Liquidity and leverage. *Journal of financial intermediation*, 19(3):418–437.

Adrian and Shin, 2013. Adrian, T. and Shin, H. S. (2013). Procyclical leverage and value-at-risk. *The Review of Financial Studies*, 27(2):373–403.

Airoldi et al., 2008. Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of machine learning research*, 9(Sep):1981–2014.

Allen and Babus, 2009. Allen, F. and Babus, A. (2009). Networks in finance. *The network challenge: strategy, profit, and risk in an interlinked world*, 367.

Allen and Gale, 2000. Allen, F. and Gale, D. (2000). Financial contagion. *Journal of political economy*, 108(1):1–33.

Amaral et al., 2000. Amaral, L. A. N., Scala, A., Barthelemy, M., and Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the national academy of sciences*, 97(21):11149–11152.

Amini et al., 2016. Amini, H., Cont, R., and Minca, A. (2016). Resilience to contagion in financial networks. *Mathematical finance*, 26(2):329–365.

Anderson, 1972. Anderson, P. W. (1972). More is different. *Science*, 177(4047):393–396.

Aymanns et al., 2016. Aymanns, C., Caccioli, F., Farmer, J. D., and Tan, V. W. (2016). Taming the basel leverage cycle. *Journal of financial stability*, 27:263–277.

Aymanns and Farmer, 2015. Aymanns, C. and Farmer, J. D. (2015). The dynamics of the leverage cycle. *Journal of Economic Dynamics and Control*, 50:155–179.

Bacharach, 1965. Bacharach, M. (1965). Estimating nonnegative matrices from marginal data. *International Economic Review*, 6(3):294–310.

Bao et al., 2013. Bao, T., Duffy, J., and Hommes, C. (2013). Learning, forecasting and optimizing: An experimental study. *European Economic Review*, 61:186–204.

Bao et al., 2012. Bao, T., Hommes, C., Sonnemans, J., and Tuinstra, J. (2012). Individual expectations, limited rationality and aggregate outcomes. *Journal of Economic Dynamics and Control*, 36(8):1101–1120.

Barabási et al., 2000. Barabási, A.-L., Albert, R., and Jeong, H. (2000). Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: statistical mechanics and its applications*, 281(1-4):69–77.

Barucca and Lillo, 2016. Barucca, P. and Lillo, F. (2016). Disentangling bipartite and core-periphery structure in financial networks. *Chaos, Solitons & Fractals*, 88:244–253.

Barucca and Lillo, 2018. Barucca, P. and Lillo, F. (2018). The organization of the interbank network and how ecb unconventional measures affected the e-mid overnight market. *Computational Management Science*, 15(1):33–53.

Barucca et al., 2018. Barucca, P., Lillo, F., Mazzarisi, P., and Tantari, D. (2018). Disentangling group and link persistence in dynamic stochastic block models. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(12):123407.

Bassett et al., 2013. Bassett, D. S., Porter, M. A., Wymbs, N. F., Grafton, S. T., Carlson, J. M., and Mucha, P. J. (2013). Robust detection of dynamic community structure in networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(1):013142.

Bassett and Deride, 2019. Bassett, R. and Deride, J. (2019). Maximum a posteriori estimators as a limit of bayes estimators. *Mathematical Programming*, 174(1-2):129–144.

Battiston et al., 2012. Battiston, S., Puliga, M., Kaushik, R., Tasca, P., and Caldarelli, G. (2012). Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Scientific reports*, 2:541.

Bauwens et al., 2006. Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate garch models: a survey. *Journal of applied econometrics*, 21(1):79–109.

Bazzi et al., 2016. Bazzi, M., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., and Howison, S. D. (2016). Community detection in temporal multilayer networks, with an application to correlation networks. *Multiscale Modeling & Simulation*, 14(1):1–41.

BCBS III, 2011. BCBS III, B. (2011). A global regulatory framework for more resilient banks and banking systems. *Bank for International Settlements, Basel Committee on Banking Supervision, Basel*.

Beale et al., 2011. Beale, N., Rand, D. G., Battey, H., Croxson, K., May, R. M., and Nowak, M. A. (2011). Individual versus systemic risk and the regulator's dilemma. *Proceedings of the National Academy of Sciences*, 108(31):12647–12652.

Bhattacharya and Majumdar, 2003. Bhattacharya, R. and Majumdar, M. (2003). Random dynamical systems: a review. *Economic Theory*, 23(1):13–38.

Bianconi and Barabási, 2001. Bianconi, G. and Barabási, A.-L. (2001). Competition and multiscaling in evolving networks. *EPL (Europhysics Letters)*, 54(4):436.

Blondel et al., 2008. Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Boccaletti et al., 2014. Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., Sendina-Nadal, I., Wang, Z., and Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122.

Bordenave et al., 2015. Bordenave, C., Lelarge, M., and Massoulié, L. (2015). Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357. IEEE.

Boss et al., 2004. Boss, M., Elsinger, H., Summer, M., and Thurner 4, S. (2004). Network topology of the interbank market. *Quantitative finance*, 4(6):677–684.

Bräuning and Koopman, 2016. Bräuning, F. and Koopman, S. J. (2016). The dynamic factor network model with an application to global credit-risk.

Brock and Hommes, 1997. Brock, W. A. and Hommes, C. H. (1997). A rational route to randomness. *Econometrica: Journal of the Econometric Society*, pages 1059–1095.

Brock et al., 2009. Brock, W. A., Hommes, C. H., and Wagener, F. O. O. (2009). More hedging instruments may destabilize markets. *Journal of Economic Dynamics and Control*, 33(11):1912–1928.

Brunnermeier and Pedersen, 2008. Brunnermeier, M. K. and Pedersen, L. H. (2008). Market liquidity and funding liquidity. *The review of financial studies*, 22(6):2201–2238.

Caccioli et al., 2012. Caccioli, F., Catanach, T. A., and Farmer, J. D. (2012). Heterogeneity, correlations and financial contagion. *Advances in Complex Systems*, 15(supp02):1250058.

Caccioli et al., 2015. Caccioli, F., Farmer, J. D., Foti, N., and Rockmore, D. (2015). Overlapping portfolios, contagion, and financial stability. *Journal of Economic Dynamics and Control*, 51:50–63.

Caccioli et al., 2014. Caccioli, F., Shrestha, M., Moore, C., and Farmer, J. D. (2014). Stability analysis of financial contagion due to overlapping portfolios. *Journal of Banking & Finance*, 46:233–245.

Caimo and Friel, 2011. Caimo, A. and Friel, N. (2011). Bayesian inference for exponential random graph models. *Social Networks*, 33(1):41–55.

Caldarelli et al., 2002. Caldarelli, G., Capocci, A., De Los Rios, P., and Munoz, M. A. (2002). Scale-free networks from varying vertex intrinsic fitness. *Physical review letters*, 89(25):258702.

Caldarelli et al., 2013. Caldarelli, G., Chessa, A., Pammolli, F., Gabrielli, A., and Puliga, M. (2013). Reconstructing a credit network. *Nature Physics*, 9(3):125.

Casella and Berger, 2002. Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Casteigts et al., 2012. Casteigts, A., Flocchini, P., Quattrociocchi, W., and Santoro, N. (2012). Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems*, 27(5):387–408.

Castellacci and Choi, 2015. Castellacci, G. and Choi, Y. (2015). Modeling contagion in the eurozone crisis via dynamical systems. *Journal of Banking & Finance*, 50:400–410.

Chatterjee et al., 2011. Chatterjee, S., Diaconis, P., Sly, A., et al. (2011). Random graphs with a given degree sequence. *The Annals of Applied Probability*, 21(4):1400–1435.

Chen et al., 2003. Chen, Z. et al. (2003). Bayesian filtering: From kalman filters to particle filters, and beyond. *Statistics*, 182(1):1–69.

Chiarella and He, 2003. Chiarella, C. and He, X.-Z. (2003). Heterogeneous beliefs, risk, and learning in a simple asset-pricing model with a market maker. *Macroeconomic Dynamics*, 7(4):503–536.

Choi and Douady, 2012. Choi, Y. and Douady, R. (2012). Financial crisis dynamics: attempt to define a market instability indicator. *Quantitative Finance*, 12(9):1351–1365.

Cifuentes et al., 2005. Cifuentes, R., Ferrucci, G., and Shin, H. S. (2005). Liquidity risk and contagion. *Journal of the European Economic Association*, 3(2-3):556–566.

Cimini et al., 2015. Cimini, G., Squartini, T., Garlaschelli, D., and Gabrielli, A. (2015). Systemic risk analysis on reconstructed economic and financial networks. *Scientific reports*, 5:15758.

Cocco et al., 2009. Cocco, J. F., Gomes, F. J., and Martins, N. C. (2009). Lending relationships in the interbank market. *Journal of Financial Intermediation*, 18(1):24–48.

Condon and Karp, 2001. Condon, A. and Karp, R. M. (2001). Algorithms for graph partitioning on the planted partition model. *Random Structures & Algorithms*, 18(2):116–140.

Constantinides, 1986. Constantinides, G. M. (1986). Capital market equilibrium with transaction costs. *Journal of political Economy*, 94(4):842–862.

Cont and Wagalath, 2013. Cont, R. and Wagalath, L. (2013). Running for the exit: distressed selling and endogenous correlation in financial markets. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 23(4):718–741.

Corsi et al., 2016. Corsi, F., Marmi, S., and Lillo, F. (2016). When micro prudence increases macro risk: The destabilizing effects of financial innovation, leverage, and diversification. *Operations Research*, 64(5):1073–1088.

Crawford, 1991. Crawford, J. D. (1991). Introduction to bifurcation theory. *Reviews of Modern Physics*, 63(4):991.

Csiszár et al., 2012. Csiszár, V., Hussami, P., Komlós, J., Móri, T. F., Rejtő, L., and Tusnády, G. (2012). When the degree sequence is a sufficient statistic. *Acta Mathematica Hungarica*, 134(1-2):45–53.

Danıelsson et al., 2004. Danıelsson, J., Shin, H. S., and Zigrand, J.-P. (2004). The impact of risk regulation on price dynamics. *Journal of Banking & Finance*, 28(5):1069–1087.

Danielsson et al., 2012. Danielsson, J., Shin, H. S., and Zigrand, J.-P. (2012). Endogenous and systemic risk. In *Quantifying systemic risk*, pages 73–94. University of Chicago Press.

Danon et al., 2005. Danon, L., Diaz-Guilera, A., Duch, J., and Arenas, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008.

De Masi et al., 2006. De Masi, G., Iori, G., and Caldarelli, G. (2006). Fitness model for the italian interbank money market. *Physical Review E*, 74(6):066112.

Decelle et al., 2011a. Decelle, A., Krzakala, F., Moore, C., and Zdeborová, L. (2011a). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106.

Decelle et al., 2011b. Decelle, A., Krzakala, F., Moore, C., and Zdeborová, L. (2011b). Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107(6):065701.

Dempster et al., 1977. Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Devaney, 2008. Devaney, R. (2008). *An introduction to chaotic dynamical systems*. Westview press.

Di Gangi et al., 2018. Di Gangi, D., Lillo, F., and Pirino, D. (2018). Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. *Journal of Economic Dynamics and Control*, 94:117–141.

Di Maggio et al., 2017. Di Maggio, M., Kermani, A., and Song, Z. (2017). The value of trading relations in turbulent times. *Journal of Financial Economics*, 124(2):266–284.

Dodds and Payne, 2009. Dodds, P. S. and Payne, J. L. (2009). Analysis of a threshold model of social contagion on degree-correlated networks. *Physical Review E*, 79(6):066115.

Duarte and Eisenbach, 2018. Duarte, F. and Eisenbach, T. M. (2018). Fire-sale spillovers and systemic risk. *FRB of New York Staff Report*, (645).

Durante and Dunson, 2014. Durante, D. and Dunson, D. B. (2014). Bayesian dynamic financial networks with time-varying predictors. *Statistics & Probability Letters*, 93:19–26.

Durante et al., 2016. Durante, D., Dunson, D. B., et al. (2016). Locally adaptive dynamic networks. *The Annals of Applied Statistics*, 10(4):2203–2232.

Dyer and Frieze, 1989. Dyer, M. E. and Frieze, A. M. (1989). The solution of some random np-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489.

Eckmann and Ruelle, 1985. Eckmann, J.-P. and Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. In *The theory of chaotic attractors*, pages 273–312. Springer.

Eilers and Marx, 1996. Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, pages 89–102.

Engle and Bollerslev, 1986. Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, 5(1):1–50.

Erdös and Rényi, 1959. Erdös, P. and Rényi, A. (1959). On random graphs, i. *Publicationes Mathematicae*, 6:290–297.

Farmer et al., 2012. Farmer, J. D., Gallegati, M., Hommes, C., Kirman, A., Ormerod, P., Cincotti, S., Sanchez, A., and Helbing, D. (2012). A complex systems approach to constructing better models for managing financial markets and the economy. *The European Physical Journal Special Topics*, 214(1):295–324.

Feigenbaum, 1978. Feigenbaum, M. J. (1978). Quantitative universality for a class of nonlinear transformations. *Journal of statistical physics*, 19(1):25–52.

Finger et al., 2013. Finger, K., Fricke, D., and Lux, T. (2013). Network analysis of the e-mid overnight money market: the informational value of different aggregation levels for intrinsic dynamic processes. *Computational Management Science*, 10(2-3):187–211.

Finger and Lux, 2017. Finger, K. and Lux, T. (2017). Network formation in the interbank money market: An application of the actor-oriented model. *Social Networks*, 48:237–249.

Fortunato, 2010. Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3-5):75–174.

Franz et al., 2001. Franz, S., Mézard, M., Ricci-Tersenghi, F., Weigt, M., and Zecchina, R. (2001). A ferromagnet with a glass transition. *EPL (Europhysics Letters)*, 55(4):465.

Fricke and Lux, 2015. Fricke, D. and Lux, T. (2015). Core–periphery structure in the overnight money market: evidence from the e-mid trading platform. *Computational Economics*, 45(3):359–395.

Friedman et al., 2001. Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

Friel et al., 2016. Friel, N., Rastelli, R., Wyse, J., and Raftery, A. E. (2016). Interlocking directorates in irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences*, 113(24):6629–6634.

Gabrielli et al., 2014. Gabrielli, A., Battiston, S., Caldarelli, G., Musmeci, N., and Puliga, M. (2014). Reconstructing topological properties of complex networks from partial information using the fitness model. In *APS March Meeting Abstracts*.

Gabrielsen, 1978. Gabrielsen, A. (1978). Consistency and identifiability. *Journal of Econometrics*, 8(2):261–263.

Gai and Kapadia, 2010. Gai, P. and Kapadia, S. (2010). Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2120):2401–2423.

Garlaschelli and Loffredo, 2004a. Garlaschelli, D. and Loffredo, M. I. (2004a). Fitness-dependent topological properties of the world trade web. *Physical review letters*, 93(18):188701.

Garlaschelli and Loffredo, 2004b. Garlaschelli, D. and Loffredo, M. I. (2004b). Patterns of link reciprocity in directed networks. *Physical review letters*, 93(26):268701.

Gauvain and Lee, 1994. Gauvain, J.-L. and Lee, C.-H. (1994). Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE transactions on speech and audio processing*, 2(2):291–298.

Geanakoplos, 2010. Geanakoplos, J. (2010). The leverage cycle. *NBER macroeconomics annual*, 24(1):1–66.

Ghasemian et al., 2016. Ghasemian, A., Zhang, P., Clauset, A., Moore, C., and Peel, L. (2016). Detectability thresholds and optimal algorithms for community structure in dynamic networks. *Physical Review X*, 6(3):031005.

Gilbert, 1959. Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144.

Giraitis et al., 2016. Giraitis, L., Kapetanios, G., Wetherilt, A., and Žikeš, F. (2016). Estimating the dynamics and persistence of financial networks, with an application to the sterling money market. *Journal of Applied Econometrics*, 31(1):58–84.

Girvan and Newman, 2002. Girvan, M. and Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826.

Gleeson, 2008. Gleeson, J. P. (2008). Cascades on correlated and modular random networks. *Physical Review E*, 77(4):046117.

Gleeson and Cahalane, 2007. Gleeson, J. P. and Cahalane, D. J. (2007). Seed size strongly affects cascades on random networks. *Physical Review E*, 75(5):056103.

Goldenberg et al., 2010. Goldenberg, A., Zheng, A. X., Fienberg, S. E., Airoldi, E. M., et al. (2010). A survey of statistical network models. *Foundations and Trends® in Machine Learning*, 2(2):129–233.

Greenwood et al., 2015. Greenwood, R., Landier, A., and Thesmar, D. (2015). Vulnerable banks. *Journal of Financial Economics*, 115(3):471–485.

Gualdi et al., 2016. Gualdi, S., Cimini, G., Primicerio, K., Di Clemente, R., and Challet, D. (2016). Statistically validated network of portfolio overlaps and systemic risk. *Scientific reports*, 6:39467.

Guegan, 2009. Guegan, D. (2009). Chaos in economics and finance. *Annual Reviews in Control*, 33(1):89–93.

Halling et al., 2016. Halling, M., Yu, J., and Zechner, J. (2016). Leverage dynamics over the business cycle. *Journal of Financial Economics*, 122(1):21–41.

Handcock et al., 2007. Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354.

Hanneke et al., 2010. Hanneke, S., Fu, W., Xing, E. P., et al. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605.

Hatzopoulos et al., 2015. Hatzopoulos, V., Iori, G., Mantegna, R. N., Miccichè, S., and Tumminello, M. (2015). Quantifying preferential trading in the e-mid interbank market. *Quantitative Finance*, 15(4):693–710.

Heaukulani and Ghahramani, 2013. Heaukulani, C. and Ghahramani, Z. (2013). Dynamic probabilistic models for latent feature propagation in social networks. In *International Conference on Machine Learning*, pages 275–283.

Heemeijer et al., 2009. Heemeijer, P., Hommes, C., Sonnemans, J., and Tuinstra, J. (2009). Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation. *Journal of Economic dynamics and control*, 33(5):1052–1072.

Hellmann and Staudigl, 2014. Hellmann, T. and Staudigl, M. (2014). Evolution of social networks. *European Journal of Operational Research*, 234(3):583–596.

Hendrickson and Leland, 1995. Hendrickson, B. and Leland, R. (1995). An improved spectral graph partitioning algorithm for mapping parallel computations. *SIAM Journal on Scientific Computing*, 16(2):452–469.

Ho et al., 2011. Ho, Q., Song, L., and Xing, E. (2011). Evolving cluster mixed-membership blockmodel for time-evolving networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 342–350.

Hoff, 2005. Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the american Statistical association*, 100(469):286–295.

Hoff, 2011. Hoff, P. D. (2011). Hierarchical multilinear models for multiway data. *Computational Statistics & Data Analysis*, 55(1):530–543.

Hoff et al., 2011. Hoff, P. D. et al. (2011). Separable covariance arrays via the tucker product, with applications to multivariate relational data. *Bayesian Analysis*, 6(2):179–196.

Hoff et al., 2002. Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098.

Holland et al., 1983. Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.

Holland and Leinhardt, 1971. Holland, P. W. and Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative group studies*, 2(2):107–124.

Holland and Leinhardt, 1981. Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association*, 76(373):33–50.

Holme and Saramäki, 2012. Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics reports*, 519(3):97–125.

Holme and Saramäki, 2013. Holme, P. and Saramäki, J. (2013). *Temporal networks*. Springer.

Hommes, 2011. Hommes, C. (2011). The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic dynamics and control*, 35(1):1–24.

Hommes, 2013. Hommes, C. (2013). *Behavioral rationality and heterogeneous expectations in complex economic systems*. Cambridge University Press.

Hommes et al., 2005. Hommes, C., Huang, H., and Wang, D. (2005). A robust rational route to randomness in a simple asset pricing model. *Journal of Economic dynamics and control*, 29(6):1043–1072.

Hommes et al., 2012. Hommes, C., Kiseleva, T., Kuznetsov, Y., and Verbic, M. (2012). Is more memory in evolutionary selection (de) stabilizing? *Macroeconomic Dynamics*, 16(3):335–357.

Hommes et al., 2007. Hommes, C., Sonnemans, J., Tuinstra, J., and Van De Velden, H. (2007). Learning in cobweb experiments. *Macroeconomic Dynamics*, 11(S1):8–33.

Hommes et al., 2008. Hommes, C., Sonnemans, J., Tuinstra, J., and Van de Velden, H. (2008). Expectations and bubbles in asset pricing experiments. *Journal of Economic Behavior & Organization*, 67(1):116–133.

Hommes and Sorger, 1998. Hommes, C. and Sorger, G. (1998). Consistent expectations equilibria. *Macroeconomic Dynamics*, 2(3):287–321.

Hommes, 1994. Hommes, C. H. (1994). Dynamics of the cobweb model with adaptive expectations and nonlinear supply and demand. *Journal of Economic Behavior & Organization*, 24(3):315–335.

Hommes and Wagener, 2009. Hommes, C. H. and Wagener, F. (2009). Bounded rationality and learning in complex markets. *Handbook of Economic Complexity*, pages 87–123.

Huang et al., 2013. Huang, X., Vodenska, I., Havlin, S., and Stanley, H. E. (2013). Cascading failures in bi-partite graphs: model for systemic risk propagation. *Scientific reports*, 3:1219.

Hurd and Gleeson, 2013. Hurd, T. R. and Gleeson, J. P. (2013). On watts? cascade model with random link weights. *Journal of Complex Networks*, 1(1):25–43.

Hurd et al., 2017. Hurd, T. R., Gleeson, J. P., and Melnik, S. (2017). A framework for analyzing contagion in assortative banking networks. *PloS one*, 12(2):e0170579.

Iba, 1999. Iba, Y. (1999). The nishimori line and bayesian statistics. *Journal of Physics A: Mathematical and General*, 32(21):3875.

Ibragimov et al., 2011. Ibragimov, R., Jaffee, D., and Walden, J. (2011). Diversification disasters. *Journal of financial economics*, 99(2):333–348.

Iori et al., 2008. Iori, G., De Masi, G., Precup, O. V., Gabbi, G., and Caldarelli, G. (2008). A network analysis of the italian overnight money market. *Journal of Economic Dynamics and Control*, 32(1):259–278.

Iori et al., 2015. Iori, G., Mantegna, R. N., Marotta, L., Micciche, S., Porter, J., and Tumminello, M. (2015). Networked relationships in the e-mid interbank market: A trading model with memory. *Journal of Economic Dynamics and Control*, 50:98–116.

Jacobs and Lewis, 1978a. Jacobs, P. A. and Lewis, P. A. (1978a). Discrete time series generated by mixtures ii: asymptotic properties. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(2):222–228.

Jacobs and Lewis, 1978b. Jacobs, P. A. and Lewis, P. A. (1978b). Discrete time series generated by mixtures. iii. autoregressive processes (dar (p)). Technical report, NAVAL POSTGRADUATE SCHOOL MONTEREY CALIF.

Janson et al., 2004. Janson, S., Mossel, E., et al. (2004). Robust reconstruction on trees is determined by the second eigenvalue. *The Annals of Probability*, 32(3B):2630–2649.

Jung et al., 2018. Jung, H., Lee, J.-G., Lee, N., and Kim, S.-H. (2018). Comparison of fitness and popularity: fitness-popularity dynamic network model. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(12):123403.

Kao et al., 2018. Kao, E. K., Smith, S. T., and Airoldi, E. M. (2018). Hybrid mixed-membership blockmodel for inference on realistic network interactions. *IEEE Transactions on Network Science and Engineering*.

Karrer and Newman, 2011. Karrer, B. and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical review E*, 83(1):016107.

Kim et al., 2018. Kim, B., Lee, K. H., Xue, L., Niu, X., et al. (2018). A review of dynamic network models with latent variables. *Statistics Surveys*, 12:105–135.

Kong et al., 2008. Kong, J. S., Sarshar, N., and Roychowdhury, V. P. (2008). Experience versus talent shapes the structure of the web. *Proceedings of the National Academy of Sciences*, 105(37):13724–13729.

Krivitsky and Handcock, 2014. Krivitsky, P. N. and Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):29–46.

Krivitsky et al., 2009. Krivitsky, P. N., Handcock, M. S., Raftery, A. E., and Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social networks*, 31(3):204–213.

Krzakala et al., 2013. Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L., and Zhang, P. (2013). Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940.

Krzakala and Zdeborová, 2009. Krzakala, F. and Zdeborová, L. (2009). Hiding quiet solutions in random constraint satisfaction problems. *Physical review letters*, 102(23):238701.

Kuehn, 2011. Kuehn, C. (2011). A mathematical framework for critical transitions: Bifurcations, fast–slow systems and stochastic dynamics. *Physica D: Nonlinear Phenomena*, 240(12):1020–1035.

Larremore et al., 2014. Larremore, D. B., Clauset, A., and Jacobs, A. Z. (2014). Efficiently inferring community structure in bipartite networks. *Physical Review E*, 90(1):012805.

Latora et al., 2017. Latora, V., Nicosia, V., and Russo, G. (2017). *Complex networks: principles, methods and applications*. Cambridge University Press.

Lee et al., 2017. Lee, J., Li, G., and Wilson, J. D. (2017). Varying-coefficient models for dynamic networks. *arXiv preprint arXiv:1702.03632*.

Liben-Nowell and Kleinberg, 2007. Liben-Nowell, D. and Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American society for information science and technology*, 58(7):1019–1031.

Little and Rubin, 1983. Little, R. J. and Rubin, D. B. (1983). On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *The American Statistician*, 37(3):218–220.

Longerstaey and Spencer, 1996. Longerstaey, J. and Spencer, M. (1996). Risk metrics ? technical document. *Morgan Guaranty Trust Company of New York: New York*, 51:54.

Luo et al., 2017. Luo, J., Qin, H., Yan, T., and Zeyneb, L. (2017). A note on asymptotic distributions in directed exponential random graph models with bi-degree sequences. *Communications in Statistics-Theory and Methods*, 46(18):8852–8864.

Malliaros and Vazirgiannis, 2013. Malliaros, F. D. and Vazirgiannis, M. (2013). Clustering and community detection in directed networks: A survey. *Physics Reports*, 533(4):95–142.

Marsili et al., 2009. Marsili, M., Raffaelli, G., and Ponsot, B. (2009). Dynamic instability in generic model of multi-assets markets. *Journal of Economic Dynamics and Control*, 33(5):1170–1181.

Masciandaro and Passarelli, 2013. Masciandaro, D. and Passarelli, F. (2013). Financial systemic risk: Taxation or regulation? *Journal of Banking & Finance*, 37(2):587–596.

Matheson, 2012. Matheson, T. (2012). Security transaction taxes: issues and evidence. *International Tax and Public Finance*, 19(6):884–912.

Matias and Miele, 2017. Matias, C. and Miele, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1119–1141.

May, 1976. May, R. M. (1976). Simple mathematical models with very complicated dynamics. volume 261, pages 459–467. Nature Publishing Group.

Mazzarisi et al., 2019a. Mazzarisi, P., Barucca, P., Lillo, F., and Tantari, D. (2019a). A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market. *European Journal of Operational Research (in press)*.

Mazzarisi and Lillo, 2017. Mazzarisi, P. and Lillo, F. (2017). Methods for reconstructing interbank networks from limited information: A comparison. In *Econophysics and Sociophysics: Recent Progress and Future Directions*, pages 201–215. Springer.

Mazzarisi et al., 2019b. Mazzarisi, P., Lillo, F., and Marmi, S. (2019b). When panic makes you blind: A chaotic route to systemic risk. *Journal of Economic Dynamics and Control*, 100:176–199.

Mendoza and Terrones, 2008. Mendoza, E. G. and Terrones, M. E. (2008). An anatomy of credit booms: evidence from macro aggregates and micro data. Technical report, National Bureau of Economic Research.

Mézard and Parisi, 2001. Mézard, M. and Parisi, G. (2001). The bethe lattice spin glass revisited. *The European Physical Journal B-Condensed Matter and Complex Systems*, 20(2):217–233.

Milgram, 1967. Milgram, S. (1967). The small world problem. *Psychology today*, 2(1):60–67.

Mistrulli, 2011. Mistrulli, P. E. (2011). Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns. *Journal of Banking & Finance*, 35(5):1114–1127.

Moreno, 1934. Moreno, J. L. (1934). Who shall survive?: A new approach to the problem of human interrelations.

Mossel et al., 2014. Mossel, E., Neeman, J., and Sly, A. (2014). Belief propagation, robust reconstruction and optimal recovery of block models. In *Conference on Learning Theory*, pages 356–370.

Mossel et al., 2015a. Mossel, E., Neeman, J., and Sly, A. (2015a). Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461.

Mossel et al., 2015b. Mossel, E., Neeman, J., and Sly, A. (2015b). Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461.

Mossel et al., 2018. Mossel, E., Neeman, J., and Sly, A. (2018). A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708.

Mucha et al., 2010. Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., and Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *science*, 328(5980):876–878.

Musmeci et al., 2013. Musmeci, N., Battiston, S., Caldarelli, G., Puliga, M., and Gabrielli, A. (2013). Bootstrapping topological properties and systemic risk of complex networks using the fitness model. *Journal of Statistical Physics*, 151(3-4):720–734.

Nadakuditi and Newman, 2012. Nadakuditi, R. R. and Newman, M. E. (2012). Graph spectra and the detectability of community structure in networks. *Physical review letters*, 108(18):188701.

Nerlove, 1958. Nerlove, M. (1958). Adaptive expectations and cobweb phenomena. *The Quarterly Journal of Economics*, 72(2):227–240.

Newman, 2010. Newman, M. (2010). *Networks: an introduction*. Oxford university press.

Newman, 2001. Newman, M. E. (2001). Clustering and preferential attachment in growing networks. *Physical review E*, 64(2):025102.

Newman, 2002. Newman, M. E. (2002). Assortative mixing in networks. *Physical review letters*, 89(20):208701.

Newman, 2006. Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582.

Newman, 2016. Newman, M. E. (2016). Equivalence between modularity optimization and maximum likelihood methods for community detection. *Physical Review E*, 94(5):052315.

Newman et al., 2002. Newman, M. E., Watts, D. J., and Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(suppl 1):2566–2572.

Neyman et al., 1948. Neyman, J., Scott, E. L., et al. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

Nicholson and Snyder, 2012. Nicholson, W. and Snyder, C. M. (2012). *Microeconomic theory: Basic principles and extensions*. Nelson Education.

Nowicki and Snijders, 2001. Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American statistical association*, 96(455):1077–1087.

Park and Newman, 2004. Park, J. and Newman, M. E. (2004). Statistical mechanics of networks. *Physical Review E*, 70(6):066117.

Peel et al., 2017. Peel, L., Larremore, D. B., and Clauset, A. (2017). The ground truth about metadata and community detection in networks. *Science advances*, 3(5):e1602548.

Peixoto, 2013. Peixoto, T. P. (2013). Parsimonious module inference in large networks. *Physical review letters*, 110(14):148701.

Peixoto, 2014a. Peixoto, T. P. (2014a). Efficient monte carlo and greedy heuristic for the inference of stochastic block models. *Physical Review E*, 89(1):012804.

Peixoto, 2014b. Peixoto, T. P. (2014b). Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X*, 4(1):011047.

Peixoto, 2017. Peixoto, T. P. (2017). Nonparametric bayesian inference of the microcanonical stochastic block model. *Physical Review E*, 95(1):012317.

Peixoto and Rosvall, 2017. Peixoto, T. P. and Rosvall, M. (2017). Modelling sequences and temporal networks with dynamic community structures. *Nature communications*, 8(1):582.

Pesaran, 2015. Pesaran, M. H. (2015). *Time series and panel data econometrics*. Oxford University Press.

Pesaran et al., 2004. Pesaran, M. H., Schuermann, T., and Weiner, S. M. (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics*, 22(2):129–162.

Petrone and Latora, 2018. Petrone, D. and Latora, V. (2018). A dynamic approach merging network theory and credit risk techniques to assess systemic risk in financial networks. *Scientific reports*, 8(1):5561.

Poledna et al., 2014. Poledna, S., Thurner, S., Farmer, J. D., and Geanakoplos, J. (2014). Leverage-induced systemic risk under basle ii and other credit risk policies. *Journal of Banking & Finance*, 42:199–212.

Polson et al., 2013. Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American statistical Association*, 108(504):1339–1349.

Porter et al., 2009. Porter, M. A., Onnela, J.-P., and Mucha, P. J. (2009). Communities in networks. *Notices of the AMS*, 56(9):1082–1097.

Raffaelli and Marsili, 2006. Raffaelli, G. and Marsili, M. (2006). Dynamic instability in a phenomenological model of correlated assets. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(08):L08001.

Reichardt and Leone, 2008. Reichardt, J. and Leone, M. (2008). (un) detectable cluster structure in sparse networks. *Physical review letters*, 101(7):078701.

Richard et al., 2014. Richard, E., Gaïffas, S., and Vayatis, N. (2014). Link prediction in graphs with autoregressive features. *The Journal of Machine Learning Research*, 15(1):565–593.

Rinaldo et al., 2013. Rinaldo, A., Petrović, S., Fienberg, S. E., et al. (2013). Maximum lilkelihood estimation in the beta-model. *The Annals of Statistics*, 41(3):1085–1110.

Robins and Pattison, 2001. Robins, G. and Pattison, P. (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25(1):5–41.

Sampson, 1969. Sampson, S. F. (1969). A novitiate in a period of change: An experimental and case study of social relationships.

Sarkar and Moore, 2006. Sarkar, P. and Moore, A. W. (2006). Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems*, pages 1145–1152.

Sarkar et al., 2007. Sarkar, P., Siddiqi, S. M., and Gordon, G. J. (2007). A latent space approach to dynamic embedding of co-occurrence data. In *Artificial Intelligence and Statistics*, pages 420–427.

Schularick and Taylor, 2012. Schularick, M. and Taylor, A. M. (2012). Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. *American Economic Review*, 102(2):1029–61.

Sewell and Chen, 2015. Sewell, D. K. and Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657.

Shannon, 1948. Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.

Snijders, 2002. Snijders, T. A. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40.

Snijders and Nowicki, 1997. Snijders, T. A. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of classification*, 14(1):75–100.

Strogatz, 2001. Strogatz, S. H. (2001). Exploring complex networks. *nature*, 410(6825):268.

Tasca and Battiston, 2016. Tasca, P. and Battiston, S. (2016). Market procyclicality and systemic risk. *Quantitative Finance*, 16(8):1219–1235.

Travers and Milgram, 1977. Travers, J. and Milgram, S. (1977). An experimental study of the small world problem. In *Social Networks*, pages 179–197. Elsevier.

Tsay, 2005. Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John wiley & sons.

Tumminello et al., 2011. Tumminello, M., Micciche, S., Lillo, F., Piilo, J., and Mantegna, R. N. (2011). Statistically validated networks in bipartite complex systems. *PloS one*, 6(3):e17994.

Upper, 2011. Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability*, 7(3):111–125.

Van Duijn et al., 2004. Van Duijn, M. A., Snijders, T. A., and Zijlstra, B. J. (2004). p2: a random effects model with covariates for directed graphs. *Statistica Neerlandica*, 58(2):234–254.

Wagner, 2008. Wagner, W. (2008). The homogenization of the financial system and financial crises. *Journal of Financial Intermediation*, 17(3):330–356.

Wang and Wong, 1987. Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19.

Watts, 2002. Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771.

Watts and Strogatz, 1998. Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *nature*, 393(6684):440.

Weisbuch et al., 2000. Weisbuch, G., Kirman, A., and Herreiner, D. (2000). Market organisation and trading relationships. *The economic journal*, 110(463):411–436.

Wilinski et al., 2019. Wilinski, M., Mazzarisi, P., Tantari, D., and Lillo, F. (2019). Detectability of macroscopic structures in directed asymmetric stochastic block model. *Phys. Rev. E*, 99:042310.

Williams et al., 2019. Williams, O. E., Lillo, F., and Latora, V. (2019). Effects of memory on spreading processes in non-markovian temporal networks. *New Journal of Physics*.

Wolf et al., 1985. Wolf, A., Swift, J. B., Swinney, H. L., and Vastano, J. A. (1985). Determining lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, 16(3):285–317.

Wolff, 1950. Wolff, K. H. (1950). *The Sociology of Georg Simmel*. Glencoe, Ill: Free Press.

Xing et al., 2010. Xing, E. P., Fu, W., Song, L., et al. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, 4(2):535–566.

Xu, 2015. Xu, K. (2015). Stochastic block transition models for dynamic networks. In *Artificial Intelligence and Statistics*, pages 1079–1087.

Xu and Hero, 2014. Xu, K. S. and Hero, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):552–562.

Yan and Xu, 2013. Yan, T. and Xu, J. (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika*, 100(2):519–524.

Yang et al., 2011. Yang, T., Chi, Y., Zhu, S., Gong, Y., and Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks: a bayesian approach. *Machine learning*, 82(2):157–189.

Young et al., 2018. Young, J.-G., St-Onge, G., Desrosiers, P., and Dubé, L. J. (2018). Universality of the stochastic block model. *Physical Review E*, 98(3):032309.

Zhang et al., 2015. Zhang, X., Martin, T., and Newman, M. E. (2015). Identification of core-periphery structure in networks. *Physical Review E*, 91(3):032803.

Zhang et al., 2017. Zhang, X., Moore, C., and Newman, M. E. (2017). Random graph models for dynamic networks. *The European Physical Journal B*, 90(10):200.

Zijlstra et al., 2006. Zijlstra, B. J., Van Duijn, M. A., and Snijders, T. A. (2006). The multilevel p2 model. *Methodology*, 2(1):42–47.