SCUOLA
NORMALE
SUPERIORE

Hierarchy and risk in financial networks

by

Elisa Letizia

PhD Thesis
in
Financial Mathematics

Supervisor
Prof. Fabrizio Lillo

Datta. Dayadhvam. Damyata.

T.S. Eliot

Acknowledgements

First of all, I would like to thank my supervisor Fabrizio Lillo for his guidance and patience during these years. His suggestions, remarks as well as encouragement has been crucial for my research and my personal growth.

I also would like to thank Stefano Battiston for allowing me to be a visiting under his supervision. During my time in Zurich, beside adding value to my research, I learnt the importance of putting research at the service of the community. I am grateful to the entire Finexus group at UZH for welcoming me as full member from the very first day, and for sharing with me the very fine art of Swiss grilling.

I would like to thank the R&D team of Unicredit in Milan for technical support and useful discussions during my short visits.

I would like to thank the people who made these years easier to go by, with a gesture of kindness, or endless debates on the deepest and silliest subjects.

Finally, I am sincerely grateful to my family. Since I can remember you have been a constant support and source of inspiration.

Contents

List of publications

Published

Letizia, E., Barucca, P., and Lillo, F. (2018). Resolution of ranking hierarchies in directed networks. *PloS One*, 13(2):1–25

Working Papers

Letizia, E. and Lillo, F. (2017). Corporate payments networks and credit risk rating. *arXiv preprint arXiv:1711.07677*.
E. Letizia, & F. Lillo. Beyond the resolution limit in hierarchy detection

Other works

Co-supervisor of master thesis: Linear Threshold Models on Directed Networks, F. Di Lauro, Physics Department, University of Pisa

Introduction

Networks, and why to use them

In finance, as well as in several other fields, a faithful description of the behaviour of a system cannot disregard the role of interactions between the parts composing it, to the point the interactions become as crucial as the parts themselves. The role of interactions can be made evident in different ways. In some cases they can explain macroscopic behaviours that are not observable at a singleton level. In others, the collective behaviour can influence each part through the same source of interactions.

Bearing this in mind, one can provide very general yet mathematically rigorous models for dealing with many phenomena in a system of interacting parts using networks. This is the case, for example, of (Cantono and Solomon, 2010), where authors focus on propagation of shocks, and despite the complexity in terms of number of participants and type of interactions, the network perspective is able to capture the main drivers and to condensate the information, and it provides a good level of predictability and control. More in general, this approach is usually referred to as *network science*. A network is the combination of both the individuals and their interactions. Individuals are represented by the nodes, and examples come from many domains: people, animals, genes, banks, countries. Interactions are represented by links, and they may stand for friendship, trophic level, co-occurrence, trade agreements, contracts, and so on. In other words, network science provides a conceptual framework to describe and model complex systems (Tumminello et al., 2005). The mathematical foundations are mostly given by graph theory, whose canonical starting point is (Euler, 1736) and which developed into many streams (Bondy et al., 1976; Biggs, 1993;

Chung, 1997; Bollobás, 2013).

For its generality and flexibility this approach is suitable for modelling a wide range of systems, and it has benefited from its ubiquity being techniques developed in specific fields successfully transferred to other domains. Applications span from biology (Salwinski and Eisenberg, 2004) to social interactions (Wasserman and Faust, 1994), from finance to air traffic (Mayer and Sinai, 2003). Using network abstraction one is able to cover both empirical investigation and modeling to address issues such as

- *how much* participants interact with each other;
- how interactions evolve in time;
- how network structure influences macroscopic behaviours or quantities;
- which (microscopic) mechanisms lead to the formation the network;
- which subgroups (communities) of nodes, if any, are homogeneous with respect to some characteristic (from the network, or from external information);
- how much the whole system is resistant to attacks (targeted or random);
- which nodes, if any, have a dominant position in the network, e.g. being the intermediary for most interactions.

Networks in finance

As for financial systems, network science has proven to be a very powerful tool and its popularity is still growing. One of the reason for this success, is that in the financial world, let one been dealing with stocks (Bonanno et al., 2004), banks (Fujiwara et al., 2009), traders (Delpini et al., 2018), the singleton's action / evolution is bond to the interaction with others. For stocks interaction is made evident by correlation, synchronous, lagged, partial (Mantegna and Stanley, 1999); for banks it may represent ownership (Glattfelder and Battiston, 2009), interbank lending (Gai et al., 2011; Bargigli et al., 2015), common asset holding (Huang et al., 2013; Caccioli et al., 2014; Levy-Carciente et al., 2015). This inherently interacting dynamic is coupled with a very large availability of data: if for biological systems one may need costly experiments to collect data that describe interactions (Li et al., 2004), or collect data a posteriori (Zachary, 1977; Girvan and Newman, 2002), financial data come either from the fact since last two decades financial systems are mostly managed electronically so data are created the daily operations or because they are collected for regulatory purposes (balance sheets, ownership).

In (Allen and Babus, 2009) authors clearly outline the potential of network approach to finance, in particular its ability to highlight the externalities (both positive and negative) due to the fact the financial institutions do not live in isolation. Authors effectively identify two main streams of research for network in finance. A large part of the literature is devoted to the study of the mechanisms that lead to the creation of networks, of various kind, among financial institutions, and in which measure this is consistent with economic theories. Central issues are the role of intermediaries in many markets, and more in general how the position in the network may influence an institution's business activity. In (Iori et al., 2008) authors asses the efficiency of the interbank market by studying the network of banks from the European overnight money market. This is achieved by focussing on the evolution of the connectivity structure of the banks active in the market. (Adams et al., 2010) also focuses on direct interactions by simulating a payment systems among banks. This approach allow to asses how liquidity is allocated in presence of different protocols for transfer the payment, e.g with or without intermediaries. Different costs for the liquidity have direct consequences on the overall topology of the network and authors are able to recover real networks' topology by only controlling the stream of transactions and rule for liquidity allocation. The approach in (Billio et al., 2012) is instead econometric: several measures are applied to time-series of returns coming from hedge funds, banks, broker/dealers, and insurance companies to quantify interactions. The resulting network gives an overview of the complex and multifaceted interconnectedness of all financial participants over a decade and offer a insight on the means of propagation of shocks.

The other main driver of network studies in finance mentioned in (Allen and Babus, 2009) has been the financial crisis of 2007. In this case the focus is on the resilience of financial networks in the case of an extreme event, and on mechanisms of propagation of shocks (distress, defaults etc.) among participants on a network that is usually static. In this perspective, the speech by Bank of England Executive Director, Financial Stability (Haldane, 2013) proposes most elements developed in the subsequent literature. Author highlights how interconnectedness among financial institutions played a mayor role in the consequences of crisis: concepts from biology, epidemiology, ecology are introduced in the financial word via network abstraction to explain how a relatively small event (the collapse of Lehman Brothers) lead to a massive distress of the entire system (total liquidity freeze). Beside their importance to model the financial system, networks are also marked as a possible tool for regulators to address systemic risk. To this extend, the assumption of homogeneity and

completeness which ensures analytical tractability and common to many early network models finds a certain degree of limitation when comparing to the empirical evidence of the complex and heterogeneous structure of real financial networks. In (Cont, 2013) author presents an overview of these results, mostly by central banks, stressing the importance of knowing the specific topology to assess the stability properties of the system as a whole. Author subsequently proposes a model to quantify the systemic importance of each institution, and results confirms the key role of the heterogeneity in the network structure in determining the importance of each institution, posing the basis for more tailored regulatory requirements. Similar reasoning is developed in (Drehmann and Tarashev, 2013), where authors, in order to asses an institution's systemic importance, consider both its tendency to propagate shocks and its vulnerability. (Li et al., 2014), have the same objective of quantify the riskiness of the financial system as a whole but rather than looking at banks only they focuses on industries and countries, and proposes a quantitative method to rank their systemic importance documenting an increasing dominance of China on the worldwide economy.

This very brief overview exemplifies the wide range of problems that can be addressed in finance using networks. Despite several different definitions for both the networks and the propagation are present in the literature, two features represent the common ground: the first is that networks can effectively capture the complexity of the financial system. The second is that the knowledge of the real organisation of interaction among financial institution is crucial to develop more effective policy as complexity *per se* is a source of instability as investigated in (Haldane and May, 2011; Gai et al., 2011) and more recently in (Battiston et al., 2016).

Contribution

This thesis contributes to the network literature in two directions. The first part, which comprises the first three chapters, deals with the general problem of characterising the organisation of nodes into groups according to their connections. The structure considered here is the hierarchy, and in particular the focus is on strategies and limitations of its inference. The approach is methodological, hence our considerations can be applied in many fields, but we also present an example of application to finance in the last chapter. The second part presents the study of the collection of payments between firms and their credit risk rating

using the tools of network science. The unique dataset gives the opportunity to shed light on a system, the Italian corporate firms, on which not much is known from the literature especially from the perspective of network science and with the same granularity of details.

Hierarchies

As mentioned at the beginning of this introduction, a recurrent goal in network literature is to identify subsets of nodes, which are homogeneous with respect to a certain criterion derived from network properties. This is convenient because a coarse grouping can summarise information about the structure otherwise undetectable, especially in the case of very large networks. In this thesis, the focus will be on the hierarchical structure of directed networks. An exact hierarchical organisation in a directed network means that the set of nodes can be divided in an ordered collection of classes such that links exist only from a node of a low rank class to a node of a higher rank class. Since real networks are not necessarily exactly hierarchical, the problem considered here will be to find an *optimal* ordered partition of nodes into classes such that the structure has a maximal level of hierarchy.

It is important to stress that the concept of hierarchy just described is different from the more common definition of nested hierarchy in networks (Clauset et al., 2008) where low-level communities of nodes are nested into bigger ones, in a way directly associated with hierarchical clustering. The former case, one is able to model graphs, representing for example social organisations, as command structure or influential communities, and is defined for directed networks. The latter makes sense also for undirected networks and look for nested clusters of nodes.

The reasons to look for hierarchies in networks are several. First its presence is quite ubiquitous, from biology to economics, as many systems tend to self-arrange in stratified organisation It also has a natural interpretation providing the dominant direction of flow for resources (information, money, credit) into the system. This may be useful especially if one has in mind problems of propagation, for example of a disease in a social network or default for financial system. In the already mentioned (Haldane, 2013), hierarchy is enlisted among the structures that naturally arise in financial networks and need to be kept into consideration in the study of systemic risk.

In the literature of network theory, there are some different definitions of hierarchy, here we focus on a method which infers the structure by suitably

penalising the links against the main direction. We find, both analytically and empirically, that the chosen metric succeeds in identifying hierarchies when the structure is strong. However we prove the existence of resolution thresholds in the model parameters such that beyond these thresholds the inferred hierarchical structures are different from the planted one. To overcome the resolution limitation we propose two strategies. One is based on the iteration of the inference on each level from first run, in order to attain a finer resolution. The second strategy is based on the proof that acting on only one parameter of the penalisation function, one is able to loosen the resolution limits. This comes at the cost of solving an optimisation problem which is NP-Hard, so we propose an heuristic algorithm to tackle multi-resolution inference. Finally we propose an application to financial networks that exploiting the knowledge of the hierarchical structure could be employed as policy protocol to control financial stability.

Risk

Assessing the risk of firms is one of the fundamental activities of the credit system. Banks spend a significant amount of resources to scrutinise the balance sheet of firms in order to obtain accurate estimations of their riskiness, the internal rating, and provide credit conditions reflecting both the capability of the firm to repay the loans and its probability of default. The riskiness of a firm depends on many idiosyncratic factors (e.g. balance sheet, structure of management, etc.) as well as the industrial sector or its geographical location. However, corporate firms interact with each other on a daily basis, and the interactions can be of different kinds, including those due to the supply chain, payments, business partnerships, financial contracts, and mutual ownership. The structure of interactions is complex and multifaceted but, as we anticipated before, its knowledge is critical both for macroeconomists and for the credit and banking industry to understand the dynamics of the economy, the business cycle, the structure of corporate control, and, of course, the risk of firms.

Taking the network perspective on the problem of risk assessment has a double aim. On the one side, we want to shift the attention from the single *isolated* firm to a *system* of interacting parts. This is interesting because a high number of firms are involved and it may be not trivial to capture the complexity of the interplay between a single or a group of firms by looking at one firm at a time. Despite the aforementioned interest in this sense in several financial domains, this appear not to be so common when dealing with corporate firms

systems. On the other side, we want to understand whether and in which measure a firm's role in the network can be informative of its riskiness. This is important for two reasons. First, even if the risk of a firm is not known to all the counterparts, it may affect its ability to interact with other firms. For example, a poor rating (i.e. high riskiness) may prevent the access to credit and as a result it may cause a reduction or delay in payments toward suppliers. If the supplier has high risk, the missing or delayed payment can prevent its own payments, increasing the likelihood of a cascade of missing payments and a propagation of financial distress. The second reason is that, in certain cases, the knowledge of the riskiness of a firm or a group of firms is lacking or imprecise. In these cases, the existence of a correlation between network properties and risk can allow or improve the assessment of risk.

Thesis organisation

The thesis is organised as follows

- **Part I** focuses oh hierarchical structure in networks
  - In **Chapter 1** we consider the problem of resolution limit of hierarchy by means of the minimisation of a score function, termed agony. This function penalises the links violating the hierarchy in a way depending on the strength of the violation. To investigate the resolution of ranking hierarchies we introduce an ensemble of random graphs, the Ranked Stochastic Block Model. We find that agony may fail to identify hierarchies when the structure is not strong enough and the size of the classes is small with respect to the whole network. We analytically characterise the resolution threshold and we show that an iterated version of agony can partly overcome this resolution limit.
  - In **Chapter 2** we extend the investigation of the resolution limit to a wider family of metrics. We analytically characterise the resolution threshold and we find that is less strict that the one presented in the previous chapter. This motivates the introduction of an heuristic for inferring the hierarchy also when the optimisation problem is NP-Hard. We describe an heuristic algorithm and test its performance on synthetic networks.

  - In **Chapter 3** we propose an application of hierarchies in network to control the financial stability of an interbank network. The proposed strategy benefits from fast implementation and relatively limited data disclosure.

- **Part II** focuses on the application of network theory and methods to a real dataset.

  - In **Chapter 4** we provide empirical evidences that corporate firms risk assessment could benefit from taking quantitatively into account the network of interactions among firms. We consider the interactions by investigating a large proprietary dataset of payments among Italian firms. We first characterise the topological properties of the payment networks, and then we focus our attention on the relation between the network and the risk of firms. Our main finding is to document the existence of an homophily of risk, i.e. the tendency of firms with similar risk profile to be statistically more connected among themselves. This effect is observed when considering both pairs of firms and communities or hierarchies identified in the network. We leverage this knowledge to predict the missing rating of a firm using only network properties of a node by means of machine learning methods.

The four chapters present the original contribution of this thesis. The relevant literature is presented at the beginning of each chapter.

Part I

Hierarchies

Part of the content of this Chapter was published in Letizia, E., Barucca, P., and Lillo, F. (2018). Resolution of ranking hierarchies in directed networks. PloS One, *13(2):1–25.*

1

Resolution of ranking hierarchies in directed networks

We consider the problem of the inference of hierarchies in directed networks via a class of metrics recently introduced and termed *agony*. Given a ranking of nodes into classes (i.e. an ordered partition), the metric penalises links against the ranking, i.e. from a high rank to a low rank node. The hierarchy of the graph is obtained by minimising the total cost, which depends on the functional form of the penalisation.

Specifically, we focus on the resolution limits when detecting ranking hierarchies with this strategy. We introduce a class of random graphs, termed *Ranked Stochastic Block Models* (RSBM) with a tunable hierarchical structure and we study the resolution limit of hierarchy detection with agony on RSBM. We find, both analytically and empirically, that agony succeeds in identifying hierarchies when the structure is strong. However we prove the existence of resolution thresholds in the model parameters such that beyond these thresholds agony minimisation identifies hierarchical structures which are different from the planted one. Using symmetry arguments we explore analytically alternative rankings, showing that they can have higher hierarchy than the planted one.

The chapter is organised as follows. In Section 1.1 we present relevant literature regarding hierarchical structure of networks and its inference. In Section 1.2 we introduce the cost functions for agony, in Section 1.3 we define the model for RSBM and we compute an estimate for the value of agonies of graphs in the

ensemble, in Section 1.4 we study the resolution limit for this class of graphs, and in Section 1.5 we present some numerical simulations which support the analytical computations. In Section 1.6 we present two strategies to improve the detection of ranking in real networks. Finally in Section 1.7 we draw some conclusions.

1.1 Literature review

The relevance of identifying ranking hierarchy is not new in the context of social network analysis (Wellman, 1983; Simon, 1991; Salthe, 2010) and its applications have been many so far. In the context of ecosystems it has been used to describe food webs, as for example in (Johnson et al., 2014; Johnson and Jones, 2017), where it was shown how species exhibit a property of trophic coherence, measuring how consistently a species falls into a distinct level of hierarchy within a food web. In social network analysis, examples of applications are (Shetty and Adibi, 2005; Nguyen and Zheng, 2014), where the importance of individuals into a corporate organisation and online community, respectively, are found by looking at the direction on the connections, assuming that the most important individuals in a networks are those to whom the other tend (or want) to connect. Finally, another field where hierarchies have been found to be particularity suitable is finance, both to study the funds flow of portfolios (Fama and French, 2002; Frank and Goyal, 2003) and in modelling the corporate cross-ownerships in economics (Glattfelder and Battiston, 2009).

Despite the intuition of a hierarchy is quite well established, there is no real consensus on its definition, especially with the formalism of network theory. A more systematic effort to characterise hierarchy in network is presented in (Corominas-Murtra et al., 2013), but is focused on undirected network. On one side, a common definition is related to *nestedness*: low-level communities of nodes are nested into bigger ones, in a way directly associated with hierarchical clustering. This approach has been developed in (Clauset et al., 2007, 2008; Newman, 2012; Peixoto, 2014b), where beside methodologies for the inference, also techniques for network reconstruction from partial data based on the knowledge of the hierarchical structure are presented. Another definition is the one used in (Langville and Meyer, 2012; Cucuringu, 2016) and related literature, where the ranking to be inferred is a complete order on a undirected network, not an ordered partition. Finally, another stream of literature deals with a definition of hierarchy closer to the one employed in this thesis,

introduced in (Simon, 1991; Krackhardt, 1994). Here hierarchy is defined in directed networks and look for ordered partitions (rankings) of nodes such that there is a primary direction of connections (and hence of the flow of information/resources). Starting from this definition, in (Maiya and Berger-Wolf, 2009) a method for inference based on maximum likelihood estimation for hierarchies in directed and weighted network is proposed and tested on real network. In (Romei et al., 2015) the inference method is instead based on the more general concept of bow-tie structure in networks (Vitali et al., 2011), and has the pitfall of not being able to deal with the strongly connected component of the graph (which is a feature that many real networks share). In (Trusina et al., 2004; Nepusz and Vicsek, 2013) the rank of a node is deduced from its degree or assigned and generative processes for hierarchical network are proposed. The strategy proposed in (Tibély et al., 2013) is specifically designed for the problem of tagging, however a general metric to measure the similarity of hierarchies is proposed, which is based on the similarity of the induced DAGs, i.e the directed acyclic graph obtained once links opposite to the hierarchy are removed. Finally in (Mones, 2013) after acknowledging three different hierarchy connotations, authors introduce a metric for the one they called flow hierarchy (which is equivalent to the last definition) based on the concept of reachable set (i.e the set of nodes to which there is a path from a given node). More recently, in (Coscia, 2018) author proposes a score for hierarchy, called in this case arborescence. From the original graph a DAG is obtained in two steps. First, each strongly connect component is substituted by a single node representing all the nodes in the component, secondly some links are further eliminated to avoid undirected cycles. The score is the ratio of number the edges in the obtained DAG and the original network.

The inference of hierarchy by minimising a specific penalisation function can be embedded into the more explored problem of communities detection in graphs (Fortunato, 2010; Botta and del Genio, 2016). An extensively used and investigated metric in that stream of literature is modularity, which is also known to have resolution limits (Kumpula et al., 2007; Fortunato and Barthelemy, 2007). The specific definition and the metric *agony* used here to measure the hierarchy of the graph has been introduced in (Gupte et al., 2011), while in (Tatti, 2017) the algorithm for solving the optimisation problem in a special case and the computational complexity is presented. Here, to investigate the possible resolution limits we proceed in a way similar to what has been done for community detection in (Bickel and Chen, 2009; Zhang and Moore, 2014; Newman, 2016).

1.2 Hierarchical structure of a network

Let $G = (V, E)$ be a binary directed graph of $N \equiv |V|$ nodes and $m \equiv |E|$ links. A rank function $r : V \to \{1, ..., R\}$ associates each node to an integer number which indicates the position of the subset (or class) containing the node in the hierarchy. Thus a rank function generates an ordered partition of the nodes into R subsets $\mathcal{C}_i$ ($i \in \{1, ..., R\}$) of size $n_i = |\mathcal{C}_i|$. From this point, we will refer to the ordered partition induced by the rank function with the term ranking. Once a ranking has been assigned to the graph G, a link between two nodes is classified as *forward* if it goes from a node in a class to one in a class with a strictly higher rank and *backward* otherwise. Identifying the optimal hierarchical structure in a directed graph means to find a ranking where the presence of backward links is suitably penalised. The penalisation will in general depend on the number of backward links as well as on the distance in rank between the connected nodes. The penalisation is of course arbitrary and it is interesting to investigate the ability of different forms of penalisation in identifying hierarchies.

The concept of agony in graphs was first introduced in (Gupte et al., 2011) and it is the weighted cost of all the backward links in a ranking. More specifically, given a graph G and a ranking r the value of agony with respect to r is given by:

$$A_f(G, r) = \sum_{(u,v)\in E} f(r(u) - r(v)), \tag{1.1}$$

where f is a penalty function such that it is zero for negative argument and non decreasing otherwise. We will consider here f of the form

$$f_d(x) = \begin{cases} (x+1)^d & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad d \geq 0, \tag{1.2}$$

and we will denote the value of agony of the ranking r on graph G with $A_d(G, r)$. The agony of the graph is defined as the minimum value of agony with respect to all possible rankings on the nodes, i.e.

$$A_d^*(G) = \min_{r\in\mathcal{R}} A_d(G, r), \tag{1.3}$$

where $\mathcal{R}$ denotes the set of all rankings. Fig 1.1 shows two examples of optimal rankings for simple graphs and illustrates the difference between backward and forward links.

Figure 1.1: Optimal rank and agony ($d = 1$) for simple graphs. On the top, the graphs are represented without any ordering, on the bottom nodes are divided according to their ranks. The red links are the backward ones, those contributing to agony, and the black links are the forward ones.

Remark.

1. When the graph is a Directed Acyclical Graph (DAG), one can always find a ranking of the nodes such that there are no backward links (see (Newman, 2014) for a simple routine to solve this problem), hence the value of agony of a DAG is 0, and we say the graph has a perfect hierarchy.

2. The presence of $+1$ in the cost function f makes sure that same same class links are penalised. Without this, the optimal partition would always be the trivial one, with all the nodes in the same class.

3. Thanks to the minimisation, for the value of (generalised) agony for the optimal partition it holds $0 \leq A_d^* \leq m$. For the trivial partition r^T, i.e. the one with all the nodes in the same class, it holds

$$A_d(G, r^T) = \sum_{(u,v)\in E} (0+1)^d = m$$

which gives the indicated upper bound for the optimum.

4. The exponent d acts as a tuning parameter: when it increases, only rankings with stronger hierarchies are privileged over the trivial one.

5. The optimal ranking may be not unique, however there exists a routine to choose the ranking with the smallest number of classes among those with the optimal value of agony (see (Tatti, 2017) for more details).

Finally, one can define the *hierarchy* of a directed graph as

$$h_d^*(G) = 1 - \frac{A_d^*(G)}{m}. \tag{1.4}$$

From the previous remark (ii) it follows that $0 \leq h_d^* \leq 1$ where $h_d^* = 1$ indicates a perfect hierarchy.

Once the penalisation has been chosen, the problem of finding the optimal ranking is quite complex. In its original version, agony was defined with the

piecewise linear cost function, i.e $d = 1$ in our notation. With this choice few exact algorithms to identify the optimal ranking of a graph are known (Gupte et al., 2011; Tatti, 2017). Ref. (Tatti, 2017) considered the computational complexity of algorithms for generic d. The case $d = 1$ is proven to be solved by an algorithm of polynomial complexity, while the case $d = 0$ can be reformulated into the minimum Feedback Arc Set problem (FAS, or equivalently into the dual problem: Maximum Acyclic Subgraph)(Slater, 1961) which is known to be NP-hard, but for which some heuristics exist (Eades et al., 1993). The intermediate cases, $0 < d < 1$, have concave cost functions, which also lead to a NP-hard problem according to (Tatti, 2017). The case $d > 1$, instead, have a convex cost function which gives a problem of polynomial complexity. However, to the best of our knowledge, no algorithm is available at the moment for these latter cases. One of the objectives of this paper is to investigate how the detected optimal ranking depends on the choice of the penalty function. For this reason we need to introduce a class of graphs which have a hierarchical structure and whose strength can be tuned by a suitable choice of parameters. This is what we do in the next Subsection.

1.3 Ranked Stochastic Block Model

Our ensemble of graphs belongs to the class of Stochastic Block Models (SBMs) (Holland et al., 1983). In this ensemble of graphs, nodes are partitioned into R disjointed subsets and the probability of having a link between two nodes depends on the classes they belong to and it is independent of all the other pairs of nodes, i.e.

$$\mathbb{P}[(u,v) \in E \,|\, u \in \mathcal{C}_i,\, v \in \mathcal{C}_j] = c_{ij}\,.$$

The $R \times R$ matrix $C = \{c_{ij}\}_{ij}$ is called the ***affinity matrix***. For our purpose we consider the directed version of SBMs, and C is not symmetric.

While the properties of SBM are easily obtained, due to the link independence assumption, the inference of the model's parameters which best fit a given empirical network is a subtle issue. Two approach are currently available in the literature. One one side, there is the method proposed in Peixoto (2014a), which is a Markov Chain Monte Carlo (MCMC) heuristic, which simultaneously infers the number of blocks, their composition, and the linking probabilities. Note that we consider the inference of a directed network. This method will be the one chosen later for inference in Sections 1.5 and 4.3.3. Another strategy

is the one proposed in (Decelle et al., 2011b) and based on belief propagation and cavity methods of statistical physics. Besides proving their methods to be asymptotically exact, authors also highlight a phase transition, such that some choices for the parameters generate blocks that are undetectable.

We choose a parametrisation of C in order to keep the number of parameters small, which allows to have both analytical tractability, and enough flexibility to model different types of hierarchies.

The ranking $r^{(p)}$, which we will refer to as *planted ranking*, is defined so that it is consistent with the labelling in the affinity matrix, i.e.

$$r^{(p)}(\mathcal{C}_i) = i \quad i = 1, \dots, R.$$

The term *ranking* is used here as synonym of ordered partition, which implies that in general the order is not complete

Note that, given the collection of subsets of nodes, any rank function with a range of values larger than $R-1$ would have a larger value of A_d.

Consider

$$\begin{aligned} p &= \mathbb{P}(\text{forward link towards a node in the nearest upper class}), \\ q &= \mathbb{P}(\text{forward link towards more distant classes}), \\ s &= \mathbb{P}(\text{backward link}). \end{aligned}$$

This gives the affinity matrix

$$C = \begin{bmatrix} s & p & & & & \\ & \ddots & \ddots & & q & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & s & & & \ddots & p \\ & & & & & s \end{bmatrix}$$

Figure 1.2: Schematic representation of the structure of a RSBM graph

In order to have a true hierarchical structure we require that the parameters p, q, s are such that

$$\mathbb{E}[\#\{\text{backward links}\}] \leq \mathbb{E}[\#\{\text{forward links}\}]. \tag{1.5}$$

Define $\forall k \in 1,..,R$,

$$b_k = \sum_{i=1}^{R-k} n_i n_{i+k}\,.$$

For any pair (i,j) $(i,j = 1,....R)$, the number of links between subset i and j $m_{i,j}$ follows a binomial distribution, $m_{i,j} \sim \text{Binom}(n_i n_j, (C)_{i,j})$, therefore the constraint (1.5) is equivalent to

$$s\sum_{k=0}^{R-1} b_k \leq p b_1 + q\sum_{k=2}^{R-1} b_k\,.$$

In the case of uniform cardinality of the subsets, $n_i = n\,\forall\, i$, which we will consider in the following, the inequality further simplifies to

$$s \leq s_{\max} := \frac{2(R-1)}{R(R+1)}p + \frac{(R-2)(R-1)}{R(R+1)}q \tag{1.6}$$

A SBM having the above structure and satisfying the constraint (1.5) will be termed Ranked Stochastic Block Models RSBM$(p,q,s,R,\{n_i\})$. In the case of uniform cardinality, we denote briefly RSBM(p,q,s,R).

Remark.

1. A possible interpretation of the RSBM is that p,q give the backbone of the hierarchical structure, while s represents the noise.

2. As mentioned in Remarks 1.2, the definition of the cost function implies that links between nodes of the same rank have a positive cost. This means that those links are classified as *backward*, and for this reason they are assigned a probability s as the other backward links in the model.

Since RSBMs are random graphs, different realisations of the model give different values of agony and hierarchy. We will compute below the expected value of these quantities.

We estimate the expected value of $h_d(G, r^{(p)})$, the hierarchy of the planted ranking of RSBM graphs. Note that we make a little abuse of notation indicating with h_d the value $1 - A_d/m$, i.e. we do not consider the minimisation of agony. For this reason h is not necessarily bounded between 0 and 1 as h^*_d.

Indicating with $\bar{h}^{(p)}_d$ the ensemble average of $h_d(G, r^{(p)})$, we obtain

$$\bar{h}^{(p)}_d = \mathbb{E}\left[1 - \frac{1}{m}A_d(G, r^{(p)})\right] = 1 - \sum_{i\geq j}(i-j+1)^d e_{ij}\,,$$

where $e_{ij} = \mathbb{E}\left[\frac{m_{ij}}{m}\right]$. In order to have closed form expressions we need to estimate the terms e_{ij}. We consider a second order Taylor expansion:

$$\mathbb{E}\left[\frac{m_{ij}}{m}\right] \approx \frac{\mathbb{E}[m_{ij}]}{\mathbb{E}[m]} - \frac{\mathrm{cov}(m_{ij}, m)}{\mathbb{E}[m]^2} + \frac{\mathrm{var}(m)\mathbb{E}[m_{ij}]}{\mathbb{E}[m]^3}. \tag{1.7}$$

If we assume that $n_i = O(N)\,\forall i$, then the last two terms in Eq. (1.7) vanish when $N \to \infty$, hence

$$e_{ij} \to \frac{\mathbb{E}[m_{ij}]}{\mathbb{E}[m]} \quad \text{as} \quad N \to \infty.$$

This gives the first order estimate for $\bar{h}_d^{(p)}$

$$\begin{aligned}\bar{h}_d^{(p)} &= 1 - \frac{\mathbb{E}[A_d(G, r^{(p)})]}{\mathbb{E}[m]} + o(N^{-1}) \\ &= 1 - \frac{s\sum_{k=0}^{R-1}(k+1)^d b_k}{pb_1 + q\sum_{k=2}^{R-1} b_k + s\sum_{k=0}^{R-1} b_k} + o(N^{-1}).\end{aligned}$$

It is possible to compute higher order estimates or estimates based on exact expected values. The expressions are however less transparent and we find in simulations that first order estimates are quite accurate, thus in the following we use them.

1.4 Looking for optimal hierarchies in RSBM

RSBMs are constructed with a specific ranking, the planted one, which is determined by the choice of the classes and the model parameters. When minimising a generalised agony A_d on realisations of such graphs, it is not *a priori* obvious that the optimal ranking is the planted one. We therefore ask the following question:

Given a RSBM$(p, q, s, R, \{n_i\})$, find the ranking r which minimises the generalised agony A_d. In particular check when the planted ranking $r^{(p)}$ is optimal.

This is in general a complicated problem and we do not have a complete answer to this question, despite the fact that it is possible, at least for $d = 1$, to find numerically the optimal ranking of a specific realisation of a RSBM. In

order to simplify the problem, in this paper we will restrict our attention to the homogeneous case $n_i = N/R$, $\forall i$. Given the form of the affinity matrix and the homogeneity assumption, we expect that the optimal solution, when different from the planted one, preserves the homogeneity of the planted ranking. Possible boundary effects (for example the first and last class have different size from the other ones) are not considered and we expect to play a minor role when the number of planted classes is large. In any case in the subsection 1.5 we use numerical simulations to test our intuition.

For this reason we shall compute the generalised agony of the following alternative rankings:

1. the number of classes changes either by merging adjacent classes or by splitting each class; due to homogeneity, merged or split classes have all the same size;

2. the rank is inverted, $r_j^{(i)} = r_{R-j+1}^{(p)}$, $\forall j = 1,..,R$, i.e. nodes in highest ranks of the planted ranking are given lowest ranks in the alternative. Moreover we consider also the case when the number of classes is arbitrary, but again their size is assumed to be uniform.

To distinguish between the two families of ranking, we will denote the former as *direct*, in contrast with *inverted* for the latter. For each of these alternative rankings we compute the value of $\bar{h}_d$ as a function of the number of classes and we look for the optimal one among these alternatives and the planted ranking. Clearly there is no guarantee that this will be the global optimum over all the possible rankings. To maintain this distinction, we will denote *optimal* the ranking with highest value of $\bar{h}_d$ within the subset of alternatives just described, while we will always refer to the best among all the rankings, i.e. that which gives h_d^*, as the *global optimum*. We will see for example that numerical simulations of some RSBM indicate that the globally optimal ranking is a *partial* inversion of the planted hierarchy. However this analysis serves to show that planted ranking might not be globally optimal for some generalised agony and to provide an upper bound for the resolution threshold as well as getting intuition on the characteristics of the optimal ranking in a RSBM.

In the following we will focus on two regions of the parameter space of RSBMs:

- $p \geq q > s$, termed a twitter-like hierarchy;
- $q = 0$, $p \neq 0$ termed a military-like hierarchy.

In the former hierarchy forward links can connect low rank nodes with nodes of any higher rank, while in the latter the forward links can connect a node only with nodes in a direct superior class. In both cases backward links can exists with a probability s. As we will see the global optimal ranking of the two hierarchies is quite different.

Finally we consider the case

$$R = 2^a,\ \tilde{R} = 2^{a-b},$$

where $\tilde{R}$ is the number of classes after splitting ($b < 0$) or merging ($b > 0$). The parameters $a > 1$ and $b < a$ are such that $2^a, 2^{a-b} \in \mathbb{N}$. We denote the direct and inverted rankings with 2^{a-b} classes as $r^{(b)}$ and $r^{(i,b)}$ respectively.

We will focus our attention on the case $d = 1$, $d = 0$, and $d = 2$. Results for other values of d are left for a future paper.

1.4.1 Agony with $d = 1$

In this case exact algorithms for its optimisation are known, allowing the comparison of calculations with numerical simulations.

Provided that the constraints in (1.6) are satisfied, one can easily verify that $\forall\, b < 0$

$$\mathbb{E}[A_1(G, r^{(b)})] > \mathbb{E}[A_1(G, r^{(p)})]$$
$$\mathbb{E}[A_1(G, r^{(i,b)})] > \mathbb{E}[A_1(G, r^{(p)})]$$

i.e. splitting is never optimal, neither in the direct nor in the inverted ranking.

As for merging ($b > 0$), the first order estimate of $\bar{h}_1$ is given by

$$\bar{h}_1(b; p, q, s, a) = \frac{2^{-b}(2^a - 2^b)(6p + 3(-2 + 2^{a+b})q - 2^a(2^a + 2^b)s)}{3\,(2^a(2p - 3q + s) + 4^a(q + s) - 2p + 2q)}. \tag{1.8}$$

Similarly, one can write the estimate for the value of hierarchy of the inverted ranking

$$\bar{h}_1^{(i)}(b; p, q, s, a) = \frac{2^{-b}\left(2^b - 2^a\right)\left(2^{a+b}(q - 3s) + (4^a - 6)\,q + 6p\right)}{3\,(2^a(2p - 3q + s) + 4^a(q + s) - 2p + 2q)}.$$

In this notation p, q, s, a are the parameters of the RSBM, while b refers to the modified ranking $r^{(b)}$ or $r^{(i,b)}$. Moreover it is clearly $\bar{h}_1(b = 0) = \bar{h}_1^{(p)}$.

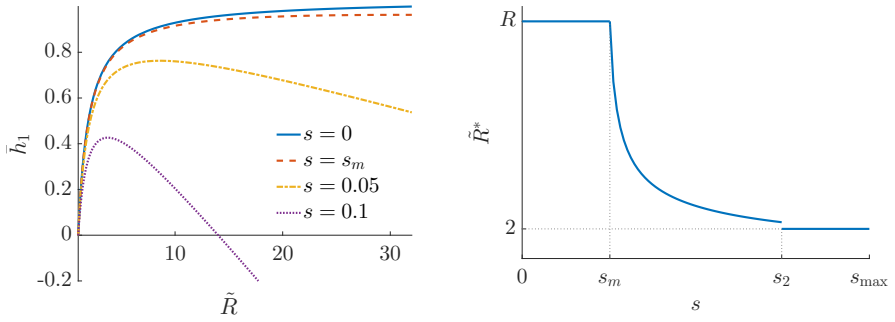


Figure 1.3: Panel (a) shows the value of the estimate of h_1 for different values of s as a function of the number of classes, $\tilde{R}$, for twitter-like graphs with parameters $p = q = 0.5$, $R = 32$. Panel (b) gives a schematic representation of the estimated optimal number of classes $\tilde{R}^*$ as s varies.

In the twitter-like hierarchy ($p \geq q > s$) it is $\bar{h}_1^{(p)} > \bar{h}_1^{(i)}(b; p, q, s, a)$, i.e. the inverted ranking is never optimal. Merging, instead, can give rankings with higher hierarchy than the planted ranking.

To show this, in the left panel of Fig 1.3 we plot the behaviour of $\bar{h}_1(b)$ as a function of the number of classes, $\tilde{R} = 2^{a-b}$, after merging. Each line is associated to a $RSBM(p, q, s, R)$. The parameters $p = q = 0.5$, $R = 32$ are fixed, while different curves refer to different values of s. We plot the variable $\tilde{R}$ as a continuous variable to help the interpretation of the observed behaviour. When s is small the maximum value of $\bar{h}_1$ is correctly identified at $\tilde{R} = R$. Above a critical value s_m of the parameter describing the probability of a backward link, the planted ranking is no longer optimal and merging classes gives a ranking with higher hierarchy. Notice that for $s > s_{\bar{h}_1^{(p)}=0}$, the hierarchy $\bar{h}_1^{(p)}$ of the planted ranking becomes negative. This might seem counterintuitive since we showed before that $h^* \in [0, 1]$. The condition $\bar{h}_1^{(p)} < 0$ simply means that putting all the nodes in the same class has a higher hierarchy than the one of the planted ranking when $s > s_{\bar{h}^{(p)}=0}$.

The right panel of Fig 1.3 shows the optimal number of classes $\tilde{R}^*$ as a function of s. As explained, when $s < s_m$ it is $\tilde{R}^* = R$, while after this value the optimal number of classes decreases and in the limit $s = s_{\max}$ it is $\tilde{R}^* = 2$. Therefore the value s_m sets a ***resolution threshold***, since twitter-like graphs with

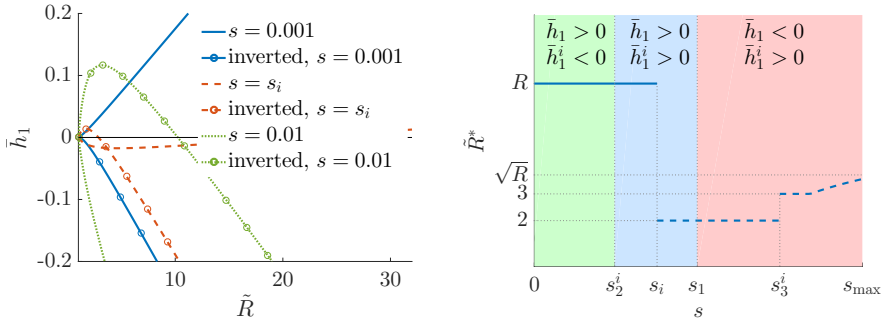


Figure 1.4: Panel (a) shows how depending on the value of s the inverted rank can give a higher value of $\bar{h}$ than the planted rank in military-like graph with parameters $p = 0.5$, $q = 0$, $R = 32$. Panel (b) gives a schematic representation of the estimated optimal number of classes $\tilde{R}^*$ as s varies, dashed lines are associated to the inverted rank.

a probability of backward links larger than s_m will not be correctly identified by agony with $d = 1$. More precisely s_m is an upper bound of the resolution threshold, since other rankings, not considered here, could have higher hierarchy than the planted and the merged ones when $s < s_m$.

Interestingly for large number of classes R, as we prove in the following Proposition 1, the resolution threshold scales as $s_m \sim (6p - 3q)/R^2$, i.e. the more communities are present the more it is difficult to detect them. The same happens for large networks ($N \to +\infty$). Taking the number of classes constant and letting p and q scale as $1/N$ to keep the connectivity fixed, one immediately sees that $s_m = O(N^{-1})$, i.e. for large networks and fixed number of classes the detectable structures are those with very strong hierarchical structure. Thus agony with $d = 1$ has strong resolution limits for large graphs, similarly to what happens with modularity and community detection.

The situation is more complex in the military-like hierarchy ($q = 0$) because for large s inverted rankings become better than direct ones. To show this, we refer to the left panel of Fig 1.4, which is the analogous of left panel of Fig 1.3. In this case, alongside $\bar{h}_1(b)$ we also plot $\bar{h}_1^{(i)}(b)$, with matching line colours to distinguish those associated to the same values of s, and circles to identify $\bar{h}_1^{(i)}$. In all cases we chose $p = 0.5$ and $R = 32$. For small values of s (solid blue lines), $\bar{h}_1$ is convex in $\tilde{R}$ and has its maximum at $\tilde{R} = R$, whereas $\bar{h}_1^{(i)}$ is negative for

inverted rankings different from the trivial one. Thus in this regime the planted ranking is optimal. When s reaches the critical value s_i (dashed red lines), the optimal choices for both the direct and inverted rankings give the same value of hierarchy. For higher s (dotted green lines) the only direct ranking with non negative hierarchy is the trivial one, i.e $\tilde{R} = 1$, while the inverted rankings are (strictly) positive for a suitable choice of b. Therefore in this regime inverted rankings outperfom the planted one.

The right panel of Fig 1.4 shows the optimal number of classes $\tilde{R}^*$ as a function of s together with an indication of the sign of the hierarchy of the optimal direct and inverted ranking. For $s < s_2^i$ the hierarchy of the optimal direct ranking is positive and the one of the optimal negative ranking is negative, for $s_2^i < s < s_1$ they are both positive, while for $s_1 < s < s_{\max}$ the inverted optimal hierarchy is positive and the optimal direct one is negative. Thus for $s < s_i$ the optimal ranking is direct and coincides with the planted one, while after this value the inverted ranking with two classes becomes optimal. This is true in the region $s_i < s < s_3^i$ after which the inverted ranking with three classes becomes optimal. By increasing s further, the optimal ranking is always inverted with an increasing number of classes up to a value smaller or equal to $\sqrt{R}$ for $s = s_{\max}$. Therefore for the military-like hierarchy the resolution threshold is s_i which for large R scales as $6p/R^2$, displaying a resolution limit similar to the twitter-like hierarchy, both for large number of classes R and for large graphs ($N \to \infty$).

We summarise the results for $d = 1$ in the following proposition.

Proposition 1. *When $d = 1$ and $p \geq q > s$,(Twitter hierarchy) the first order estimate for the optimal value of h*

$$\bar{h}_1^* = \begin{cases} \bar{h}_1^{(p)} & s \leq s_m \\ \bar{h}_1(b = b^*) & s_m < s < s_2 \\ \bar{h}_1(b = 2) & s \geq s_2\,, \end{cases} \tag{1.9}$$

where

$$s_m = \frac{6\,(2^a - 1)\,p - 3\,(2^a - 2)\,q}{2^a - 4^a + 8^a}, \quad s_2 = \frac{3}{7}\frac{(4^a - 12)\,q + 12p}{4^a},$$
$$b^* = \frac{1}{2}\log_2 \frac{2^{2a}s + 6(q - p)}{3q - s}.$$

Furthermore, when $q = 0$, (Military hierarchy)

$$\bar{h}_1^* = \begin{cases} \bar{h}_1^{(p)} & s \le s_i \\ \bar{h}_1^{(i)}(b = a-1; q = 0) & s_i < s_3^i \\ \bar{h}_1^{(i)}(b = b^{i,*}; q = 0) & s > s_3^i, \end{cases}$$

where

$$s_3^i = \frac{12}{2^{2a}} p, \quad s_i = \frac{12p}{3\ 2^a + 2^{2a+1} - 2}, \quad b^{i,*} = \frac{1}{2} log_2 \frac{2p}{s}.$$

The proof and the extended expression for $\bar{h}_1^*$ are given in Appendix 1.A.

In conclusion, we explicitly showed that for RSBMs there exist alternative rankings with a smaller agony ($d = 1$) than the planted one. The merging of the classes for the twitter hierarchy is due to fact that for a large number of classes it might be more convenient to aggregate classes paying a penalty equal to one than to leave them separate but paying a higher penalty for the distant backward links. Similarly, for the military hierarchy, when the number of backward links is relatively large, it is more convenient (in terms of agony) to invert the ranking because forward links do not enter the cost minimisation. Thus even if p is much larger than s and the number of forward links is much larger than the number of the backward links, it is more convenient to invert the ranking to avoid to pay large penalties of backward links between very distant classes.

Thus our results depend on the choice of the penalisation function and on the choice of the affinity matrix. In the next Subsection we show indeed that a very different result is obtained for $d = 0$. Changing the affinity matrix, for example introducing a probability of backward links which depends on the distance between classes, and changing the penalty function by including the negative cost of forward links is left for a future study.

1.4.2 Agony with $d = 0$

This case corresponds to the FAS problem. The optimal ranking is obtained when each node is in a different class, $\tilde{R} = N$, and the inverted ranking is never optimal as stated by the following:

Proposition 2. *When $d = 0$, $\forall\, RSBM(p, q, s, R = 2^a)$ the optimal value for the first order estimate of h is given by (for both Twitter and Military hierarchy)*

$$\bar{h}_0^* = \bar{h}_0\left(b = -\log_2 \frac{N}{R}\right) \geq \frac{1}{2}.$$

See Appendix 1.A for the proof. The reason for this result is that backward links are weighted in the same way irrespectively from the distance between the ranks of the nodes connected by the link. Thus, for example, the naive ranking with all nodes in one class has a agony equal to the number of links, while the ranking where each node is in one class has an agony equal to the number of backward links, which is smaller than the total number of links.

Finally we note that the value of $\bar{h}_0$ increases very slowly when $\tilde{R}$ approaches N, so in specific realisations of the RSBM the optimal ranking can have a number of classes smaller than N.

1.4.3 Agony with $d = 2$

Finally, we consider the case of $d = 2$. Similarly to the case $d = 1$, splitting is never optimal, both for the direct and inverted rankings, while merging can give rankings with higher value of $\bar{h}_2$ than the planted one. One can proceed as before, considering the expressions for the alternative rankings when $b > 0$:

$$\bar{h}_2(b; p, q, s, a) = -\frac{2^{-2b-1}\left(2^a - 2^b\right)\left(2^{a+2b+1}(2s - 3q) + 5s2^{2a+b} + 8^a s - 3\ 2^{b+2}(p - q)\right)}{3\left(2^a(2p - 3q + s) + 4^a(q + s) - 2p + 2q\right)},$$

and

$$\bar{h}_2^{(i)}(b; p, q, s, a) = \frac{2^{-2b-1}\left(2^b - 2^a\right)\left(2^{a+2b+1}(2q - 3s) + (5\ 4^a - 36)\,2^b q + 8^a q + 9\ 2^{b+2} p\right)}{3\left(2^a(2p - 3q + s) + 4^a(q + s) - 2p + 2q\right)}.$$

As before we describe the behaviour for the two considered hierarchies and then we state the proposition summarising our results. For the twitter-like hierarchy ($p \geq q > s$), the behaviour is similar to the $d = 1$ case. Since $\bar{h}_2^{(p)} > \bar{h}_2^{(i)}(b)$, $\forall b$, inverted rankings are never optimal. The planted ranking is optimal up to the critical value $s_{2,m}$ for the probability of backward links. After that, merged rankings outperform the planted one, and the number of classes decreases with s. When $s_{2,1} < s \leq s_{\max}$ the optimal choice is the trivial ranking, i.e. $\tilde{R} = 1, h_2 = 0$. Despite the similarity with the $d = 1$ case, the resolution threshold is now higher, since it can be shown that $s_{2,m} \leq s_m$. Moreover, while,

as noted before, in the $d = 1$ case $s_m = O(R^{-2})$, in the $d = 2$ case the resolution threshold is not only stricter but also it decreases faster as the number of classes increases, since it scales as $s_{2,m} \sim \frac{2p-q}{2R^3} = O(R^{-3})$. Finally, when $d = 2$ the large s case has the trivial ranking as the optimal one, whereas in the $d = 1$ case the optimal ranking has two classes.

For the military-like hierarchy ($q = 0$), the planted ranking is proven to be optimal with respect to the direct rankings up to the critical value $s^0_{2,1}$. After this value the optimal choice is the trivial ranking. Then when $s > s^i_{2,2}$ it becomes optimal to merge inverted rankings and the optimal number of classes increases with s, starting from $\tilde{R} = 2$. Differently from the case $d = 1$, in this case it holds $s^0_{2,1} < s^i_{2,2}$, hence for $s \in (s^0_{2,1}, s^i_{2,2})$ the optimal rank is the trivial one, and the resolution threshold is given by $s^0_{2,1}$, which scales as $12p/R^3$, while inverted rankings are to be preferred for any $s > s^i_{2,2}$.

We summarise the results for $d = 2$ in the following proposition.

Proposition 3. *When $d = 2$ and $p \geq q > s$ (Twitter hierarchy), the first order estimate for the optimal value of h*

$$\bar{h}^*_2 = \begin{cases} \bar{h}^{(p)}_2 & s \leq s_{2,m} \\ \bar{h}_2(b = b^*_2) & s_{2,m} \leq s \leq s_{2,1} \\ 0 & s > s_{2,1}\,, \end{cases}$$

where

$$s_{2,m} = \frac{6\left(2^{1-a}(q-p) + 2p - q\right)}{-3\ 2^a + 2^{3a+1} + 4^a + 4}\,, \quad s_{2,1} = \frac{2^{2a}q + 4p - 4q}{3\,2^{2a}}$$

and b^*_2 is given Appendix 1.A.

Furthermore, when $q = 0$ (Military hierarchy),

$$\bar{h}^*_2 = \begin{cases} \bar{h}^{(p)}_2 & s < s^0_{2,1} \\ 0 & s^0_{2,1} \leq s \leq s^i_{2,2} \\ \bar{h}^{(i)}_2(b = a - 1; q = 0) & s^i_{2,2} < s < s^i_{2,3} \\ \bar{h}^{(i)}_2(b = b^{i,*}_2; q = 0) & s \geq s^i_{2,3}\,, \end{cases}$$

where

$$s^0_{2,1} = \frac{3\ 2^2 p}{2^a(5\ 2^a + 4^a + 4)}\,, s^i_{2,2} = \frac{12}{2^{2a}}p\,, s^i_{2,3} = 3s^i_{2,2}\,, b^{i,*}_2 = \frac{1}{2}\log_2\left(\frac{6p}{s}\right).$$

With this last proposition we showed that hierarchy detection with quadratic cost function has a behaviour very similar to the linear case. However the resolution limits we highlighted before escalates in this case, and, as a result, only very strong hierarchies are detected correctly when the number of class is large. The same computations can be done also for greater integers d, for which the sums in the estimates of agony have a closed formula. Intuitively as $d \in \mathbb{N}$ increases, backward links to distant classes are given a larger penalisation, hence rankings with merged classes become more convenient than the planted one even for smaller values of s. In other words agonies with $d > 1$ are strongly suboptimal and are able to identify very strong structures.

Following this remark, better candidates as penalty functions are likely those with $0 < d < 1$. For at least some of those d one can expect to soften the resolution limits associated to integer d. However the approach to study the regime cannot rely on analytical formulae.

1.5 Numerical Simulations

In this Subsection we show the results of numerical simulations to test the propositions we presented before. This is important for two reasons. First, to show that the guessed rankings, obtained by merging, splitting, or inverting the planted one, are indeed the optimal ones or have a hierarchy close to the optimal one. Second, to prove that the first order approximation and other simplifying assumptions give analytic expressions close to numerical simulations.

We use *igraph* (Csardi and Nepusz, 2006) to sample a graph from the RSBM ensemble. For computing agony we use the algorithm described in (Tatti, 2017), which we will refer to it as *agony* (in italics) for brevity, and which gives the exact solution for the optimisation problem when $d = 1$. Finally, we use the MCMC algorithm in the *GraphTool* (Peixoto, 2014a) package for the inference of the SBM (without constraint on the structure of the affinity matrix).

We perform the same analysis with different choices for the parameters p, q, s, R, N and the results are consistent, thus in the following we present only representative cases. We use the adjusted Rand Index (RI) (Hubert and Arabie, 1985) to measure the similarity between the planted and the inferred ranking. The RI is 0 between independent rankings and 1 when each pair of elements that are in the same class in one ranking are also in the same class in the other. Ordering of classes does not matter in computing RI, thus the RI between a ranking and its inverted version is 1. Nevertheless we checked that high values

of RI do not correspond to inverted rankings.

1.5.1 Twitter-like hierarchy

We generate RSBM with parameters $p = 0.5$, $q = 0.5$, $R = 32$, $\frac{N}{R} = 128$, $N = 2^{12} = 4096$, and we vary the value of s. Fig 1.5 shows the heat maps of the classes found by *agony* for different values of s. The heat-maps are constructed as follow: a square in position (i, j) refers to the number of nodes that belong to class i in the planted rank and are placed in class j by *agony*: the darker the colour, the higher the number. For small s (almost DAG structures) the algorithm recovers faithfully the planted ranking and the RI is high. When the hierarchical structure becomes weaker, the ranking obtained by agony is the merging of contiguous classes in the hierarchy, as postulated in the theoretical part above. For this choice of p, q, R the resolution threshold for s is $s_m = 0.00151$ consistently with our simulations. As we predicted, classes merge more and more when s increases. The inferred rankings are close to uniform, and the main exception is the first and last class which are smaller than the other ones.

We show numerically that the ranking we proposed as optimal in the previous Section has indeed a value of hierarchy very close to the one obtained from simulations. In Fig1.6 we show a scatter plot of the true value of h_1^* computed with *agony* on the simulated graphs against the hierarchy of the planted rank $h_1^{(p)}$ (circles), and against $\bar{h}_1^*$, the hierarchy computed with Eq. (1.9)(stars). To evaluate the latter we use the coefficients of the RSBM estimated from the sample graph with *GraphTool*. We estimate $p = q$ and s as the average elements of the inferred affinity matrix on the corresponding classes and we leave free the number of classes. For $s < s_m$ (red symbols) the two methods agree and give a value of hierarchy consistent with the real one. When $s > s_m$ (green and blue symbols depending on whether s is smaller or larger of $s_{\bar{h}^{(p)}=0}$) the hierarchy of the planted ranking is significantly smaller than h_1^*, showing that another ranking is optimal. This has a value of hierarchy which is very close to the one computed from Eq. (1.9), even when the coefficients of the RSBM are estimated from data. It is interesting to note that this is true also for s very close to s_{max} where the number of classes detected by *GraphTool* is significantly smaller than R. This is due to the fact that the analytical expression in Eq. (1.9) of the value of hierarchy of the merged ranking depends weakly on the number of classes. This is a strong indication that the ranking we suggested, and obtained

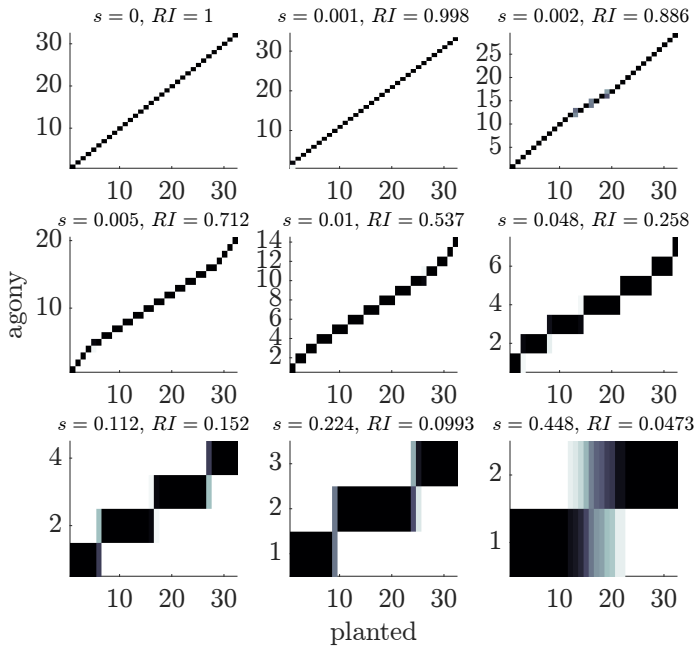


Figure 1.5: Heat maps comparing the planted ranking with the ranking inferred with *agony* for twitter-like hierarchy. In each panel a square in position (i, j) contains the number of nodes that belong to class i in the planted rank and are placed in class j by *agony*: the darker the colour, the higher the number. The parameters are $p = q = 0.5$, $R = 32$ and 9 values of s. Each plot refers to a single realisation from the ensemble.

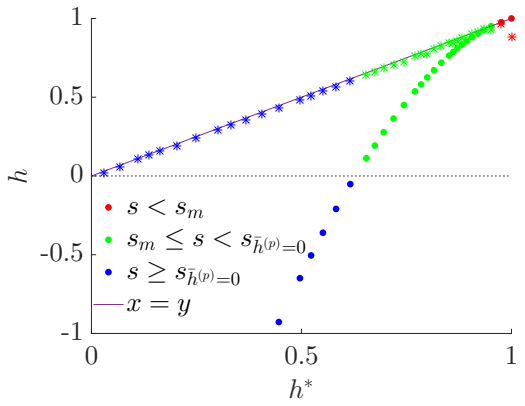


Figure 1.6: Comparison of hierarchies for twitter-like RSBMs. The parameters are $p = q = 0.5$, $R = 32$, s varies in $[0, s_{\max}]$, with $s_{\max} = 0.448$. Each point refers to a single realisation of the ensemble. The circles represent the pairs $(h_1^*, h_1^{(p)})$, i.e. the optimal hierarchy h_1^* computed with *agony* and the one of the planted hierarchy $h_1^{(p)}$. The stars represent $(h_1^*, \bar{h}_1^*)$ where $\bar{h}_1^*$ is the theoretical hierarchy of Eq. (1.9) with the parameters of the SBM estimated via *GraphTool*. Finally, s_m is the theoretical resolution threshold and $s_{\bar{h}^{(p)}=0}$ is the theoretical value of s for which the estimate for the planted hierarchy is zero.

by merging the classes, has a value of hierarchy which is indeed very close to the globally optimal one. In conclusion, the planted hierarchy is optimal for a very small range of values of s and, as we expected, it gives negative values of h_1 when s is large enough. On the other side, our estimate for optimal h_1 is accurate for all the value of s considered.

Finally in Fig 1.7 we show that the resolution problem is due to the choice of the method (agony with $d = 1$) and not necessarily to the model itself. In fact it is well known that SBM have a resolution threshold both when inference is done using Maximum Likelihood methods (Decelle et al., 2011a) and spectral methods (Nadakuditi and Newman, 2012). To this end we infer a SBM on the adjacency matrix, keeping free the number of classes (see (Peixoto, 2014a) for the model selection adopted by *GraphTool*) and we compute the RI of the planted ranking versus the one obtained with *agony* and the SBM fit. The result is shown in Fig 1.7 for different values of s. We see that the SBM fit outperforms *agony*. It is clear that, since we are using SBM for generating the graphs, its fitting will be better. However what we want to stress is that there is remarkably wide interval

of values of s for which *agony* is not able to detect a hierarchical structure even if it is strong enough to be detected by another method. Hence the limit in resolution is not embedded in the RSBM but in the objective function associated to agony.

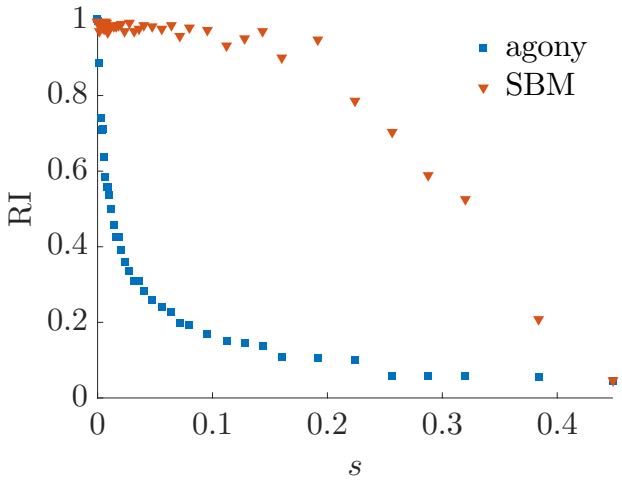


Figure 1.7: The figure shows the value of the Rand Index between the planted ranking and the inferred ones. The blue squares considers the ranking obtained with *agony* (hence $d = 1$), while the red triangles considers the ranking obtained with a RSBM fit via *GraphTool*. The parameters of the twitter-like hierarchy are $p = q = 0.5$, $R = 32$, s varies in $[0, s_{\max}]$, with $s_{\max} = 0.448$, and each point refers to a single realisation of the ensemble.

1.5.2 Military-like hierarchy

For the military-like hierarchy things are more complicated. Fig 1.8 shows the heat map of the classes for $p = 0.5$ and nine values of s. With these parameters our formulas give $s_i = 0.00280$ and $s_1 = 0.00284$. We see that for strong hierarchical structures (small s) *agony* recovers well the classes. However when s increases a *partial* inversion of the hierarchy is observed and only for large s we recover the fully inverted ranking we studied in the previous Section. Thus simulations show that the latter is not always the optimal ranking but rather there are partially inverted rankings with a larger hierarchy. The purpose of the above analysis on the military-like hierarchy is to show that there exist

values of the parameters for which the planted ranking is not optimal and to demonstrate that partial inversion can outperform the planted one. Moreover the partial inversion is observed for $s = 0.002 < s_1$, hence our computations provide a upper bound of the true resolution threshold.

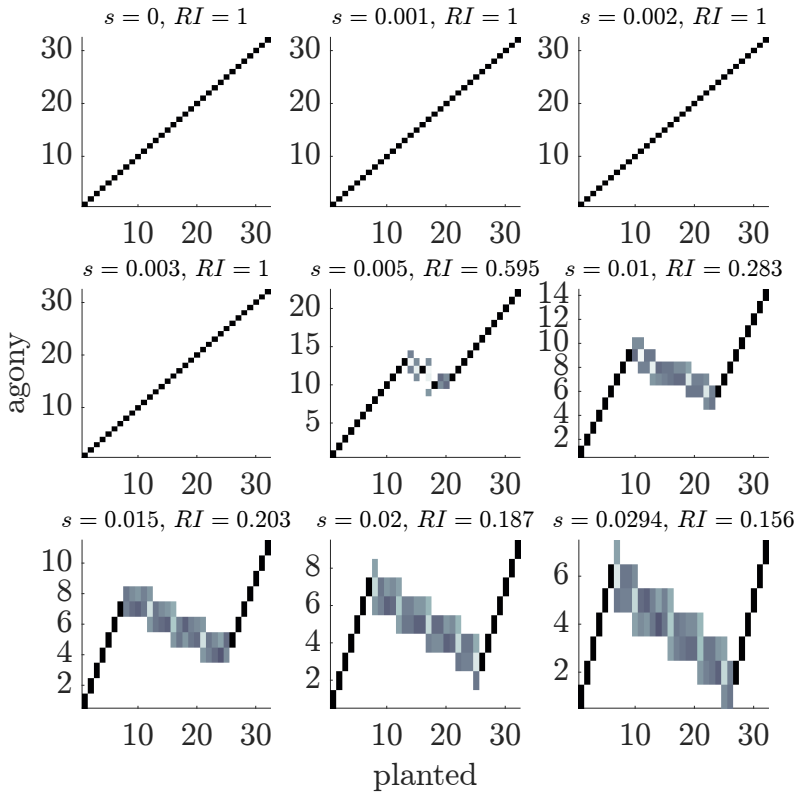


Figure 1.8: Heat maps comparing the ranking inferred using *agony* with the planted ranking for military-like hierarchy. In each panel a square in position (i, j) contains the number of nodes that belong to class i in the planted rank and are placed in class j by agony: the darker the color, the higher the number. The parameters are $p = 0.5$, $q = 0$, $R = 32$, s varies in $[0, s_{\max}]$, with $s_{\max} = 0.0294$, and each plot refers to a single realisation of the ensemble.

Fig 1.9 shows, similarly to Fig 1.6, the scatter plot of the true value of h_1^* computed via *agony* on the simulated graphs against the hierarchy of the planted rank $h_1^{(p)}$ (circles), and against $\bar{h}_1^*$, the hierarchy computed with Eq. (1.9) using the coefficients of the SBM estimated from the sample graph with *GraphTool*. The main message of the Fig is that, despite the fact the symmetrically inverted

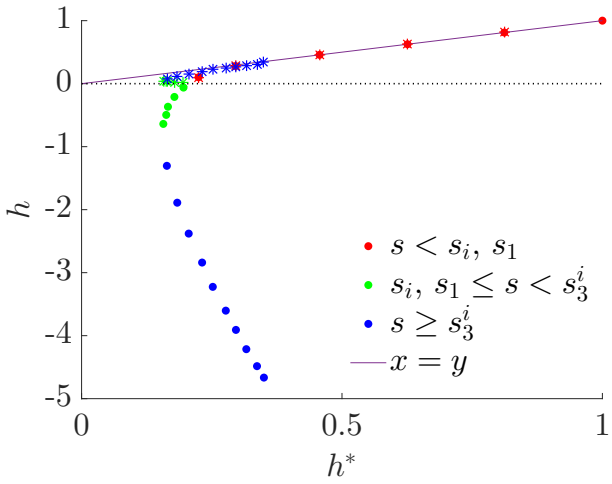


Figure 1.9: Comparison of hierarchies of military-like RSBMs. The parameters are $p = 0.5$, $q = 0$, $R = 32$, s varies in $[0, s_{\max}]$, with $s_{\max} = 0.0294$, and each point refers to a single realisation of the ensemble. The circles represent the pairs $(h_1^*, h_1^{(p)})$, i.e. the optimal hierarchy h_1^* computed with *agony* and the one of the planted hierarchy $h_1^{(p)}$. The stars represent $(h_1^*, \bar{h}_1^*)$ where $\bar{h}_1^*$ is the theoretical hierarchy with the parameters of the SBM estimated via *GraphTool*.

ranking is not the optimal one according to numerical simulations, its value of hierarchy is very close to the one of the optimal ranking, while the planted one strongly mis-estimates the value of h. Thus our computation in the previous Section can be used to reliably estimate the hierarchy of a military-like ranking. This is obviously a partial answer and analytical calculations of the hierarchy of partially inverted rankings are left for a future study.

1.6 Improving hierarchy detection

In this short section we propose to strategies to improve hierarchy detection, on one side regarding the resolution limit, and the by proposing a way to include statistical significance by means of bootstrap resampling.

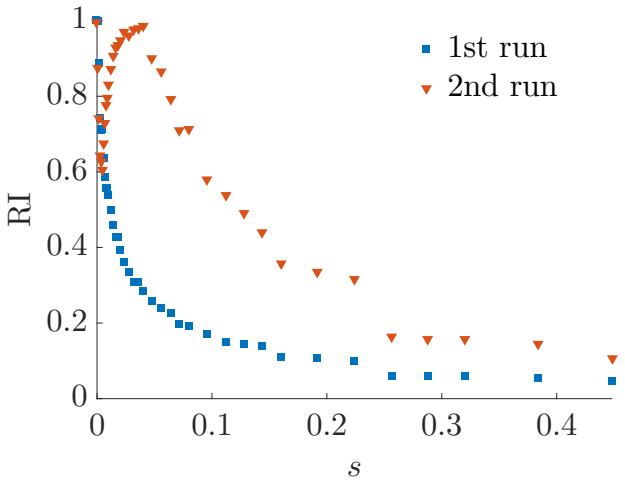


Figure 1.10: Comparison of the Rand Index between the planted ranking and one (blue squares) or two (orange triangles) iterations of *agony*. Data refers to simulation of twitter-like HSBM with parameters $p = q = 0.5$, $R = 32$, $s \in [0, s_{\max}]$, with $s_{\max} = 0.448$, and each point refers to a single realisation of the ensemble.

1.6.1 Iterated agony

In the previous Sections we have shown theoretically and numerically that inference of ranking hierarchies based on agony suffers from significant resolution limit. In twitter-like hierarchies, the identified classes are merging of adjacent classes and thus small classes are not identified. In military-like hierarchies inversions start to play a significant role.

An heuristic method to overcome this problem is to iterate the application of agony. As done with modularity, one can apply agony to each class found in the first iteration of the algorithm, in order to find subclasses. In principle one could continue to iterate, even if the fact that agony finds two classes in an Erdös-Renyi graph suggests a careful design of the stopping criterion. The purpose of this Section is not to propose a full criterion for the improvement of agony via iteration, but to show that indeed improvement is possible, both considering model graphs and real networks.

We first consider the model graphs with twitter-like hierarchy we presented in the previous Section. Fig 1.10 shows the RI between the planted ranking and the one inferred with one (as in the previous Section) and two iterations of

agony with $d = 1$. For small values of s the second iteration does not improve the inference because one iteration already recovers the planted structure. For larger values of s, i.e. weaker structures, the second iteration dramatically outperforms the result of the first one, indicating that iterated applications of *agony* can significantly improve the hierarchies detection. Fig 1.1 shows some details of the obtained results. It is worth noticing that the value of h after the second run is actually smaller than the one from the first run, despite the fact that the RI follows the opposite pattern. This is expected since *agony* finds the optimal value of h, while the RI looks at the similarity with the planted ranking. A closer look to the results of the two iterations (see Tables 1.3,1.4) highlights that high number of classes after the second iteration and high hierarchy in each subclass are associated to the cases for which there is no significant improvement in the RI, hence a successful routine would rely on the control of these two quantities to stop the iterations.

Table 1.1: Simulated graphs, output of the two runs of agony

	1st run			2nd run		
s	h^*	RI	R	h	RI	R'
0	1	1	32	> 0.99	> 0.99	32
0.001	0.98	> 0.99	34	0.93	0.87	97
0.002	0.95	0.89	29	0.81	0.74	128
0.005	0.91	0.71	20	0.51	0.60	160
0.01	0.85	0.54	14	0.51	0.83	102
0.048	0.62	0.26	7	-0.14	0.90	40
0.112	0.41	0.15	4	-0.16	0.54	17
0.224	0.20	0.10	3	-0.20	0.31	9
0.448	0.03	0.05	2	-0.19	0.11	4

We now show that the same phenomenon is relevant also for real networks. We investigate four datasets from SNAP, Stanford Network Analysis Platform (Leskovec and Krevl, 2014), which were also used in (Tatti, 2017). Note that these datasets have been updated since they have been used in (Tatti, 2017) so our results are slightly different.

The networks are quite different in size (from a minimum of 7K nodes to almost 400K nodes) but they are all quite sparse.

- **Wiki vote.** The network contains all the Wikipedia voting data from the inception of Wikipedia till January 2008. Nodes in the network represent

Wikipedia users and a directed edge from node i to node j represents that user i voted for user j.

- **Higgs Reply.** The network contains replies to existing tweets: nodes are users and i is linked to j if i replied to a j's tweet.
- **Higgs mention.** Similar to the previous case, here links represent mentions: a link from i to j means that user i mentioned user j.
- **Amazon.** Network was collected by crawling the Amazon website. It is based on Customers Who Bought This Item Also Bought feature of the Amazon website. If a product i is frequently co-purchased with product j, the graph contains a directed edge from i to j.

Table 1.2 reports some properties of the networks alongside the output of one and two iterations of the *agony* algorithm. Specifically, for each network the table contains: the number of nodes N, the density ($\frac{m}{N(N+1)}$, where m is the number of edges), the percentage of nodes in the largest strongly connected component (*SCC*), the value of h_1^*, the number of classes inferred in the first run (R) and the total number of classes after the second run (R') of *agony*.

Table 1.2: Networks summary. SCC is the percentage of nodes in the largest strongly connected component, h_1^* is the hierarchy of the ranking obtained with one iteration of *agony*, R is the number of classes in the globally optimal ranking, and R' is the number of classes after two iterations of *agony*.

network	nodes	density	SCC	h_1^*	R	R'
Wikivote	$7,115$	$2*10^{-3}$	18%	0.83	12	49
HiggsReply	$38,918$	$2*10^{-5}$	0.8%	0.82	13	27
HiggsMention	$116,408$	$1*10^{-5}$	1%	0.89	20	59
Amazon	$403,394$	$2*10^{-5}$	98%	0.42	17	69

It is clear that the second application the algorithm to the classes detected in the first iteration increases significantly the number of classes, suggesting that the classes identified in the first iteration could be aggregation of smaller classes. In Table 1.5we report more details on the classes identified in the iteration and on the subclasses identified by the second iteration.

Since *agony* penalises links among nodes in the same class, the subgraphs in some cases have no links (those with $*$ in Table 1.5. Notice this would be the

case for any class in a DAG. Thus, a low value of h in each class and a number of sub classes larger than 2 indicate a non trivial and not completely resolved structure of the class.

1.6.2 Bootstrap

Hierarchy detection via agony minimisation misses statistical significance of the inferred partition. It can be shown that when $d = 1$ one can always find a partition with two classes with positive h even if there is no real hierarchical structure. More generally, it may be possible that a particular in-degree or out-degree favour high h, even if there is no real structure underneath.

In this case there is a high number of nodes with only out-degree which are put in the bottom class and certainly contribute to h without showing real structure.

A benchmark partition (call it LCT) is the following:

$$\begin{aligned} L &= \{v \in V : k_u^{\text{in}} = 0\} & r(L) &= 0 \\ T &= \{v \in V : k_u^{\text{out}} = 0\} & r(T) &= 2 \\ C &= V \setminus (L \cup T) & r(C) &= 1 \end{aligned}$$

whose associated h is given by

$$h_{LCT} = \frac{m_{lc} + m_{lt} + m_{ct}}{m} > 0,$$

where

$$\begin{aligned} m_{lc} &= |\{(u,v) : u \in L, v \in C\}| \\ m_{lt} &= |\{(u,v) : u \in L, v \in T\}| \\ m_{ct} &= |\{(u,v) : u \in C, v \in T\}| \\ m &= |E|. \end{aligned}$$

Notice that, if one similarly defines the other subset of edges, it follows $m_{ll} = m_{cl} = m_{tl} = m_{tc} = m_{tt} = 0$ and $m_{cc} > 0$.

This partition is very similar to the bow-tie structure, but it has the advantage that one can easily compute the value of h_{LCT}, whose expression holds whichever d, and it is always possible to define it.

Define

$$\Delta_{LCT} = h^* - h_{LCT}\,.$$

which gives a first measure of the strenght of hierarchy solely due to this coarse partition.

To assess the significance of this value, one can simulate a collection of synthetic networks is obtained from the original one by random rewiring. More specifically, two edges are drawn uniformly at random and their end nodes are swapped, i.e

$$\begin{cases} e_1 = (s_1, t_1) \\ e_2 = (s_2, t_2) \end{cases} \quad \rightarrow \quad \begin{cases} e_1' = (s_1, t_2) \\ e_2' = (s_2, t_1) \end{cases} \quad .$$

All the randomised graphs have the same associated LCT partition in order to compare Δ_{LCT} of the real network with the randomised version ($\Delta_{LCT}^{(r)}$), but preserving the degree distributions (as rewiring does) is not sufficient. Depending on the classes of the connected nodes, each link can be of one of the following types

$$(L,C) \quad (L,T) \quad (C,C) \quad (C,T),$$

where

$$(L,C) = \{(u,v) :, u \in L\,, v \in C\},$$

and similarly for the others. There are 10 ($=\binom{4+2-1}{2}$) possible unordered pairs of types, of those only

$$\begin{cases} (L,C) \\ (C,C) \end{cases} \qquad \begin{cases} (C,C) \\ (C,C) \end{cases} \qquad \begin{cases} (C,T) \\ (C,C) \end{cases}$$

preserves h_{LCT} when swapped, so only these combinations are considered.

In practice, fixed a number of moves (n_{MC}), a number M of randomised replicas of the original network are generated and their h is computed, obtaining the distribution oh h for the specific level of rewiring, from which one can compute the p-value associated to the measured $Delta$ The purpose of this approach is not only to achieve some measure of statistical significance for h, but also to understand what percentage of links has to be changed in order to destroy the hierarchical structure. As one can expect, increasing the number of pair swapped, $h^{(r)}$ decreases, up to a limit $h*_{LCT} \geq h_{LCT}$.

1.7 Conclusions

In this Chapter we considered the problem of hierarchy inference in directed networks using a non parametric method, the metric agony, which penalises links contrary to the hierarchy. To asses its ability to correctly identify hierarchical structures we introduce a generative model for hierarchical networks, RSBM, for which we are able to characterise analytically the optimal partition. We have shown both analytically and empirically that when the hierarchy is not strong enough alternative rankings, obtained by merging, splitting or inverting classes in the original ranking, may have a value of the hierarchy larger than the planted one, hence signalling the presence of resolution limit for this method of inference. This is somewhat similar to the well known resolution limit of modularity in community detection. We showed two possible strategies to enrich the information on the inferred partition without the need of changing algorithm or metric. The first partially overcomes the resolution limit by iteratively apply agony, and we show that indeed one can obtain a significant improvement. The second strategy uses bootstrap to asses the statistical significance of the inferred partition.

There are several directions along which extend the present work. The natural extension of considering non integer d, in particular in the interval $(0, 1)$ will be addressed in the next Chapter. Then, one may investigate, with the same analytical formulae the case of non uniform cardinality of the classes, such as a pyramidal hierarchy with a small top class and larger bottom classes. With a careful choice of the sizes one might be able to maintain analytical tractability. Finally, other methods to identify ranking hierarchies could be investigated, for example suitably modifying the agony function or by considering optimisations for a set of functions.

Appendix 1.A Detailed proofs

In this Supporting material we present details and extended formulae for the propositions.

To start, we consider the values of agony for general d depending on the choice of the alternative rankings.

- **No inversion and splitting**. When $b < 0$, each class is divided into 2^{-b} classes. As for the affinity matrix, the only part affected by the change in

the ranking is the one above the diagonal, which has no impact on the computation of $\mathbb{E}[A_d(G, r^{(b)})]$. Hence one has

$$\mathbb{E}[A_d(G, r^{(b)})] = s\left(\frac{N}{R}2^b\right)^2 \sum_{k=0}^{2^{a-b}-1} (k+1)^d (2^{a-b}-k). \tag{1.10}$$

- **No inversion and merging**. When $b \geq 0$, for any pair (i, j) it holds:

$$\mathbb{E}[m_{ij}] = \left(\frac{N}{R}\right)^2 \begin{cases} 2^{2b}s & j < i \\ (2^b-1)p + 2^{b-1}(2^b+1)s + (2^{b-1}-1)(2^b-1)q & j = i \\ p + (2^{2b}-1)q & j = i+1 \\ 2^{2b}q & j > i+1\,, \end{cases} \tag{1.11}$$

which gives

$$\begin{aligned}\mathbb{E}[A_d(G, r^{(b)})] =& s\left(\frac{N}{R}2^b\right)^2 \sum_{k=1}^{2^{a-b}-1} (k+1)^d (2^{a-b}-k) + \\ &+ 2^{a-b}\left((2^b-1)p + 2^{b-1}(2^b+1)s + (2^{b-1}-1)(2^b-1)q\right)\end{aligned}$$

- **Inversion and merging**. When $b \geq 0$ the expression for agony of the inverted ranking becomes

$$\begin{aligned}\mathbb{E}[A_d(G, r^{(i,b)})] =& 2^{2b}\left(\frac{N}{R}\right)^2 q \sum_{k=2}^{2^{a-b}-1} (k+1)^d \left(2^{a-b}-k\right) + \\ &+ 2^d \left(\frac{N}{R}\right)^2 \left(2^{a-b}-1\right)\left(\left(2^{2b}-1\right)q + p\right) + \\ &+ 2^{a-b}\left(\frac{N}{R}\right)^2 \left(\left(2^b-1\right)p + \left(2^{b-1}-1\right)\left(2^b-1\right)q + 2^{b-1}\left(2^b+1\right)s\right)\end{aligned}$$

- **Inversion and splitting** When $b < 0$

$$\mathbb{E}[A_d(G, r^{(i,b)})] = \left(\frac{N}{R}2^b\right)^2 \sum_{k=0}^{2^{-b}-1} (k+1)^d \left(2^a\left(2^{-b}-k\right)s + (2^a-1)\,kp\right) + \\ + \left(\frac{N}{R}2^b\right)^2 \sum_{k=0}^{2^{-b}-1} \left(k+1+2^{-b}\right)^d \left((2^a-1)\left(2^{-b}-k\right)p + (2^a-2)\,kq\right) + \\ + \left(\frac{N}{R}2^b\right)^2 q \sum_{k=0}^{(2^a-2)2^{-b}} \left(k+1+2^{1-b}\right)^d \left((2^a-2)\,2^{-b}-k\right).$$

Then, we present the proofs of the propositions.

Proof of Proposition 1

We explicitly show that in the $d = 1$ case there exists critical values for s at which the planted ranking ceases to maximize hierarchy both for Twitter-like and Military-like hierarchies.

To determine the optimal number of classes we first treat b as a continuous variable and compute the derivative oh $\bar{h}_1$ with respect to it. The unique critical point is denoted by b^* and it is given by

$$b^* = \frac{1}{2}\log_2 \frac{2^{2a}s + 6(q-p)}{3q-s}.$$

Note that it must hold

$$0 \leq b \leq a$$

and we want to avoid the continuous relaxation at the boundaries so we consider the extreme values separately.

When $p \geq q > s$ (*Twitter-like hierarchy*), we first notice that

$$\frac{\partial \bar{h}_1}{\partial b}|_{b=b*} < 0$$

Moreover, it holds

$$\bar{h}_1(b = a-1) > \bar{h}_1(b = a),$$

that is the trivial ranking is never better than that with two classes.

Moreover, we denote with s_2 the value of s such that the rankings with two and three classes have the same value of hierarchy, i.e.

$$\bar{h}_1\left(b = a - \log_2 3\right) = \bar{h}_1(b = a - 1)\,,$$

since for any fixed $b > 0$, $\bar{h}_1$ is monotone decreasing with respect to s,

$$\bar{h}_1\left(b = a - \log_2 3\right) < \bar{h}_1(b = a - 1)\,\forall s \geq s_2\,.$$

Similarly, one can find the critical value s_m such that the ranking with of $R-1$ classes shares the value of hierarchy with the planted one,

$$\bar{h}_1(b = 0) = \bar{h}_1\left(b = a - \log_2\left(2^a - 1\right)\right).$$

Finally, we can combine the results to obtain the optimal number of classes for the direct ranking in the region $p \geq q > s$:

$$\tilde{R}^* = \begin{cases} R & s \leq s_m \\ 2^{a-b^*} & s_m < s < s_2 \\ 2 & s \geq s_2\,, \end{cases} \tag{1.12}$$

where

$$\begin{aligned} s_m &= \frac{6\left(2^a - 1\right)p - 3\left(2^a - 2\right)q}{2^a - 4^a + 8^a} \\ s_2 &= \frac{3}{7}\frac{\left(4^a - 12\right)q + 12p}{4^a} \end{aligned} \tag{1.13}$$

With a reasoning similar to the one carried before, one gets that when $p \geq q > s$ the optimal number of classes for the inverted ranking is such that

$$1 \leq \tilde{R}^* \leq 2$$

hence,

$$h_1^{i,*} \leq 0,\, \forall\, p \geq q > s,\, \forall\, a\,.$$

One can conclude that the optimal ranking for the twitter-like hierarchy is the direct one with a number of classes which depends on s, according to (1.12).

When $q = 0$ (***Military-like hierarchy***), when it is defined, we have

$$\frac{\partial^2 \bar{h}_1}{\partial b^2}\Big|_{b=b*} > 0\,,$$

so, to obtain the optimal directed ranking we only need to check the extreme values for b, i.e. $b = 0, b = a$. The optimal number of classes for the direct ranking is given by

$$\tilde{R}^* = \begin{cases} R & s \leq s_{m|_{q=0}} \\ 1 & \text{otherwise}, \end{cases}$$

where

$$s_1 = \frac{6p}{2^a(1+2^a)}.$$

Then, one can consider the inverted ranking.
It easy to verify that

$$\mathbb{E}[A_1(G, r^{(i,b)})] > \mathbb{E}[A_1(G, r^{(p)})], \forall\, b < 0,$$

that is, also for the inverted ranking splitting is never optimal on average.

As for merging, the optimal choice for b is given by

$$b^{i,*} = \frac{1}{2} log_2 \frac{2p}{s},$$

which is well defined when $s > \frac{2}{4^a}p$ and satisfies $\frac{a}{2} \leq b^{i,*} \leq a$. The optimal number of classes fro the inverted ranking is given by

$$\tilde{R}^{i,*} = \begin{cases} 1 & s \leq s_2^i \\ 2 & s_2^i < s \leq s_3^i \\ 2^{a-b^{i,*}} & s > s_3^i \end{cases},$$

where

$$s_2^i = 2^{2-2a}p$$
$$s_3^i = 3s_2^i.$$

When $s \leq s_1$, the planted ranking is optimal and non zero and decreasing, and

$$s_2^i < s_1 < s_3^i. \tag{1.14}$$

Denote by s_i the value of s such that

$$\bar{h}_1^i(b = a-1) = \bar{h}_1(b = 0).$$

One gets

$$s_i = \frac{12p}{3\ 2^a + 2^{2a+1} - 2},$$

and when $s > s_i$ the optimal inverted ranking has a higher value of hierarchy than the planted, which is the optimal directed one.

Finally, one can write the expression for the estimate of the optimal value of h in proposition 1.

For $p \geq q > s$,

$$\bar{h}_1^* = \begin{cases} -\frac{(2^a-2)(-6(2^a-1)q+2^a(2^a+2)s-6p)}{6(2^a(2p-3q+s)+4^a(q+s)-2p+2q)} & s \leq s_m \\ \frac{3((4^a+2)q-2p)\sqrt{\frac{4^a s-6p+6q}{3q-s}}-2^{a+1}(4^a s-6p+6q)}{3\sqrt{\frac{4^a s-6p+6q}{3q-s}}(2^a(2p-3q+s)+4^a(q+s)-2p+2q)} & s_m < s < s_2 \\ \frac{4^a(q-s)+4p-4q}{2(2^a(2p-3q+s)+4^a(q+s)-2p+2q)} & s \geq s_2\,. \end{cases}$$

When $q = 0$,

$$\bar{h}_1^* = \begin{cases} \frac{2^a(6p+s)-8^a s-6p}{6(2^a-1)p+3\ 2^a(2^a+1)s} & s \leq s_i \\ \frac{4^a s-4p}{2(2^a(2p+s)+4^a s-2p)} & s_i < s \leq s_3^i \\ \frac{-2^{a+\frac{3}{2}} s\sqrt{\frac{p}{s}}+4^a s+2p}{2^a(2p+s)+4^a s-2p} & s > s_3^i\,. \end{cases}$$

Proof of Proposition 2

We here proceed to show that in the $d = 0$ case (FAS), both for Twitter-like and Military-like hierarchies, agony is minimized by the ranking where nodes are partitioned in singletons. When $b > 0$, the derivative of h with respect to b is negative hence the planted ranking is better that any other with a fewer number of classes. Instead, when $b < 0$ one has

$$\mathbb{E}[A_0(G, r^{(b)})] = s(2^a + 2^b)\left(\frac{N}{R}\right)^2$$

which implies

$$\mathbb{E}[A_0(G, r^{(b)})] < \mathbb{E}[A_0(G, r^{(p)})],\ \forall\, b < 0\,,$$

and

$$\frac{\partial \bar{h}_0}{\partial b} = -\frac{2^{a+b-1}}{m} s < 0 \quad \forall\, b < 0$$

So the optimal ranking is obtained for the limit value of b

$$b^* = -\log_2 \frac{N}{R}, \quad \tilde{R}^* = N\,.$$

Similar computations give that any inverted ranking (i.e $\forall\, b$) has never a higher value of hierarchy than the the ranking we just discussed.

One get the formula in proposition 2

$$h_0^* = 1 - \frac{2^{2a}(N+1)s}{(2^{2a}(q+s) + 2^a(2p-3q+s) - 2p + 2q)N}$$

Proof of Proposition 3

For the case $d = 2$ one can follow the same procedure we showed for $d = 1$ and find the critical values for resolution threshold.

When $p \geq q > s$, the optimal number of classes is given by

$$\tilde{R}_2^* = \begin{cases} R & s \leq s_{2,m} \\ 2^{a-b_2^*} & s_{2,m} \leq s \leq s_{2,1} \\ 1 & s \geq s_{2,1}\,, \end{cases}$$

where

$$\begin{aligned} b_2^* = \log_2\Big(&\frac{2\sqrt[3]{2}\left(2^{2a}s - 3p + 3q\right)}{\sqrt[3]{\beta + 3^5\ 2^{3a}q^2s - 3^4\ 2^{3a+2}qs^2 + 3^3\ 2^{3a+2}s^3}} + \\ &+ \frac{\sqrt[3]{\frac{1}{3}\beta + 2^4\ 2^{3a}q^2s - 3^3\ 2^{3a+2}qs^2 + 3^2\ 2^{3a+2}s^3}}{\sqrt[3]{2\,3^2}(3q-2s)}\Big), \\ \beta = &\sqrt{3^6\ 2^{6a}s^2(3q-2s)^4 - 2^5\,3^3(3q-2s)^3\,(4^as - 3p + 3q)^3}\,. \end{aligned} \tag{1.15}$$

is the unique zero of the first order derivative of $\bar{h}_2$ with respect to b, and

$$\begin{aligned} s_{2,m} &= \frac{6\left(2^{1-a}(q-p) + 2p - q\right)}{-3\ 2^a + 2^{3a+1} + 4^a + 4} \\ s_{2,1} &= \frac{2^{2a}q + 4p - 4q}{3\,2^{2a}} \end{aligned}$$

with $s_{2,1}$ being the value of s such that

$$\bar{h}_2(b = a-1) = \bar{h}_2(b = a) = 0\,.$$

When $q = 0$, the planted ranking is optimal and gives positive $\bar{h}_2$ when $s < s^0_{2,1}$, where

$$s^0_{2,1} = \frac{3\ 2^{2-a}p}{5\ 2^a + 4^a + 4}\,.$$

For the inverted ranking instead one can compute the optimal choice for the number of classes, that is

$$\tilde{R}^{i,*}_2 = \begin{cases} a & s \le s^i_{2,2} \\ a-1 & s^i_{2,2}p < s < s^i_{2,3} \\ \frac{\log\left(\frac{6p}{s}\right)}{\log(4)} & s > s^i_{2,3}\,, \end{cases}$$

where

$$b^{i,*}_2 = \frac{\log\left(\frac{6p}{s}\right)}{2\log(2)}\,,$$

and

$$s^i_{2,2} = \frac{12}{2^{2a}}p$$
$$s^i_{2,3} = 3s^i_{2,2}\,.$$

For any choice of p and a, it holds

$$s_{2,1} < s^i_{2,2}\,,$$

so the inverted ranking is optimal for $s > s^i_{2,2}$.

Appendix 1.B Numerical results

For each network and for each class, the table contains the size of the class, n_i, as a percentage of the total number of nodes), the value of h, and the number of sub-classes inferred (R).

* empty

Table 1.3: **Simulated graphs, details for classes**

	$s = 0.001$			$s = 0.002$			$s = 0.005$			$s = 0.01$		
cl.	$n_i(\%)$	h^*	R	$n_i(\%)$	h^*	R	$n_i(\%)$	h^*	R	$n_i(\%)$	h^*	R
1	<0.01	1*	1	0.03	1	3	0.03	1	11	0.03	0.95	12
2	0.03	1	3	0.03	1	3	0.03	1	8	0.06	0.93	5
3	0.03	1	4	0.03	0.88	3	0.03	1	7	0.06	0.93	6
4	0.03	1	3	0.03	1	4	0.03	0.98	7	0.09	0.93	5
5	0.03	1	3	0.03	1	3	0.06	0.97	7	0.09	0.94	5
6	0.03	1	3	0.03	1	4	0.06	0.97	7	0.095	0.95	5
7	0.03	1	3	0.03	1	5	0.06	0.97	10	0.09	0.94	6
8	0.03	1	2	0.03	1	5	0.06	0.97	9	0.09	0.93	6
9	0.03	1	3	0.03	1	3	0.06	0.96	8	0.09	0.94	7
10	0.03	1	2	0.03	1	4	0.06	0.97	8	0.09	0.94	6
11	0.03	1	3	0.03	1	4	0.06	0.97	11	0.06	0.93	6
12	0.03	1	2	0.045	0.99	5	0.06	0.97	6	0.06	0.94	7
13	0.03	1	4	0.05	0.99	6	0.06	0.97	10	0.03	0.87	10
14	0.03	1	2	0.05	0.99	8	0.06	0.97	8	0.03	0.91	16
15	0.03	1	3	0.05	0.98	7	0.06	0.97	7			
16	0.03	1	2	0.05	0.99	6	0.06	0.97	7			
17	0.03	1	3	0.04	0.99	5	0.03	0.94	7			
18	0.03	1	3	0.03	1	5	0.03	1	6			
19	0.03	1	3	0.03	1	4	0.03	1	9			
20	0.03	1	3	0.03	1	5	0.03	1	7			
21	0.03	1	3	0.03	1	4						
22	0.03	1	3	0.03	1	4						
23	0.03	1	3	0.03	1	4						
24	0.03	1	3	0.03	1	3						
25	0.03	1	5	0.03	1	4						
26	0.03	1	3	0.03	1	5						
27	0.03	1	3	0.03	1	4						
28	0.03	1	3	0.03	1	3						
29	0.03	1	2	0.03	1	5						
30	0.03	1	3									
31	0.03	1	4									
32	0.03	1	3									
33	0.03	1	3									
34	<0.01	1*	1									

Table 1.4: **Simulated graphs, details for classes**

	$s = 0.048$			$s = 0.112$			$s = 0.224$			$s = 0.448$		
cl.	$n_i(\%)$	h	R	$n_i(\%)$	h	R	$n_i(\%)$	h	R	$n_i(\%)$	h	R
1	0.06	0.71	4	0.18	0.44	4	0.27	0.21	3	0.51	0.03	2
2	0.16	0.70	6	0.32	0.42	5	0.48	0.20	3	0.49	0.03	2
3	0.22	0.67	7	0.32	0.42	4	0.26	0.21	3			
4	0.22	0.67	7	0.18	0.44	4						
5	0.19	0.69	7									
6	0.12	0.72	4									
7	0.03	0.35	5									

Table 1.5: **Real networks: details for classes**

	Wikivote			HiggsReply			HiggsMention			Amazon		
cl.	$n_i(\%)$	h^*	R	$n_i(\%)$	h^*	R	$n_i(\%)$	h^*	R	$n_i(\%)$	h^*	R
1	0.67	1*	1	0.60	0.03	2	0.77	0.13	2	0.02	< 0.01	3
2	0.01	0	1	0.31	0.34	2	0.16	0.78	4	0.03	0.01	3
3	< 0.01	0.24	3	0.04	0.64	2	0.03	0.78	3	0.10	0.01	3
4	0.01	0.38	6	0.01	1	2	0.01	0.80	2	0.20	0.01	5
5	0.02	0.26	5	< 0.01	0.50	2	< 0.01	0.81	2	0.25	0.05	6
6	0.04	0.23	6	< 0.01	1	2	< 0.01	0.85	2	0.20	0.08	6
7	0.06	0.20	5	< 0.01	1	3	< 0.01	0.53	2	0.11	0.08	6
8	0.09	0.32	8	< 0.01	0.67	5	< 0.01	0.65	3	0.06	0.07	5
9	0.08	0.72	10	0.01	0.24	2	< 0.01	0.66	4	0.02	0.06	5
10	0.04	1	2	< 0.01	0.83	2	< 0.01	0.55	4	0.01	0.05	5
11	< 0.01	0	1	< 0.01	1*	1	< 0.01	0.59	7	< 0.01	0.06	4
12	< 0.01	1*	1	< 0.01	1*	1	< 0.01	0.46	6	< 0.01	0.05	4
13				< 0.01	1*	1	< 0.01	0.60	6	< 0.01	0.04	3
14							< 0.01	0.66	6	< 0.01	0.05	4
15							< 0.01	0.82	1	< 0.01	0.07	3
16							< 0.01	0.69	1	< 0.01	0.05	3
17							< 0.01	1*	1	< 0.01	0	1
18							< 0.01	1*	1			
19							< 0.01	1*	1			
20							< 0.01	1*	1			

* empty

2

Beyond resolution limits in hierarchy detection

In this Chapter we propose an heuristic algorithm for hierarchy detection with varying resolution. First, we extend the analysis of previous Chapter 1 to non integer d. We show analytically that they allow to loosen the resolution limit of the benchmark. Also, the parameter d acts as resolution parameter: acting on it changes the ability of the optimisation procedure to recover small classes. This class of metrics is characterised by NP-hard optimisation problem and to our knowledge no algorithm is present in the literature. This justify the need for an algorithm for the inference of hierarchies for general d. We propose a greedy heuristic to solve the detection problem which can be used for any value of the resolution parameter. By testing the heuristic on synthetic graphs we are able to assess its performance: the behaviour is very close to analytic estimate in most cases, and it effectively address the problem resolution as it is able to recover the planted structure better in conditions where the benchmark metric fails to do so.

The interest in this investigation is that when community detection is applied to real data, one should be able understand the level of confidence of the output partition. Thanks to the resolution parameter, the strategy for hierarchies is very similar to what is usually done with modularity: one infers the optimal partition according to different values of the parameters and compare the results. It must be noted that in the case of modularity the sets obtained

for finer resolutions are nested into the one obtained for coarser levels, while this is not necessarily the case for hierarchies due to the specific form of the optimisation problem.

This chapter is divided into two parts. For the relevant literature we refer to the previous chapter. In Section 2.1 we tackle the problem analytically: we show that the resolution limit can be loosened by acting on the resolution parameter. In Section 2.2 we propose an heuristic algorithm to solve the optimisation and we test its performance. In Section 2.3 we draw conclusions.

2.1 Resolution limits for non integer d

In this section we characterise the resolution limits of hierarchy detection using agony metric with non integer d. Recall that the metric agony is defined as following

$$A_d^*(G) = \min_{r \in \mathcal{R}} \sum_{(u,v)\in E} f_d(r(u) - r(v)), \tag{2.1}$$

where $\mathcal{R}$ denotes the set of all rankings and

$$f_d(x) = \begin{cases} (x+1)^d & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad d \geq 0,$$

Equivalently, one can define the *hierarchy* of a directed graph as

$$h_d^*(G) = 1 - \frac{A_d^*(G)}{m}.$$

In previous Chapter 1 the behaviour of A_d for integer d was studied, and the choice $d = 1$ resulted to be the best both for resolution and computational feasibility. Here we want to perform a similar analysis for non integer $d \in (0, 1)$. The interest in this interval is given by the fact that the two extreme cases of $d = 0$ and $d = 1$ showed an opposite behaviour with respect to the structure of the optimal partition, so we expect to move from one to the other as d changes and to be able to loosen the resolution limit of the benchmark $d = 1$. For the case $0 < d < 1$ one cannot rely on analytical expressions since the sum in (1.10) has no close form when $d \notin \mathbb{N}$.

In the following Propositions 4-7 we show that the resolution threshold for twitter-like hierarchies could be improved by decreasing the value of d.

More specifically, we first characterise the behaviour $\bar{h}^*_d$ in the neighbourhoods of $d = 0$ and $d = 1$: we find that when d is sufficiently close to 0, then split ranking has lower cost than the planted one, while for d sufficiently close to 1, the planted rank outperforms any split rank. This is consistent with the intuition of transition between the two extremes. Moreover, for military-like hierarchy, we find that for d sufficiently close to 0, the planted ranking cannot be outperformed by any inverted one, instead in a left neighbourhood of $d = 1$, we find a limit value for s above which the planted ranking cannot be globally optimal since it is outperformed by the optimal inverted ranking. Secondly we focus on the twitter hierarchy and we find an estimate for the limit of value of s above which the planted ranking is outperformed by the merged one and we show that this limit $s_{d,m}$ is decreasing in d . Finally, for $d \geq \frac{1}{2}$ we show that $s_{d,m}$ is the resolution limit.

To improve readability, all the proofs are presented in the Appendix 2.A.

Proposition 4 (Splitting characterisation on neighboorhood on 0 and 1). $\forall RSBM(p, q, s, R = 2^a)$ *satisfying Eq.* (1.6), $\forall\, b < 0 \,\exists\, \underline{\epsilon}_{a,b} : \forall\, d \leq \underline{\epsilon}_{a,b}$ *it is* $\bar{h}_d(b) > \bar{h}^{(p)}_d$, *i.e. any splitted ranking outperforms the planted one. The explicit value of the bound in* d *is*

$$\underline{\epsilon}_{a,b} = \frac{\log \frac{1+2^a}{2^a+2^b}}{(a-b)\log 2}$$

Moreover, $\forall RSBM(p, q, s, R = 2^a)$ *satisfying Eq.* (1.6), $\forall\, b < 0 \,\exists\, \overline{\epsilon}_{a,b} : \forall\, d \geq 1 - \overline{\epsilon}_{a,b}$ *it is* $\bar{h}^{(p)}_d > \bar{h}_d(b)$, *i.e. the planted rank outperforms the split ranking. Finally*

$$\overline{\epsilon}_{a,b} = \frac{\log \frac{(2^a+2^b)(2^a+2^{1+b})}{2^b(1+2^a)(2+2^a)}}{(a-b)\log 2}$$

The proposition holds for any RSBM and sets the intervals around $d = 0$ where the planted ranking cannot be optimal because outperformed by any split ranking. The second part of the proposition shows the existence of a left neighbourhood of $d = 1$ where the planted ranking cannot be outperformed by any split one.

The following proposition sets the intervals around $d = 0$ where the planted ranking cannot be outperformed by any inverted one. The second part of the proposition shows the existence of a left neighbourhood of $d = 1$ where for any $s > s_{d,i}$ the planted ranking cannot be globally optimal since it is outperformed by the optimal inverted ranking.

Proposition 5 (Inversion characterisation on neighbours of 0 and 1 for military hierarchy). $\forall RSBM(p, q = 0, s, R = 2^a)$ *satisfying Eq.*(1.6) *(i.e. military-like hierarchy),* $\exists \underline{\epsilon}_a : \forall d \leq \underline{\epsilon}_a$ *and* $\forall b \forall s$ *it is* $\bar{h}_d^{(i)}(b) < \bar{h}_d^{(p)}$ *, with*

$$\underline{\epsilon}_a = \frac{\log \frac{2^a+1}{2^a-1}}{(a-1)\log 2},.$$

i.e. the inverted rankings are never better than the planted one. Moreover, define

$$s_{d,i} = \frac{4p\left(2^a + 2^d - 2\right)}{4\sum_{k=1}^{2^a-1}(2^a-k)(k+1)^d - 2^a(2^a-2)}. \tag{2.2}$$

Then $\exists \overline{\epsilon}_a : \forall d \geq 1 - \overline{\epsilon}_a, \forall s > s_{d,i} \exists b_d^{i,*} > 0 : \bar{h}_d^i(b_d^{i,*}) > \bar{h}^{(p)}$ *, with*

$$\overline{\epsilon}_a = \frac{\log \frac{(2^a+4)(2^a-1)}{2(2^a+1)}}{(a-1)\log 2}.$$

i.e. for any $s > s_{d,i}$ *the planted ranking cannot be globally optimal because there exists an inverted ranking which outperforms it.*

We now consider the merged case ($b > 0$) and focus on the twitter-like hierarchy. The critical value $s_{d,m}$ of s for which merged partitions outperform the planted rank is characterised by the following proposition.

Proposition 6 (Merging characterisation for twitter-like hierarchy). *Given a* $RSBM(p, q, s, R = 2^a)$ *satisfying* $p \geq q > s$ *(i.e. twitter-like hierarchy). Define*

$$s_{d,m} = \frac{1}{f_s}\left(p - \frac{2^a - 2}{2(2^a-1)}q\right), \tag{2.3}$$

where

$$f_s = \sum_{k=1}^{2^a-2}(k+1)^d\left(k\frac{2^{a+1}-1}{(2^a-1)^2} - \frac{2^a}{2^a-1}\right) + 2^a + 2^{ad} - \frac{2^{a-1}(2^{a+1}-1)}{2^a-1}.$$

Then $\forall s > s_{d,m} \exists\, b_d^* > 0 : \bar{h}_d(b_d^*) > \bar{h}_d^{(p)}$*. Moreover,* $s_{d,m}$ *decreases with* d*.*

We note that when $d = 1$, $f_s > 0$ and one recovers the expression for $s_m = s_{1,m}$, while when $d = 0$, also $f_s = 0$, so there is no solution, which is consistent with the fact the merging is never optimal in this case.

Proposition 7 (Resolution limit for twitter hierarchy). *Consider a* $RSBM(p, q, s, R = 2^a)$ *such that* $p \geq q > s$ *(Twitter). If* $R \geq 14$ *and* $d \geq \frac{1}{2}$, $s_{m,d}$ *defined in Eq.* (2.3) *of Proposition 6 sets the resolution limit. Moreover,* $\exists\, b_d^* > 0$ *such that*

$$\bar{h}_d^* = \begin{cases} \bar{h}_d^{(p)} & s < s_{m,d} \\ \bar{h}_d(b = b_d^*) & s \geq s_{m,d} \end{cases}$$

The consequence of the last two proposition is that by decreasing d up to 0.5 one obtains a looser resolution threshold, i.e. hierarchies are better identified with smaller values of d.

2.2 An heuristic for generalised agony

In this section we propose a greedy heuristic algorithm for the optimal partition for generic d, we call it *Generalised Agony Heuristic*, *GAH*. We first present the algorithm, providing the pseudo code, then we test its performance by running it on synthetic graphs drawn from RSBM. Comparing the partitions obtained on the same graph for different d one is able to capture the change in resolution. In real application one can not necessarily rely on the optimal d^* (which is unknown) but a possible strategy would be to try different d and choose among the resulting partitions.

2.2.1 The algorithm

The greedy heuristic is an iterative procedure based on the observation that agony metric in first approximation tends to split nodes according to the *net* degree $\tilde{\Delta}(v) = k_v^{(\text{out})} - k_v^{(\text{in})}$: nodes with $\tilde{\Delta} > 0$ tend to be put int lower classes, those with $\tilde{\Delta}$ close to 0 are in the middle classes, while those with $\tilde{\Delta} < 0$ are more likely to be put into upper classes. This can be seen for example by finding a partition with positive h in a Erdös-Réyni graph, i.e. a graph with no hierarchical structure. It is sufficient to define the partition:

$$C_1 = \{v \in V : \tilde{\Delta}(v) > 0\}$$
$$C_2 = \{v \in V : \tilde{\Delta}(v) \geq 0\}.$$

From the definition above follows

$$m_{11} + m_{21} < m_{11} + m_{12}$$

and

$$m_{21} + m_{22} \leq m_{22} + m_{12}$$

where $m_{i,j}$ is the number of link from class i to class j in the realisation (not the expected value) and $m = \sum m_{ij}$. Finally,

$$h_d = \frac{m_{12} - (2^d - 1)m_{21}}{m},$$

which means $h_d > 0$ when

$$m_{12} > (2^d - 1)m_{21}$$

The heuristic *GAH* comprises three parts. At first we separate the strongly connected component (SCC) from the out and in components using the algorithm in *Extract Layer* from Romei et al. (2015). Next, we use the function *DegreeCut* which split each set C^x of the SCC (at first iteration, SCC itself) into two according to the value of $\tilde{\Delta}$ of each node:

$$C_o^x = \{v \in C^x : \Delta(v) > 0\}$$
$$C_i^x = \{v \in C^x : \Delta(v) \leq 0\}$$

Rather then simply using net degree, we found more effective to use a modification of agony cost function, i.e.

$$\Delta(v) = \sum_{(v,u)\in\bar{E}} \tilde{f}_d(r(v) - r(u), \alpha) - \sum_{(u,v)\in\bar{E}} \tilde{f}_d(r(u) - r(v), \alpha)$$

In the computation of Δ, line 23, $\tilde{f}_d(x, \alpha)$ is the modified cost function of the optimisation:

$$\tilde{f}_d(x, \alpha) = \begin{cases} (x+1)^d & x \geq 0 \\ -\frac{\exp(x)}{\alpha} & x < 0, \end{cases}$$

and we put α equal to the number of sets in the core. The intuition behind this choice is that using only the degree (i.e without considering the distance in rank between the connected nodes) one would lose the information on the order of classes, which is crucial in hierarchy. The proposed modification, which gives negative weights to links in the *correct* direction, is intended to improve

recovery for military hierarchy (for twitter hierarchy the function f_d in (1.2) actually gives slightly better results). Without the correction the contribution of links seconding the hierarchy would not play a role in the computation of Δ. This is not a problem in the general formulation because one has to optimise on the set of all the possible partition, but becomes an issue in the approach proposed here because only a subset of the possible partition is explored .

The splitting is repeated until there is no improvement in the value of h. Note that each iteration only considers splitting existing sets from previous iteration, and to speed up the procedure, the obtained partition is either accepted or rejected as a whole, without trying to include only some of the modifications. Finally, the rank is compressed in order to avoid gaps in the numbering of classes. Note that the procedure just described can be easy generalised to the weighted case, it is sufficient to update the definition of Δ.

In the following we present the pseudo code for the heuristic. (A, B, C) indicates an *ordered* set and the union $\cup$ preserves the order, i.e

$$(A, B, C) \cup (D, E) = (A, B, C, D, E).$$

Algorithm 1 Generalised Agony Heuristic

```
1:  procedure GAH(G, d)
2:      {L_1, ..., L_l, C, T_1, ..., T_t} ← ExtractLayer(G)
3:      coreList ← (C)
4:      ranking ← (L_1, ..., L_l, C, T_1, ..., T_t)
5:      newRanking ← ranking
6:      h_old ← −1
7:      h ← h(G, d,ranking)
8:      while h > h_old do
9:          coreList ← DegreeCut(G,d,coreList)
10:         ranking ← newRanking
11:         newRanking ← (L_1, ..., L_l) ∪ coreList ∪ (T_1, ..., T_t)
12:         h_old ← h
13:         h ← h(G, d,newRanking)
14:     ranking ← CompressRanking(G,ranking,(T_1, ..., T_t))
15:     return h_old,ranking
```

17: **function** DegreeCut(G, d,coreList)
18: $\bar{V} \leftarrow \{v : \exists\, C \in \text{coreList}, v \in C\}$
19: $\bar{E} \leftarrow E \cap (\bar{V} \times \bar{V})$
20: $r \leftarrow$Rank($\bar{V}$,coreList)
21: $\alpha \leftarrow |\text{coreList}|$
22: **for** $v \in \bar{V}$ **do**
23: $\Delta(v) = \sum_{(v,u)\in\bar{E}} \tilde{f}_d(r(v) - r(u), \alpha) - \sum_{(u,v)\in\bar{E}} \tilde{f}_d(r(u) - r(v), \alpha)$
24: newCoreList $\leftarrow \{\}$
25: **for** $C \in$ CoreList **do**
26: $C_o \leftarrow \{v \in C : \Delta(v) > 0\}$
27: $C_i \leftarrow \{v \in C : \Delta(v) \leq 0\}$
28: newCoreList $\leftarrow$ newCoreList $\cup(C_o, C_i)$
29: **return** newCoreList

30: **function** CompressRanking(G,ranking,$(\mathcal{T}_1, \ldots, \mathcal{T}_t)$)
31: $r \leftarrow$ Rank($\bar{V}$,ranking)
32: $V_T \leftarrow \cup_k \mathcal{T}_k$
33: **for** $v \in V_T$ **do**
34: $r(v) \leftarrow \max_{\{u:(u,v)\in E\}} r(u) + 1$
35: newRanking $\leftarrow$ Rank $^{-1}(r)$
36: **return** newRanking

37: **function** h(G, d,ranking)
38: $a \leftarrow 0$
39: $m \leftarrow |E|$
40: $r \leftarrow$ Rank($\bar{V}$,ranking)
41: **for** $e = (u, v) \in E$ **do**
42: $a \leftarrow a + f_d(r(u) - r(v))$
43: **return** $1 - \frac{a}{m}$

44: **function** Rank($V, (\mathcal{S}_1, \ldots, \mathcal{S}_R)$)
45: **for** $\mathcal{S}_i \in (\mathcal{S}_1, \ldots, \mathcal{S}_R)$ **do**
46: $r(v) \leftarrow i\ \forall v \in \mathcal{S}_i$
47: **return** r

2.2.2 Tests

The last step is to test the heuristic on graphs for which the structure is known. We employ once again the RSBM. For brevity we do not present a wide exploration of the parameters space but we choose two representative sets of parameters, one for twitter and one for military hierarchy. We choose triples that from previous Chapter 1 are in the regions beyond detectability for $d = 1$. In both cases we generate a graph of $N = 1024$ nodes from the ensemble and we infer the hierarchical structure from the ensemble using GAH with different values for d(including the extreme values 0 and 1 for comparison, with a step of 0.01). The parameters of the RSBM are $R = 32$, $p = 0.5$, $s = 0.01$, $q = 0.5$ for twitter hierarchy, and $q = 0$ for military hierarchy. We repeat 25 times to control for statistical fluctuation in the graph drawing.

Figure 2.1 shows the number of inferred classes (top) and the adjusted Rand Index (RI) Hubert and Arabie (1985) (bottom) with respect to the planted partition (i.e a measure of the similarity of the two partitions). Similarly to what happens in multi-resolution modularity, the number of inferred classes presents peculiar plateaux, i.e. decreasing d it does not increase with a constant rate but it stays constant for wide intervals and or it has abrupt changes (steps). This can be useful in real application to determine how to choose d. The trend for RI presents critical points in correspondence of the jump in the number of classes, and apart from values very close to 0 is non decreasing. In the case of twitter hierarchy, recovery decisively improves after the first discontinuity (around 0.62) and outperforms the exact solution of $d = 1$. In the case of military, similarity it is lower than the case $d = 1$ with exact partition, even though the number of classes is closer to reality. We believe this is the joint effect of the optimal solution being approximated by the greedy approach and the peculiar and highly symmetric structure of this type of hierarchy.

Figure 2.2 shows the value of h as a function of d. In the case of twitter hierarchy the empirical results (solid blue) follow quite well the analytically estimated optimum (dashed red) found in Section 2.1 and for $d = 1$ it is also very close to value of the exact solution (orange circle). For military hierarchy the behaviour is less straightforward: for $d = 1$ the value from heuristic is comparable to the exact solution, but for small d the solution is quite far from the optimum. Also, d approximately larger than 0.5 the empirical value of h is consistently greater than the estimated optimum, meaning that simply inverting the ranking is just a first approximation of the optimal partition, which is more complex (see Figure 2.4, and Chapter 1 for similar discussion).

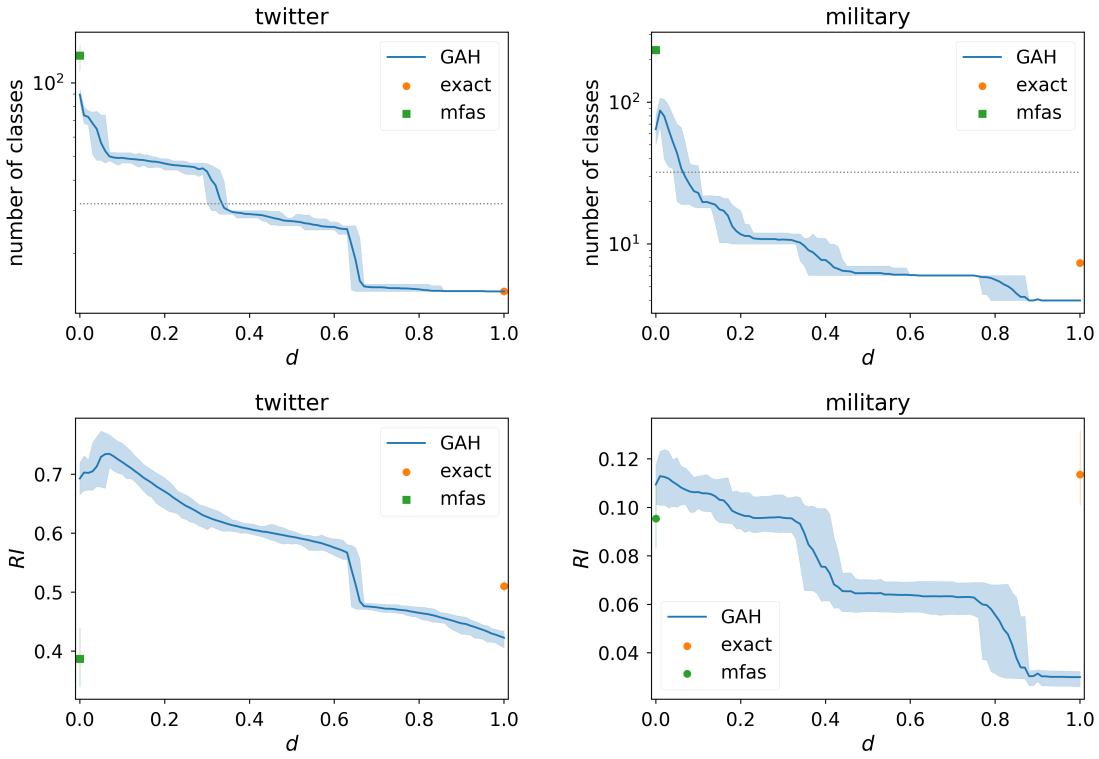


Figure 2.1: Number of classes (top, log scale) and Rand Index (bottom) for different value of d. Left panels refers to twitter hierarchy, right panels to military. The solid line indicates the average results from *GAH* inference, the shaded area indicates the values between 5% and 95% quantiles. The isolated markers indicates the same quantity associated to other algorithms for inference: green square for Minimum Feedback Arc Set ($d = 0$) Eades et al. (1993), and orange circle for the exact solution for $d = 1$ Tatti (2017). The dotted grey lines in top panels, indicated the plangent value for the number of classes (32).

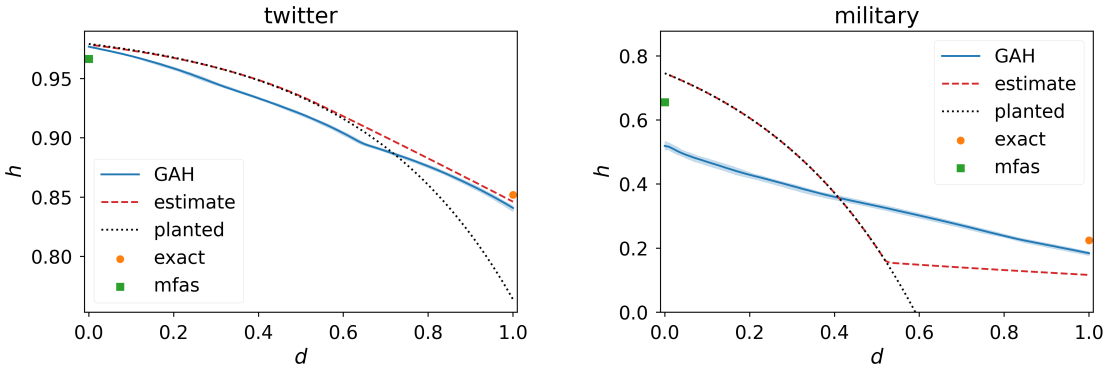
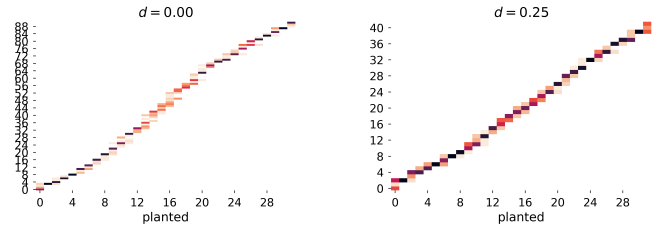
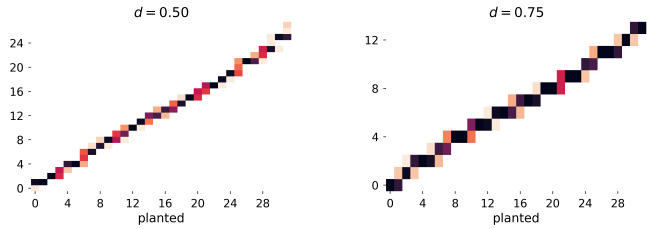


Figure 2.2: Value of h as a function of d for the heuristic (solid blue, and shaded area for values between 5% and 95% quantiles), the analytically estimated optimum dashed red), the planted partition (dotted grey). The isolated markers indicates the same quantity associated to other algorithms for inference: green square for Minimum Feedback Arc Set ($d = 0$), and orange circle for the exact solution for $d = 1$. Left panel refers to twitter hierarchy, right panel to military.

Figures 2.3 (twitter) and 2.4 (military) show the heatmaps comparing the inferred and the planted for a single realisation. For twitter hierarchy, the comparison of the four panels clearly show how classes are progressively split when choosing smaller d. For military hierarchy, as anticipated in previous discussion, the partition inferred by the heuristic presents only a partial inversion of the classes (and also splitting, as in the previous case).

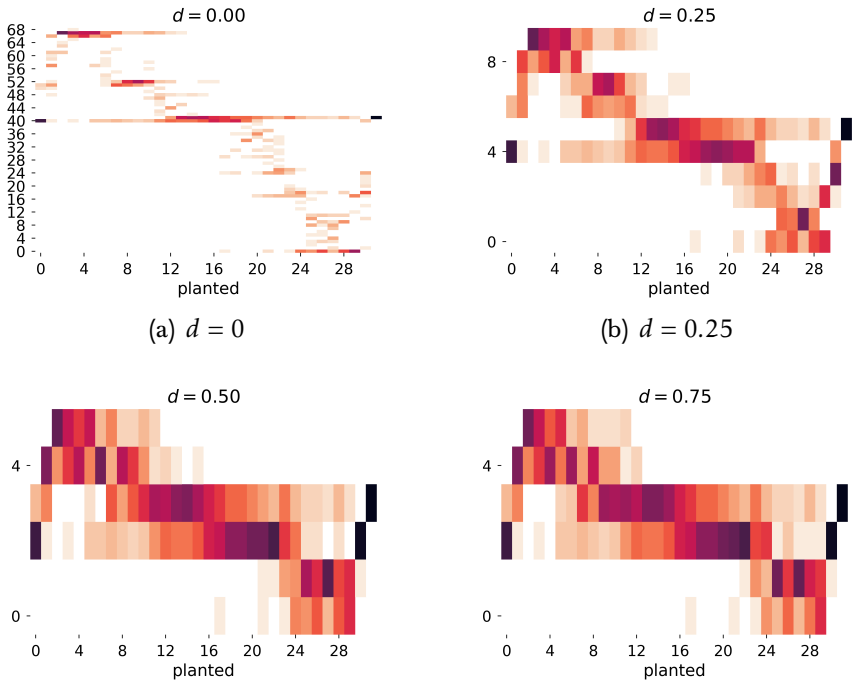


(a) $d = 0$ (b) $d = 0.25$



(c) $d = 0.5$ (d) $d = 0.75$

Figure 2.3: Twitter hierarchy: heatmaps comparing planted and inferred partitions. Each cell c_{ij} indicates the number of nodes that from class j in the planted partition are put in class i by the heuristic. A darker color indicates a higher value.



(a) $d = 0$ (b) $d = 0.25$

(c) $d = 0.5$ (d) $d = 0.75$

Figure 2.4: Military hierarchy: heatmaps comparing planted and inferred partitions. Each cell c_{ij} indicates the number of nodes that from class j in the planted partition are put in class i by the heuristic. A darker color indicates a higher value.

2.3 Conclusions

In this Chapter we characterised the resolution limits for hierarchy detection using agony metric with non integer resolution parameter d and we proposed an greedy heuristic to find the hierarchical partition for varying d.

For twitter-hierarchy, a proposition proves that the resolution limit can be softened by decreasing d, and it gives the critical value of s after which the planted partition is no longer agony-optimal. The test of the heuristic on graph sampled from RSBM are quite satisfying as one can observe both the increasing number of classes inferred for decreasing d, and a value of h close to the optimum from the analytic part. Also the Rand Index indicates the the recovery of the planted partition can improve with $d < 1$.

For military-hierarchy the results are less decisive. A proposition states the critical value of s for inversion, and the estimate for the optimal h, but its validity is quite restrictive. Indeed as for $d = 1$, already present in the literature, from simulation it is clear that the partially inverted partition is better that both the planted and the inverted (so the limit value for s is actually quite optimistic). This is also observable from the test of the heuristic, that consistently finds the partially inverted partition.

From these first two Chapters it is quite evident that the method for hierarchy detection analysed here has both strengths and pitfalls. One one side, with respect to the other methods present in the literature it is able to give more detailed information as it does not condensate into one group nodes belonging to the same strongly connected components, can be applied to both unweighted and weighted networks, it gives a synthetic measure of the overall strength of the hierarchy, and it has an straightforward interpretation. On the other side, the best choice for d that assures polynomial complexity is affected by resolution limitation that can be remarkable when the network is very large. Our proposal for a multi-resolution inference gives a partial solution to this issue but it is inherently affected by the approximations needed to address a problem otherwise computationally unfeasible. As mentioned at the end of the previous chapter one may think to reconsider the cost function in order to trade off between computational feasibility and resolution limits.

Appendix 2.A Proofs

Proof of Proposition 4

Case $0 < d < 1$, $b \leq 0$. For the first part of the proposition , we need to prove that $\exists\, \underline{\epsilon}_{a,b}$ such that $\forall\, d \leq \underline{\epsilon}_{a,b}$

$$2^{2b} \sum_{k=0}^{2^{a-b}-1} (k+1)^{1-d}(2^{a-b}-k) < \sum_{k=0}^{2^{a}-1} (k+1)^{1-d}(2^{a}-k), \tag{2.4}$$

i.e the value of agony for a split ranking is lower than that of the planted ranking, which, then, is not optimal. Assume d satisfies

$$(k+1)^d \leq c_2 \,\forall\, k \in (0, 2^{a-b})$$
$$(k+1)^d \geq c_1 \,\forall\, k \in (0, 2^{a}),$$

which gives

$$c_2 \geq \max_{k\in[0,2^(a-b)]} (k+1)^d = (2^{a-b})^d$$
$$c_1 \leq \min_{k\in[0,2^(a)]} (k+1)^d = 1\,.$$

Then

$$2^{2b} \sum_{k=0}^{2^{a-b}-1} (k+1)^{d}(2^{a-b}-k) < 2^{2b} c_2 \sum_{k=0}^{2^{a-b}-1} (2^{a-b}-k),$$

and

$$\sum_{k=0}^{2^{a}-1} (k+1)^{d}(2^{a}-k) > c_1 \sum_{k=0}^{2^{a-b}-1} (2^{a}-k).$$

Equation (2.4) becomes

$$(c_2 - c_1) < c_1 - 2^b c_2\,.$$

We substitute the limit value fr c_1, c_2 into this last inequality and get the bound for d:

$$d \leq \underline{\epsilon}_{a,b} = \frac{\log \frac{1+2^a}{2^a+2^b}}{(a-b)\log 2}.$$

For the second part, we need to prove that $\exists\, \overline{\epsilon}_{a,b}$ such that $\forall\, d > 1 - \overline{\epsilon}_{a,b}$

$$2^{2b} \sum_{k=0}^{2^{a-b}-1} (k+1)^{1-d}(2^{a-b}-k) > \sum_{k=0}^{2^a-1} (k+1)^{1-d}(2^a-k), \tag{2.5}$$

i.e any split ranking has a value of agony higher than the planted, so they are not optimal. We proceed as in the previous case, and assume

$$(k+1)^{-d} \leq c_2 \,\forall\, k \in (0, 2^a)$$
$$(k+1)^{-d} \geq c_1 \,\forall\, k \in (0, 2^{a-b}),$$

which implies

$$c_2 \geq \max_{k\in[0,2^a]} (k+1)^{-d} = 1$$
$$c_1 \leq \min_{k\in[0,2^{a-b}]} (k+1)^{d} = (2^{a-b})^{-d}.$$

Substituting into (2.5) one gets the bound for d

$$d \geq 1 - \overline{\epsilon}_{a,b} = 1 - \frac{\log \frac{(2^a+2^b)(2^a+2^{1+b})}{2^b(1+2^a)(2+2^a)}}{(a-b)\log 2}.$$

Proof of Proposition 5

$0 < d < 1$. When $q = 0$, the critical value of s for the planted ranking to be positive is given by

$$s_{d,1} = \frac{2\,(2^a-1)\,p}{2\sum_{k=0}^{2^a-1}(2^a-k)\,(k+1)^d - 2^a - 4^a}.$$

The estimate for the value of hierarchy for the inverted rank is

$$h^i(b;p,s,a) = \frac{2^{-b}\left(2^a-2^b\right)\left(s2^{a+b}-2\left(2^d-1\right)p\right)}{2^a(2p+s)+4^a s-2p}$$

from which one can compute the optimal number of classes

$$\tilde{R}_d^{i,*} = \begin{cases} 1 & s \leq s_{d,2}^i \\ 2 & s_2^i < s < s_{d,3}^i \\ 2^{a-b_d^{i,*}} & s > s_{d,3}^i, \end{cases}$$

where

$$b_d^{i,*} = \frac{1}{\log(4)} \log\left(\frac{2\left(2^d - 1\right)p}{s} \right),$$

and

$$\begin{aligned} s_{d,2}^i &= \frac{4}{2^{2a}}\left(2^d - 1\right)p \\ s_{d,3}^i &= 3s_{d,2}^i. \end{aligned}$$

Since $s_{d,2}$ is increasing in d and $s_{d,m}$ is decreasing in d, and (1.14), it holds

$$s_{d,1} \geq s_{d,2}^i.$$

Hence, one denotes the critical value for inversion $s_{d,i}$ as the value of s for which

$$\bar{h}_d^i(b = a - 1) = \bar{h}_d(b = 0),$$

that is

$$s_{d,i} = \frac{4p\left(2^a + 2^d - 2\right)}{4\sum_{k=1}^{2^a-1}(2^a - k)(k+1)^d - 2^a(2^a - 2)}$$

Finally we have to check that the critical value satisfies the bound in (1.6), which, in this case is equivalent to verify that

$$\sum_{k=1}^{2^a-1}(1+k)^d(2^a - k) \geq \frac{2^d 2^a(2^a+1)}{2}.$$

For the first part of the proposition, we show that $\exists\, \underline{\epsilon}_a$ such that $\forall\, d \leq \underline{\epsilon}_a$

$$\sum_{k=1}^{2^a-1}(1+k)^d(2^a - k) < \frac{2^d 2^a(2^a+1)}{2}. \tag{2.6}$$

If we denote

$$\begin{aligned} c_1 &= \min_{k\in[1,2^a-1]}(k+1)^d = 2^d \\ c_2 &= \max_{k\in[1,2^a-1]}(k+1)^d = 2^{ad} \end{aligned}$$

and substitute into inequality (2.6) we get the critical value $\underline{\epsilon}_a$

$$\underline{\epsilon}_a = \frac{\log \frac{2^a+1}{2^a-1}}{(a-1)\log 2}.$$

For the second part we show $\exists\, \overline{\epsilon}_a$ such that $\forall\, d = 1-\epsilon \geq 1-\overline{\epsilon}_a$

$$\sum_{k=1}^{2^a-1}(1+k)^{1-\epsilon}(2^a-k) \geq 2^{-\epsilon}2^a(2^a+1). \tag{2.7}$$

We denote

$$c_1 = \min_{k\in[1,2^a-1]}(k+1)^{-\epsilon} = 2^{-a\epsilon}$$
$$c_2 = \max_{k\in[1,2^a-1]}(k+1)^{\epsilon} = 2^{\epsilon}$$

Substituting into inequality (2.7) and rearranging one gets

$$\overline{\epsilon}_a = \frac{\log \frac{(2^a+4)(2^a-1)}{2(2^a+1)}}{(a-1)\log 2}.$$

Proof of Proposition 6

$0 < d < 1$. Similarly to the case $d = 1$, the critical value $s_{d,m}$ is the value of s such that

$$\bar{h}_d(b=0) = \bar{h}_d\left(b = a - \frac{\log(2^a-1)}{\log 2}\right).$$

To show that f_s is non-negative and strictly increasing for $d \geq 0$ we compute its derivative with respect to d and show it is bounded from below by a strictly positive function.

$$\frac{\partial f_s}{\partial d} = \sum_{k=1}^{2^a-2}(k+1)^d \log(k+1)c_k + a2^{ad}\log 2,$$

where

$$c_k = k\frac{2^{a+1}-1}{(2^a-1)^2} - \frac{2^a}{2^a-1}.$$

It easy to verify that

$$\begin{cases} c_k < 0 & k \leq 2^{a-1} - 1 \\ c_k > 0 & k \geq 2^{a-1}, \end{cases}$$

so we split the sum accordingly and isolate the factor with d

$$\frac{\partial f_s}{\partial d} \geq (2^{a-1})^d \sum_{k=1}^{2^a-2} \log(k+1) c_k .$$

It holds

$$\begin{aligned} \sum_{k=1}^{2^a-2} \log(k+1) c_k = & \frac{2^{a+1}\left(2^{a-1}\right)^d \zeta^{(1,0)}(-1, 2^a)}{(2^a-1)^2} - \frac{\left(2^{a-1}\right)^d \zeta^{(1,0)}(-1, 2^a)}{(2^a-1)^2} + \\ & + \frac{2^{a+1}\left(2^{a-1}\right)^d \log(\mathcal{G})}{(2^a-1)^2} - \frac{\left(2^{a-1}\right)^d \log(\mathcal{G})}{(2^a-1)^2} \\ & + \frac{\left(2^{a-1}\right)^d}{12(2^a-1)^2} - \frac{\left(2^{a-1}\right)^{d+1}}{3(2^a-1)^2} + \\ & - \frac{2^a\left(2^{a-1}\right)^d \log(\Gamma(2^a))}{(2^a-1)^2} - \frac{4^a\left(2^{a-1}\right)^d \log(\Gamma(2^a))}{(2^a-1)^2} + \\ & + \frac{\left(2^{a-1}\right)^d \log(\Gamma(2^a))}{(2^a-1)^2}, \end{aligned}$$

where $\mathcal{G}$ is the Glaisher constant, $\zeta^{(1,0)}(\sigma, \alpha)$ is the derivative of the generalized zeta function with respect to σ, and Γ is the gamma function. The right hand side is a strictly positive function when $a \geq 3$ so one can conclude

$$\begin{aligned} \frac{\partial f_s}{\partial d} > 0 \quad & \forall\, a \geq 3, \forall\, d \geq 0 \\ f_s \geq 0 \quad & \forall\, d \geq 0 \quad \text{and} \ = 0 \quad \text{iff } d = 0. \end{aligned}$$

Proof of Proposition 7

In order to prove that $s_{m,d}$ is the resolution limit, it is sufficient to analyse the behaviour of $h^{(b)}$ with respect to splitting (because merging is characterised by

proposition 6). In particular, we show that any split ranking has a higher agony than the planted for a suitable choice of d and R.

We proceed by induction on the number of classes in the split ranking:

$$\mathcal{P}(n) : \mathbb{E}[A_d(\tilde{R} = R + n)] \geq \mathbb{E}[A_d(\tilde{R} = R)].$$

To simplify notation, denote

$$s\bar{A}_d^{(n)} = \mathbb{E}[A_d(\tilde{R} = R + n)].$$

Case $n = 1$. $\mathcal{P}(1)$ is equivalent to

$$\begin{aligned} \bar{A}_d^{(1)} - \bar{A}_d^{(0)} &\geq 0 \\ \left(\frac{R}{R+1}\right)^2 \sum_{k=0}^{R} (k+1)^d (R+1-k) - \sum_{k=0}^{R-1} (k+1)^d (R-k) &\geq 0 \\ \sum_{k=0}^{R-1} (k+1)^d c_k + \left(\frac{R}{R+1}\right)^2 (R+1)^d &\geq 0, \end{aligned} \tag{2.8}$$

where

$$c_k = k\frac{2R+1}{(R+1)^2} - \frac{R}{R+1}.$$

It holds

$$\begin{cases} c_k < 0 & k \leq \frac{R}{2} \\ c_k > 0 & k > \frac{R}{2}. \end{cases}$$

One can split the sum in Eq. (2.8) at $k^* = \frac{R}{2}$, assuming R is even, in order to find a lower bound for both parts.

$$\sum_{k=0}^{k^*} (k+1)^d c_k \geq \left(\frac{R}{2}+1\right)^d \sum_{k=0}^{k^*} c_k = -\left(\frac{R}{2}+1\right)^d \frac{R(R+2)(2R+3)}{8(1+R)^2},$$

$$\sum_{k=k^*+1}^{R-1} (k+1)^d c_k \geq \left(\frac{R}{2}+2\right)^d \sum_{k=k^*+1}^{R-1} c_k = \left(\frac{R}{2}+2\right)^d \frac{R(R-2)(2R-1)}{8(1+R)^2}.$$

It follows,

$$A_d^{(1)} - A_d^{(0)} \geq f(d, R),$$

where

$$f(d,R) = -\left(\frac{R}{2}+1\right)^d \frac{R(R+2)(2R+3)}{8(1+R)^2} + \left(\frac{R}{2}+2\right)^d \frac{R(R-2)(2R-1)}{8(1+R)^2} + (R+1)^d \left(\frac{R}{R+1}\right)^2.$$

So a sufficient condition for $\mathcal{P}(1)$ to be true is $f(d,R) \geq 0$.

Denote by $d^* = d^*(R)$ the value of d such that

$$f(d^*,R) = 0. \tag{2.9}$$

Clearly, there is no close formula for d^*, however Eq. (2.9) can be solved numerically (see figure 2.5 for numerical results) and it is possible to characterise the solution.

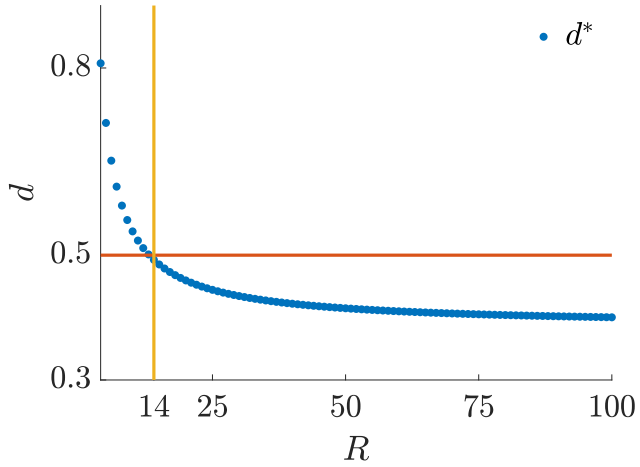


Figure 2.5: Critical value for d as R changes.

At first we note that, for fixed R, f can be rewritten as

$$f(d;R) = -\alpha^- f^-(d) + \alpha^+ f^+(d),$$
$$\alpha^-, \alpha^+ > 0,$$
$$f^-(d), f^+(d) > 0 \,\forall d.$$

Moreover f^-, f^+ are stricly monotone increasing, hence f has at most one zero.

One can easily verify that

$$\begin{cases} f(0,R) < 0 & \forall R \\ f(1,R) > 0 & \forall R > 2, \end{cases}$$

and f is continuous in d, so Bolzano's theorem applies.

Summarising, for any $R > 2$, d^* exists and it is unique, moreover $0 < d^* < 1$.

It holds

$$f(0.5,R) > 0 \quad \forall\, R \geq 14.$$

The same reasoning can be done when R odd, in that case the sum is split at $k^* = \frac{R-1}{2}$.

We can conclude that $\mathcal{P}(1)$ is true $\forall R \geq 14 \,\forall\, 0.5 \leq d \leq 1$.

Now, assume $\mathcal{P}(n-1)$ true. We need to prove that this implies $\mathcal{P}(n)$ true $\forall\, n \geq 2$. $\mathcal{P}(n)$ rewrites as

$$\begin{aligned} A_d^{(n)} - A_d^{(0)} &\geq 0 \\ A_d^{(n)} - A_d^{(n-1)} + A_d^{(n-1)} - A_d^{(0)} &\geq 0 \end{aligned}$$

From the inductive hypothesis follows $A_d^{(n-1)} - A_d^{(0)} \geq 0$, so to prove $\mathcal{P}(n)$ true it is sufficient to show that $A_d^{(n)} - A_d^{(n-1)} \geq 0$.

We proceed similarly to the case $n = 1$, obtaining

$$A_d^{(n)} - A_d^{(n-1)} = R^2 \sum_{k=0}^{R+n-2} (k+1)^d c_k^{(n)} + \left(\frac{R}{R+n}\right)^2 (R+n)^d,$$

where

$$c_k^{(n)} = \frac{R+n-k}{(R+n)^2} - \frac{R+n-1-k}{(R+n-1)^2}.$$

It holds

$$\begin{cases} c_k^{(n)} < 0 & k \leq \lfloor \frac{R+n}{2} - 1 \rfloor \\ c_k^{(n)} > 0 & k \geq \lfloor \frac{R+n}{2} \rfloor \end{cases}$$

As before, we only show the explicit computations for even $R+n$, but the same results holds also when $R+n$ odd, it is sufficient to opportunely adjust the limits of the sums.

$$
\begin{aligned}
A_d^{(n)} - A_d^{(n-1)} &= R^2 \sum_{k=0}^{\frac{R+n}{2}-1} (k+1)^d c_k^{(n)} + R^2 \sum_{k=\frac{R+n}{2}}^{R+n-2} (k+1)^d c_k^{(n)} + \left(\frac{R}{R+n}\right)^2 (R+n)^d \\
&\geq R^2 \left(\left(\frac{R+n}{2} - 1\right)^d \sum_{k=0}^{\frac{R+n}{2}-1} c_k^{(n)} + \left(\frac{R+n}{2}\right)^d \sum_{k=\frac{R+n}{2}}^{R+n-2} c_k^{(n)} + \frac{(R+n)^d}{(R+n)^2} \right) \\
&\geq R^2 g(d, R, n).
\end{aligned}
$$

The sums in g can be computed analytically, moreover, for fixed R, n, g can be rewritten as

$$
\begin{gathered}
g(d; R, n) = -\beta^- g^-(d) + \beta^+ g^+(d), \\
\beta^-, \beta^+ > 0, \\
g^-(d), g^+(d) > 0 \, \forall d \, .
\end{gathered}
$$

Moreover g^-, g^+ are strictly monotone increasing, hence g has at most one zero.

One can easily verify that

$$
\begin{cases}
g(0, R, n) < 0 & \forall R \, n \\
g(1, R, n) > 0 & \forall R > 2\, n,
\end{cases}
$$

and g is continuous in d, so Bolzano's theorem applies. Furthermore, it holds

$$
g(0.5, R, n) > 0 \quad \forall \, R \geq 14 \, \forall \, n \geq 2 \, .
$$

This conclude the induction step and the proof.

3

Controlling financial instability: a fast network strategy

In this final Chapter of Part I, we propose a way to exploit the hierarchical structure of a financial network in order to control its stability. The starting point is the work by (Bardoscia et al., 2017), where authors address the problem of identifying financial stability without relying on a specific contagion model over the financial network. Instead, authors propose the leading eigenvalues of the matrix describing the interbank network as measure of instability.

Here we propose a mechanism based on the knowledge of the hierarchical structure to control the evolution of the financial network with the aim to prevent, or at least slow down, the evolution towards network structures which are particularly prone to transmit distress. After describing the strategy, we show its applicability by using real data of interbank exposures.

The Chapter is organised as follows. In Section 3.1 we present the relevant literature. In Section 3.2 we propose our strategy to control the evolution of the network and we show its application to real data of interactions between banks. In Section 3.3 we draw conclusions.

3.1 Literature review

As mentioned in the introduction, there is a well established evidence (Haldane and May, 2011; Battiston et al., 2016) that the interconnectedness of financial institutions is one of the sources of financial instability, as it may cause the propagation of distress from one institution to its neighbours. Current policies tend to the implementation of capital requirements on the financial institutions that usually do not take into account the vulnerability of a single entity coming from the presence of network interaction and viceversa the contribution of each institution to the risk at system level, moreover they do not address directly the building up of such risk.

A seminal paper which address this is (Acharya et al., 2009). Authors claim the need for a regulation that induce financial institution to *internalise* the cost of systemic risk arising from its operation as it act as negative externality on the system. They propose a measure for systemic risk and also suggest a strategy for regulations to limit it. The idea is to quantify the contribution of each institution to a systemic loss a to impose them to buy an insurance to cover their own losses, whose beneficiary is the regulator in charge of stabilising the financial sector. This would provide incentives for a company to limit systemic risk, provide a market-based estimate of the risk and avoid moral hazard.

(Acharya et al., 2012) goes in the same direction of quantifying the contribution to general loss, but authors focus on a specific type of loss, capital shortfall, and propose an econometric approach instead of a network one. A systematic ex-post analysis of the contribution of systemic loss by financial institutions in US and Europe using those tools is presented in (Acharya et al., 2013; Acharya and Steffen, 2013), where it is clear that at the time of the crises both propagation modelling and data collection was not extensive enough for a prompt regulatory response.

In the aftermath of the crisis many models have been proposed to estimate, by simulation, the danger of contagion derived from exposures in the interbank loan market, and network tools have been largely at use for this purpose. An overview is presented in (Upper, 2011). The paper summarises the results as well as the assumptions on which the models are based, and discusses their use in financial stability analysis. On the whole, results suggest that contagious defaults are unlikely but cannot be fully ruled out, at least in some countries. Robustness tests indicate that the models might be able to correctly predict whether or not contagion could be an issue and, possibly, also identify banks

whose failure could give rise to contagion.

In (Gauthier et al., 2012) authors extend the idea of capital requirements that cover a bank's contribution to the whole system's risk by including capital reallocation as response to a extreme event. They employ a network based structural model to measure systemic risk and how it changes with reallocation. The model is then tested on a sample of Canadian banks, observing that the capital levels, are not trivially related to bank size or individual bank default probability.

Poledna and Thurner (2016) extend the idea further, by quantifying the contribution to overall risk of each transaction. Also in this case, authors rely on network formalisation and propose a tax on individual transactions that is proportional to their marginal contribution to overall risk. The tax would allow to reshape the network avoiding concentration of risk, but the implementation in practice would need a availability of granular data.

3.2 Controlled dynamics

From the previous chapters it is clear that the presence of cycles disrupts the hierarchical structure of a network, and thanks to the metric h one is able to condense the information on the extent of the impact in one quantity.

On the other side, as it has been pointed out in Battiston et al. (2016) that the presence of cycles in a network may cause instability. In several works, see for example Gai and Kapadia (2010); Acemoglu et al. (2012); Georg (2013) and related literature, this effect has been highlighted by defining contagion models and simulating or analytically quantifying the propagation of default/distress from a single (or small group) to the entire network. Even though this approach can give interesting insight, the interpretation is model dependent, so another approach that has been proposed in Caccioli et al. (2014) is to look at the leading eigenvalue of the adjacency matrix. A detailed description of this mechanism is beyond the scope of this chapter, but it is based on the observation that the propagation of contagion can be opportunely expressed in first linear approximation as an iterative application of the map described by the adjacency matrix, therefore the results regarding branching processes apply, see Mode (1971) and related literature. Indeed, it can be shown that condition on the leading eigenvalue $|\lambda_{\max}| > 1$ is sufficient and necessary to have instability.

Bardoscia et al. (2017), build on this approach to study the evolution of financial networks, described by the leverage matrix. In particular, authors

show that the issuing of new contracts leads any stable network to become unstable after enough new links are added to the network. The starting point of the evolution in that case was a DAG, which means that for the leading eigenvalue of the adjacency matrix it holds $\lambda_{\max} = 0$ and the value of hierarchy is $h^* = 1$, and the dynamic carried on until obtaining a complete network by successively adding links to simulate diversification.

Here, we propose a strategy based on the knowledge of the hierarchical structure of the network, such that it is possible to control the emergence of cycles by choosing counterparts according to the respective position in hierarchy. Note that, in a real implementation, only the control authority is likely to have enough data to build the graph (and therefore of the rank of the counterparts).

To fix ideas, in the following we describe our proposed controlled evolution of the network. As in Bardoscia et al. (2017) at each time step a new contract is issued, hence a new link is added to the network.

Suppose at time $t = k-1$, the network has adjacency matrix $A^{(k-1)}$, it is connected, with $m^{(k-1)} = \sum_{i,j} A^{(k-1)}_{i,j}$ links and it has optimal ranking $r^{(k-1)}$, and value of hierarchy $h^{*(k-1)}$. At the next time step $t = k$, a new contract is issued between counterparts i and j. The resulting link can be either forward or backward with respect to $r^{(k-1)}$. Consider the two cases separately. If the link (i, j) is forward, i.e. $r^{(k-1)}(i) < r^{(k-1)}(j)$, the new link is accepted, and the new value of h is given by

$$
\begin{aligned}
h' &= 1 - \frac{1}{m^{(k-1)}} \min_{r \in \mathcal{R}} \sum_{r(u)>r(v)} A^{(k)}_{u,v}(r(u) - r(v) + 1) \\
&\leq 1 - \frac{1}{m^{(k-1)}} \sum_{r^{(k-1)}(u)>r^{(k-1)}(v)} A^{(k)}_{u,v}(r^{(k-1)}(u) - r^{(k-1)}(v) + 1) \\
&= 1 - \frac{1}{m^{(k-1)}} \sum_{r^{(k-1)}(u)>r^{(k-1)}(v)} A^{(k-1)}_{u,v}(r^{(k-1)}(u) - r^{(k-1)}(v) + 1) \\
&= 1 - \frac{(1 - h^{*(k-1)})m^{(k-1)}}{m^{(k-1)} + 1}.
\end{aligned}
$$

Indeed equality holds: suppose it exists a ranking $\tilde{r}$ such that $h(\tilde{r}, A^{(k)}) < h(r^{(k-1)}, A^{(k)})$ then it will also hold $h(\tilde{r}, A^{(k-1)}) < h(r^{(k-1)}, A^{(k-1)})$ as $\{A^{(k)}_{u,v}\}_{u,v} \setminus \{A^{(k)}_{u,v}\}_{u,v} = \{(i, j)\}$ and we assumed (i, j) is a forward link for $r^{(k-1)}$ so it does not contribute to h. This means the the ranking $r^{(k-1)}$ stays optimal at time

$t = k$. Hence $h^{*(k)} = h'$, and it holds $h^{*(k-1)} \geq h^{*k}$.

If in addition the network is a DAG, one has the certainty that no loop is created. If backward links are already present, then for the probability to create a loop adding the link (i, j) to the network it holds (we dropp the superscript $(k-1)$ for readability)

$$\begin{aligned}
\mathbb{P}(\text{loop}) &= \mathbb{P}(\exists \text{ path from } j \text{ to } i \text{ in } A \,|\, r(i) < r(j)) \\
&= \sum_{l=1}^{r(j)-r(i)} \mathbb{P}(\exists \text{ path from } j \text{ to } i \text{ of length } l \text{ in } A \,|\, r(i) < r(j)) \\
&= \sum_{l=1}^{r(j)-r(i)} \mathbb{P}(l \text{ adjacent backward links}) \\
&= \sum_{l=1}^{r(j)-r(i)} \frac{m}{n^2}(1-h^*)^k \\
&= \frac{m}{n^2}\frac{1-(1-h^*)^{l-1}}{h^*}.
\end{aligned}$$

If the link (i, j) is backward in the ranking $r^{(k-1)}$, then our strategy tries to replace the counterpart j, assuming that it is i initiating the transaction, with another counterpart $\tilde{j}$ such that the link $(i, \tilde{j})$ is forward. If such $\tilde{j}$ does not exist, for example if i belong to a class with high rank in the hierarchy, one has two choices: either to make the transaction to fail or to allow for a backward link. In the first case, if no link is removed, the nodes in the upper classes would have small or no possibility to issue new contracts and the system will soon freeze. In the second case one allows for the build up of fragility but the system continues to work. In this scenario, one has to be careful on the choice of counterparts, whether $\tilde{j}$ exist or not, for two reasons. First, even if the addition of a backward link do not necessarily implies the creation of a loop it increases the probability for a loop in a successive iteration. For this reason an heuristic way would be to choose a counterpart close in rank so that long loops are avoided. Secondly, the new counterpart need to have sufficient funding for the new contract. In other words, one would like to avoid that re-distributing the links, one entity moves from a short to a long position when netting all the outstanding contracts. Note that, choosing a counterpart that is positioned in a higher rank when rewiring (i.e. trying to have a forward link) partially avoids this issue, because nodes in high rank in the hierarchy are already characterised

by an excess of incoming links with respect to outgoing links[1].

At this point it is clear that this procedure will eventually result into an unstable system as in the uncontrolled evolution, but this will occur after a higher number of contracts (links) are issued. In other words for the same level of density, the controlled network has a lower level of instability measured by the leading eigenvalue. [2]

One by-product of this scheme is the possibility to monitor the instability by keeping track of only a scalar quantity, which is relatively fast to compute. The reasons to choose the hierarchy to implement the monitoring are several. First, the algorithm we employ is linear in the size of the network, this allows real time feedback for new contract, even if the network is large. This is an advantage with respect to other method proposed such as (Poledna and Thurner, 2016). Secondly, even if the eigenvalue is more precise in quantifying the instability, the computation is more intense then the hierarchy because its has complexity quadratic in the size of the network. Furthermore, algorithms for eigenvalue computation can present numerical issues, even for a small number of nodes, when the network is particularly sparse, which is the case for many real financial networks. Also, with the eigenvalue alone, one could perform the tracking of instability but it is less clear how to suggest alternative counterparts without generating several networks and selecting the one with lower leading eigenvalue (which would imply further computational burden). The other reason is the small amount of data needed to compute the ranking. Here we employ the unweighted version so there is no need for the counterparts to disclose the amount of their transactions and it works also with partial information on the network of interactions. This last point highlights that this method is complementary not alternative to capital requirements tailored to contribution to global risk as proposed in (Acharya et al., 2009).

To clarify the applicability of this approach we propose an empirical exercise using real data of interaction between banks. The data are from E-MID platform for interbank lending. This platform started in 1999 and it is mostly employed for overnight deposit especially by Italian banks (Iori et al., 2008) and less intensively from other EU banks. The dataset contains daily transaction, each with anonymised lender and borrower, amount and expiration. We choose to restrict the analysis to the year 2011 following (Barucca and Lillo, 2018) for

[1]This feature of the optimal rank has already been discussed in Chapter2

[2]Following Achlioptas et al. (2009) one may argue that even if the building up of instability is delayed, once it reached the structure is prone to larger propagation, but the delay would allow in case more time to intervene with other measures.

two reasons. First, authors found that after 2011 volume traded on E-MID declined due to the sovereign debt crisis and ECB extraordinary monetary policies; secondly, in that year authors are able to clearly detect two communities that act as borrowers and lenders respectively. This is particularly interesting for our application because it means that there is a dominant direction for the money in the system, which is an indication of the presence of a hierarchical structure.

To perform the simulation of the controlled dynamic we update the network at each new transactions and we eliminate the link corresponding to expired loans at the end of each day (or we adjust the weights if multiple transactions with different expiration result in the same link). This is different from (Bardoscia et al., 2017), where authors assume to be far from expiration and hence never remove links. We consider two controlled dynamics: in one case we infer the rank at the end of each day and use the rank in the next day accordingly. In the second case we update the hierarchy after each transaction. As authors note in (Barucca and Lillo, 2018), there is high persistence in the preferred side for banks (at least up to 2011), so we use the first days of data to build the network and obtain the optimal hierarchy. This will help avoiding that, while rewiring, entities change significantly the their role in the market.

In (Bardoscia et al., 2017), the leverage matrix defined as exposure (from the single contract) over equity is employed. We do not have the information on the balance sheet, as the dataset is anonymised, so we employ the median sum of the daily incoming and outgoing volume over the entire year as a proxy for the size of the liabilities of each bank, as it has been observed in (Bargigli et al., 2015) that large banks are also more active in the short term lending. Then, following (Wolski and van de Leur, 2016) we compute the normalised weights using the fact that interbank lending represents less than 1% of the liabilities and that the leverage ratio is bounded to be $> 3\%$ (Bucalossi and Scalia, 2016)[3].

In the following figures 3.1,3.2, we show the leading eigenvalues and the value of h for each network snapshot, for both the real dataset (blue) and the controlled one (orange). We do not put any budgetary constraint on the choice of the counterpart, to avoid that an entity pass from net borrower to net lender, nonetheless, this occurs very rarely, only to less than 3% of nodes. The first observation is that the value of h is very close to 1 for all the networks built

[3]It is worth noticing that, even though the leverage has crucial role in the propagation of shocks, especially with respect to amplification, for the strict purpose of the control it has limited effect as in the proposed strategy the direction of the contract drives the acceptance of the link, not its weight.

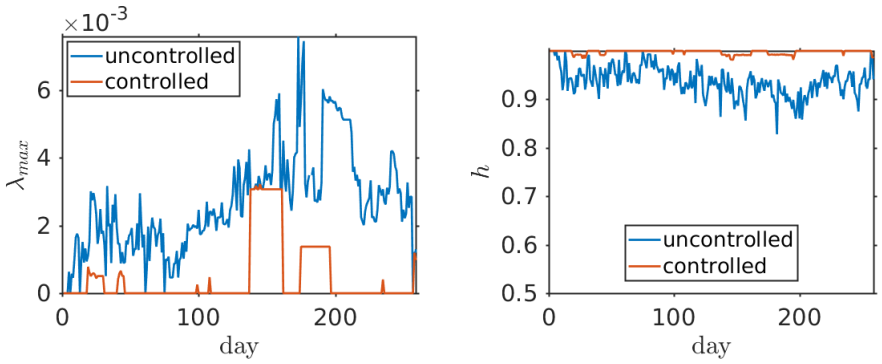


Figure 3.1: Evolution of the leading eigenvalue $\lambda_{\max}$ (top) and hierarchy h, for real data (blue) and controlled (orange) with update of the optimal ranking once a day.

from the original dataset. This is consistent with the aforementioned bipartite structure highlighted in (Barucca and Lillo, 2018). Secondly, we note that the value of $\lambda_{\max}$ is quite low in general, but there is a large interval (from day 80 to day 180 approximately) where the trend is decisively increasing, signalling a build up of instability coming from the complexity of the topology.

The second observation is that as requested we are able to effectively control the eigenvalue, so that for a large part of the year the controlled graph results into a DAG ($\lambda_{\max} = 0$) in the case of one daily update and always for the case of immediate update after each transaction. The difference is explained by the fact the in the first case the optimal hierarchical partition is inferred after removing the links corresponding to expired loans, hence in a less dense, and possibly less representative, network. An alternative would be to compute a the partition on a network, which keep track of previous interactions, but this is beyond the scope of this exercise.

3.3 Conclusions

In this conclusive Chapter of Part I we showed a way to leverage the knowledge of the hierarchical organisation of a interbank network in order to control its evolution and avoid the build up of instability. The effort is consistent with the arising awareness that the control of systemic risk in finance cannot disregard the importance of the complex collections of interactions among financial

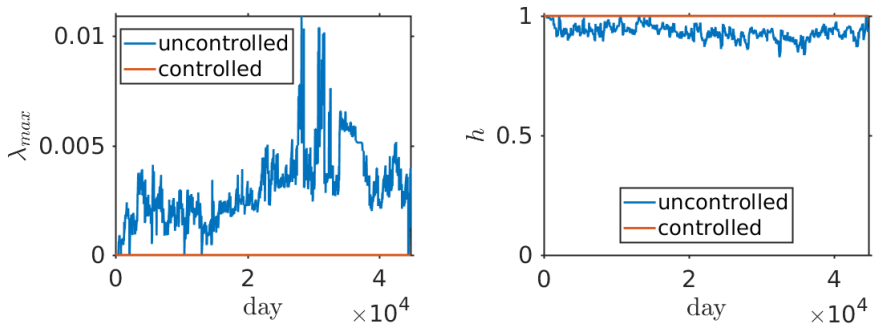


Figure 3.2: Evolution of the leading eigenvalue $\lambda_{\max}$ (top) and hierarchy h, for real data (blue) and controlled (orange) with update of the optimal ranking after each transaction.

institutions.

Instead of relying on a specific model for distress contagion, following existing literature we used the leading eigenvalue as a measure of instability. We consider a dynamic scenario, where new contracts between banks induce new links in the interbank network. The eigenvalue increases as the complexity of the network increases, likewise the presence of cycles disrupt the hierarchy and the metric h, introduced in previous chapters, decreases. Hence, controlling the choice of counterparts of contracts allow to limit the presence of cycles. This is done by favouring connection that follow the hierarchy, and forcing participants, if possible, to change the counterparts otherwise. By applying the protocol to a real dataset of interaction among banks, we show that the proposed strategy effectively succeed in limiting the growth of the eigenvalue, without decreasing the overall liquidity of the system.

A possible extensions would be to improve the choice of the replacing counterparty in order to consider budgetary or liquidity constraint and make the dynamic more realistic. Future developments could consist in embedding this strategy in a more general model for a Systemic Risk Tax, where transactions that increases the overall instability, i.e those corresponding to backward link in the hierarchy, are taxed accordingly to the increased riskiness of the system.

Part II

Risk

Part of the content of this Chapter is available as preprint in Letizia, E. and Lillo, F. (2017). Corporate payments networks and credit risk rating. arXiv preprint arXiv:1711.07677.

4

Corporate payments networks and credit risk rating

We study a system of firms from a network perspective, and more specifically, we investigate the interplay between the risk of firms and the interlinkages connecting them. The network is built from a large proprietary dataset provided by a major European bank. The dataset contains the payments collected at daily granularity between more than two million Italian firms together with the information on internal risk rating for a large fraction of them.

First, we investigate the topological properties of payment networks by considering standard network metrics, such as degree and strength distribution and components decomposition. We find that the large payment networks investigated here share the properties observed in other complex networks, namely they are sparse but almost entirely made of a single component, they are scale free and small world. Then, we look into the distribution of risk of firms in the network of payments in order to quantify the dependence between the network property of a node or a group of nodes and the risk of the firm represented by the node(s). We find an homophily of risk, i.e. the tendency of a firm to interact with firms with similar risk. This is a two nodes properties, but a similar behaviour is observed, even more clearly, also at larger aggregation scales. Communities of firms, detected by using different methods, often display a statistically significant abundance of firms of a specific risk class, indicating the tendency of firms with similar rating to be linked together through payments.

Risk is therefore not spread uniformly on the network, but rather it is concentrated in specific *areas*. The last contribution, is to exploit this correlation between risk of a firm and network characteristics of the corresponding node to predict the risk rating of the firm using network properties *alone*. To this end, we employ machine learning techniques to build classifiers for risk rating whose inputs are only network properties (e.g. degree, community, etc.). We show that our classification method has a good performance both in terms of accuracy and of recall and that outperforms significantly the random assignment (which is the natural benchmark as the data usually employed for risk assessment are not available to us).

The Chapter is organised as follows. In Section 4.1 we present the relevant literature. In Section 4.2 we describe the data, how to construct the network of payments, and its main topological characteristics. In Section 4.3 we study how the risk of firms is spread in the nodes of the networks (i.e. the firms), with a specific attention to the similarity or complementarity of risk profile of firms which are close in the network of payments. In Section 4.4 we use methods from machine learning to predict missing ratings from the network metrics studied before. In Section 4.5 we draw conclusions.

4.1 Literature review

Being credit risk rating a central task of banks and rating agencies, the literature regarding this topic is very broad and is deeply influenced by the regulators' demands over the past decades (on Banking Supervision, 2000; Estrella and on Banking Supervision, 2000; Santos, 2001; on Banking Supervision, 2003). For a review of the literature on credit risk modelling see for example (Altman and Saunders, 1997). On range of models actually employed we refer to the seminal paper (Treacy and Carey, 2000), where authors provide a comparative review of models for rating in use at major US Banks. (Crouhy et al., 2000, 2001) also provide the prototypical features of most used models in the field, while (Lopez and Saidenberg, 2000) presents a strategy to evaluate the ability of those model to predict bankruptcy, which is of particular interest for the regulators. In (Lehmann and Neuberger, 2001) authors highlight that for SME, credit rating may be affected by variable that are not necessarily measurable such as social interaction between the management of the firm and loans officers. More recently there has been an effort to improve the standard models under several aspects: in (Zhang et al., 2016) authors include dynamic incentives for

the growth of firm to be evaluated, while (Lohmann and Ohliger, 2017) includes more economic ratios in the evaluation of the rating. Finally, (Berg and Koziol, 2017) highlighted that there is an issue of consistency between model adopted in different banks, with predicted probabilities of default that may differ significantly.

What is clear from this brief overview is that the role of the interactions among firms has been so far largely disregarded. This despite the fact that financial networks found recently a widespread use. A fundamental work in this perspective is (Acemoglu et al., 2012), where the presence of inter-sectoral linkages is related to the propagation of shocks. Interestingly, even if the theoretical framework applies also to single firms, the empirical part focuses on the aggregate, sector network, due to lack of more granular data. Other works on networks of firms focussed mainly on ownership. (Kogut and Walker, 2001) highlights small-world property for the ownership network of German firms. A similar analysis in (Souma et al., 2006) for the ownership network of Japanese firms highlight a power law distribution for the number of connection, a feature common to real networks from other domains. In (Vitali et al., 2011) authors enlarge the perspective to global ownership, and highlight the so called bow-tie structure, with a dense core of firms which controls most of the network. Also related to the bow-tie structure in an ownership network, this time of Italian firms, is the already mentioned (Romei et al., 2015), which also provide an algorithm to isolate the core.

To our knowledge, much less is known about the payment network between firms, mostly because of lack of data. Exception are the empirical studies (Ohnishi et al., 2009; Watanabe et al., 2012), where links represent buyer-supplier relationship between Japanese firms, but without the information on the amount of money exchanged. In the first case, once again a scale free behaviour is found, and the role of firms is characterised by their *PageRank* (Page et al., 1999) centrality, while the second one deals with the problem of diffusion (in particular of money) within the network. Finally, in (Huremovic and Vega-Redondo, 2016) authors present a theoretical model of equilibrium between competitive firms, based on the knowledge of the payment network.

Nevertheless the use of payments as a proxy of interactions between economic entities is not new and has been employed extensively, mainly for the investigation of the interactions among banks. One of the first work the highlighted, before the crisis, how the US interbank network of payments is a non trivial nexus is (Soramäki et al., 2007). (Rørdam et al., 2009) has a similar purpose but is restricted to Danish market. More recently, in (Bargigli et al.,

2015) authors improve the insight on the interbank network by treating it as a multiplex, i.e. by taking into account separately the different type of relations.

Finally, one of the objective of this chapter is to use information from the network to predict missing ratings which will be addressed using methods from machine learning. In particular we refer to the now broad literature regarding classification problems (Friedman et al., 2001). Here, we would like to recall in particular similar efforts to employ machine learning techniques in credit rating scoring. The seminal paper (Altman et al., 1994) propose a neural network classifier, where however the predictors for the rating are all derived from balance sheets, so the results are not comparable with ours. Similarly, (Wilson and Sharda, 1994; Lee, 2007) use financial ratio as predictors in a neural network and a Support Vector Machine respectively, and their performance is significantly better that a more traditional discriminant analysis. In other works more heterogeneous information have been employed to predict the rating: (Grunert et al., 2005) includes measure for the market position and management quality besides standard financial ratios and are able to improve prediction; (Parnes, 2012) includes again management quality measures and subjective opinion on the firm in the analysis.

4.2 The network of payments

4.2.1 The dataset

The investigated dataset contains information on payments between more than two million Italian firms and is built from transactional data of the payment platform of a major European bank. Transactions are registered with daily granularity for the year 2014, for a total of 47M records, each of which includes the two counterparts involved, date, type, amount and number of transactions in the same day. Transactions are originally identified by account, but in case of customers and former customers, multiple accounts associated to the same firm are aggregated into a single entity[1]. This results in a total of 2.4M entities (which will be referred to as firms, for brevity) operating through the platform during the whole investigated period. The counterparts can be of different types. In principle, any firm or public body can make use of the platform, but in practice in most cases at least one is a customer of the bank. Similar considerations hold

[1]In order to comply with privacy regulation any payment from or to physical persons is excluded, moreover the filter is implemented to exclude any ambiguous record.

for the total amount exchanged: in each month more than 50% of the volume is transferred between customers, and it rises to above 95% when considering transaction with at least one customer involved. More details on the dataset and some descriptive statistics is presented in 4.A. For customers, the dataset contains information on the economic sector and on the internal rating of the firm on a three value scale: Low (L), Medium (M), and High (H) risk.

4.2.2 Networks definition and basic metrics

Starting from transaction data, payment networks are constructed as follows: given a time window, each node represents a firm active in that period, if there is payment between two firms a link from the source to the recipient is added, with weight equal to the payment amount. If multiple transactions occur between the same (ordered) pair of nodes, the weight of the link is the sum of the amounts of the payments. Therefore for each time period we construct a directed and weighted network. The time window of analysis may vary depending on the type of information one wants to extract from the dataset. In the following the focus will be on monthly networks, for which results are quite stable, at the cost of dealing with fewer and larger graphs. For the period covered by the dataset, each monthly network consists on average of $n = 1\text{M}$ nodes and $m = 3.2\text{M}$ links with the lowest activity in August and the highest in July (see 4.A.1). The density $\rho = \frac{m}{n(n-1)}$ is thus small, resulting in a so called sparse network. Nevertheless this low density does not imply a disaggregated system. Indeed for all the networks the diameter is very small compared to the size: on average starting from a node, one has to pass at most 19 links to reach any other node in the weakly connected component (see Table 4.1). Thus the networks have the so called small-world property.

4.2.3 Networks topology

When considering a small number of firms, one would expect simple topologies: one firms is the supplier of intermediate products for another firm, resulting in a line (the simplest supply chain), or one firm is a supplier or a buyer for many others firms, resulting in a star network. Instead what is observed is a much more complex organisation, with a non negligible presence of cycles.

At a very coarse level, it is possible to identify two large classes of firms. The first constitute the core of the network, which includes approximately 20% of the nodes and more than half of the links. This core has a density an order of

Table 4.1: Basic metrics of the network of payments

	nodes	links	in-degree	out-degree	density	diameter
month	n	m	$\mathbb{E}[k^{(in)}]$	$\mathbb{E}[k^{(out)}]$	ρ	d
Jan	1,000,555	3,271,861	6.61	4.20	$3.27 \cdot 10^{-6}$	20
Feb	997,006	3,067,029	6.11	3.98	$3.09 \cdot 10^{-6}$	18
Mar	1,018,164	3,146,559	6.19	3.98	$3.04 \cdot 10^{-6}$	18
Apr	1,047,706	3,346,763	6.52	4.09	$3.05 \cdot 10^{-6}$	19
May	1,048,803	3,359,315	6.58	4.08	$3.05 \cdot 10^{-6}$	20
Jun	1,039,876	3,239,886	6.30	4.02	$3.00 \cdot 10^{-6}$	19
Jul	1,091,393	3,510,435	6.44	4.14	$2.95 \cdot 10^{-6}$	20
Aug	891,587	2,319,697	5.21	3.44	$2.92 \cdot 10^{-6}$	19
Sep	1,041,124	3,465,233	6.80	4.25	$3.20 \cdot 10^{-6}$	20
Oct	1,066,044	3,289,946	6.11	4.00	$2.89 \cdot 10^{-6}$	18
Nov	1,023,692	3,103,365	6.15	3.90	$2.96 \cdot 10^{-6}$	18
Dec	1,052,975	3,000,284	5.60	3.74	$2.71 \cdot 10^{-6}$	19

magnitude larger than that of the whole network and it is characterised by the fact that any pair of firms is connected, directly or via intermediaries. Around 60% of the total volume circulates among the nodes of the core. The other class is made of payers-only, i.e nodes that have no incoming links. These represent each month about one half of the active firms and their activity is sporadic. To better understand the role of this significant subset of firms we check their customer status and we find that the majority of them are unclassified, and that their number is larger than one expects from the unconditional distribution among all the firms. This means that likely they are not customers and, more importantly, almost no information, for example about risk, is available on them. For further details on this refer to 4.A.3.

We now turn our attention to the distribution of degree and strength. In our case the in- (out-) degree is the number of payers (payees) of a given firm and the corresponding amount of Euro. For the monthly aggregation case the average in- and out-degree of a firm is 6 and 4, respectively (see Table 4.1). These low values are a direct consequence of the low density of the network. However the degrees and the strengths are extremely heterogeneous as testified by the degree and strength distribution.

Figure 4.1 shows the empirical cumulative distribution for these two quantities in a double logarithmic scale. The approximately straight line indicates

the presence of a fat tail with a power law behaviour. The fit of the exponent supports the observation that in- and out- degree distribution data are consistent with a power-law tail and the estimated exponents are around 2.6 and 2.8, respectively. Similarly, in-strength and out-strength are well fitted by power-law distributions of exponents around 2.1 and 2, respectively. Despite the fact that a large fraction of nodes is different in each month, the tail exponents are remarkably stable (see Table 4.7 of the 4.A.3).

This scale free behaviour is quite ubiquitous in complex networks has been found in many other real economic and financial networks (Serrano and Boguná, 2003; Garlaschelli and Loffredo, 2005; Boginski et al., 2005; Boss et al., 2004; Kim et al., 2002; Huang et al., 2009; Ohnishi et al., 2009). The fat-tailed distribution for the degree has two interesting consequences: first, there is no characteristic scale for the average degree or strength; second, there are a few nodes that act as hubs for the system, in the sense that, having a large amount of connections, many pairs of nodes are connected through them. This partially explains the low values for the diameter.

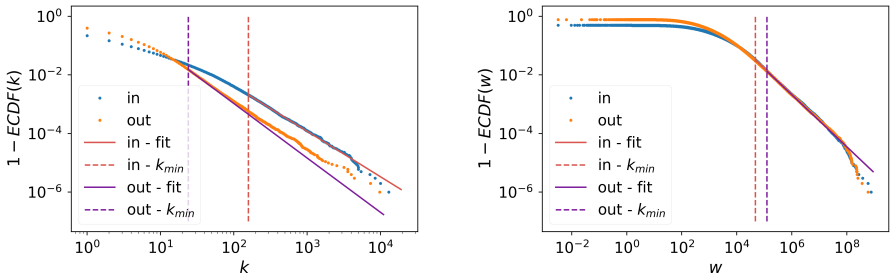


Figure 4.1: Empirical complementary cumulative degree (left) and strength (right) distributions and their power law fit. The scale is logarithmic for both axes. Data refers to January, but results are similar for the other months.

Finally, we measure the tendency of firms to be connected to firms which are similar with respect to some attribute, namely the number and the total volume of connections (i.e. degree and strength). Following (Newman, 2002), we compute the assortativity coefficient for a categorical variable,

$$r = \frac{\sum_i e_{ii} - a_i b_i}{1 - \sum_i a_i b_i} \tag{4.1}$$

where e_{ij} is the fraction of edges connecting vertices of type i and j, $a_i = \sum_j e_{ij}$ and $b_j = \sum_i e_{ij}$. It is $r_{\max} = 1$ for perfect mixing, while when the network is

perfectly disassortative (each node connects to a node of a different type) it is $r_{\min} = -\frac{\sum_i a_i b_i}{1-\sum_i a_i b_i}$. Using the number of connections as categorical variable, an high value for the assortativity coefficient indicates that highly connected firms tend to interact significantly more than average with other highly connected firms. Similar reasoning holds using the volume exchanged as categorical variable.

Beside the entire graph, we also consider the subgraph of firms with rating and the subgraph of customers. The assortativity coefficient is consistently slightly negative for both attributes, for all months and graphs, namely around -0.03 for the entire graph and the subgraph of firms with rating, and -0.04 for the subgraph of customers, with no strong differences among months and attributes. Table 4.8 of 4.A reports the summary of values of the assortativity coefficient for each month. A possible explanation can be that large, very interconnected firms are connected to many subsidiaries which in turn do not engage with many other firms, being their business almost exclusively focussed on the relationship with the large and central firms.

To summarise, each month the payment network of firms is very sparse but almost entirely connected. Half of the firms appear in the network as payers only (no incoming links) and they are mainly unclassified with respect to customer status, so no much information is available on them. Of the remaining nodes, almost half constitutes the denser core of the network where more than a half of the transactions occur and above 60% of the volume circulates. Finally the network is small world, scale free, and slightly disassortative both for degree and for strength.

4.3 Risk distribution and network topology

In this Section we investigate the distribution of risk of firms in the network of payments. We are interested in measuring the dependence between the network property of a node or a group of nodes and the risk of the firm represented by the node(s). We proceed in a bottom-up fashion, zooming out from single nodes to subsets. At first we consider a firm's local property (the number of connections) and we check if it correlates with the risk. Then we consider pairs of linked firms and measure the homophily in risk, i.e. whether firms with similar risk profile tend to do business together and thus to be linked. Finally, we divide firms into subsets induced by the network structure and we check whether the inferred subsets are informative with respect to the riskiness of the composing

firms. Specifically, we partition the network in groups (or communities) of firms by using only network information, and we test if the distribution of risk within each group is statistically different from the global one. Thus the goal is to understand if the inferred communities are homogeneous with respect to the risk profile of the composing firms: a community with many firms with high risk rating is a clear indication of financial fragility and a possible source of instability, since the distress of one or few firms of the community is likely to propagate to the other firms.

For the sake of brevity, in the following the analysis is carried on for one month, but results are consistent for all the months.

4.3.1 Degree and risk

The first investigation is on the relation between the degree of a firm and its risk. The probability for each risk level $r \in L, M, H$ conditional to the out-degree is computed[2] and plotted against the degree. The results are shown in Figure 4.2. We notice an interesting correlation between degree and risk: small degree nodes are more likely medium risk firms, whereas large degree nodes are more likely low risk firms. The high risk firms are more evenly spread across degrees, even if a larger fraction is observed for low degree nodes. To assess if the three curves are statistically different we perform a multinomial logistic regression on data (Greene, 2003) (the solid lines in the plot). This choice is justified by the fact the quantities just described are the probabilities of outcomes in a multi-class problem given an independent variable (the degree). The estimated probabilities follow quite close the trend of the empirical distribution and the coefficients are all significant. More detailed results of the fit are given in table 4.9 of 4.B.1 (first two columns).

The correlation just highlighted can be, at least in part, influenced by the effect of the size of the firm (in term of assets value from the balance sheet): a large firm is usually considered less risky than a small one, at the same time, a larger size generally implies a higher number of connections, as seen for example in the interbank network (Bargigli et al., 2015). As the size of firms is not available to us, we use the sum of the incoming and outgoing amounts as proxy. Defined in this way, the size has a Pearson correlation of around $0.19, 0.15$ with in- and out- degree, but a Spearman rank correlation of $0.67, 0.57$ respectively. To control for the effect of the size, we repeat the same

[2]The results for the in-degree are qualitatively very similar.

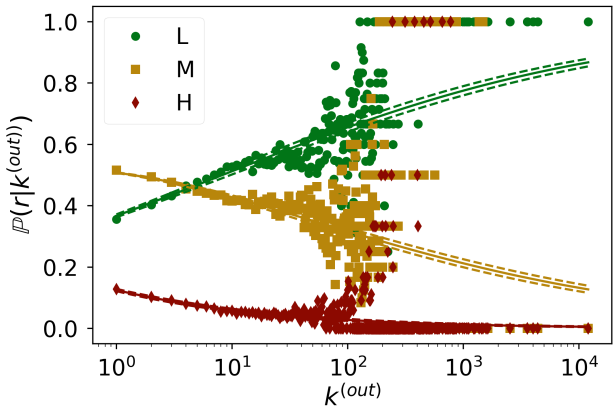


Figure 4.2: Probability of rating of a firm conditional to its out-degree. The solid lines show the fitted multinomial logistic distribution, with its confidence intervals (dashed lines) in matching colours.

procedure on subsets of firms, grouping according to their size into tertiles. The results, not shown here but available upon request, hold true also on the subsamples, but slightly less sharply. We repeat the multinomial logistic regression adding the size tertiles among the predictors, and we still obtain statistically significant coefficients (last four column in 4.9 of 4.B.1).

Similarly, the three conditional degree distributions given the rating result statistically different, as for every month all pairs reject the null hypothesis in the 2-sample Kolmogorov-Smirnov test (Smirnov, 1939). Therefore topological characteristics (the degree) of the node can be used to obtain information on the riskiness of the corresponding firm. From a risk management perspective this is an important results, since on average highly connected nodes are also less risky.

4.3.2 Assortative mixing of risk

The next step is to check whether risk is correlated with direct connection preferences. To clarify this point, two features are considered: the assortativity mixing of the risk and the conditional distribution of rating given the distance.

In the first case we compute a weighted variant of the assortativity coefficient in Eq. (4.1) using as categorical variable the risk rating rather than the

degree or the strength. In practice, the quantities e_{ij} are substituted by $\tilde{e}_{ij}$, the fraction of volume from nodes of type i to nodes of type j. The reason for this choice is to mitigate the impact of the aforementioned large number of uncategorised payers. In most cases their links are associated with low volume and few transactions. Also, customer firms, even if they represent only around 1/3 of the firms, exhibit a generally more intense activity, both in terms of number of transactions and of volume, hence accounting for the stronger ties between the firms.

The metric is positive for all the three graphs, 0.070, 0.157, 0.163 for the whole set, the nodes with rating, and the customers, respectively, with significant variability across the months but always positive sign [3]. In Table 4.10 of 4.B.2, the summary of values of the assortativity coefficients for each month is presented.

With the same quantities $\tilde{e}_{ij}$ we define metrics to assess different preferences in connection between incoming and outgoing payments. The intuition is that a firm may be more concerned for the risk of payers because a default would mean losing revenues. However, a total indifference toward the risk of suppliers (i.e. the outgoing payments) is unlikely, as the lost/shortage of supplies can induce distress if perfect substitute is not available (or it is available at a cost). In practice, for each node i we compute the percentage excess of volume with respect to the average toward nodes in certain risk class and we group according the rating of the node. The distributions are compared using Mann-Whitney U test (Mann and Whitney, 1947). This non-parametric test allow to assess if one distribution is stochastically greater than the other. Details on the metrics and the test performed are given in 4.B.2. We find that it is likely that firms are, at least in part, aware of the riskiness of their counterparts and results suggest they use this information in choosing their business partners. However the intuition that incoming payments show a more marked preference for low risk is not supported by data, moreover the overall positive assortativity is mainly lead by low risk nodes.

The quantities considered so far in this section are a pairwise comparison between the rating of nearest neighbours, and give an aggregate measure. A

[3] Results for the standard assortativity coefficient are quite different, and the choice of the subgraph appears to be crucial. When considering the entire network, the assortativity coefficient is negative, around -0.07, hence indicating a slightly disassortative behaviour with respect to risk. The subgraphs, instead, show an assortative tendency, with coefficients around 0.025 and 0.038 for the nodes with rating and for customers, respectively. This shift can be explained again by the impact of the large number of uncategorised nodes.

possible way to enrich this information is to consider the distribution of rating for nodes at a given distance[4]. Here, we look at the conditional distribution of ratings given the distance, and we compare it to the unconditional distribution. In the case of no influence of the rating on the connection pattern, the conditional distribution of risk given the distance should be statistically undistinguishable from the null unconditional distribution. To test if this is the case, we first compute the distance between all the nodes for which the rating is available. The distance between nodes in a network is defined as the length of the shortest directed path connecting two nodes, where a path is a sequence of links. Clearly, in a directed network in general $d(u,v) \neq d(u,v)$ and moreover $d(u,v)$ can be not defined (or ∞) if there is no path from u to v. Then for any fixed k, the occurrences of ratings are computed by looking at the set of pairs at distance k. Finally, the estimated distributions are tested against the null one with an hyper-geometric test, as explained in details in 4.B.3.

Results for April are summarised in right panel of Figure 4.3. We show the plot for a source node in class L, but results are similar when considering a medium or high risk source. For each k a marker indicates the percentage of nodes with low (green circles), medium (yellow squares) or high (red diamonds) risk at distance k. A marker is full when the percentage is statistically different from the null distribution (the dashed lines, with matching colours).

We note that from any starting class, up to distance 5 the class of low risk firms is significantly over-represented in the distributions. At greater distances, medium and high risk groups are over-represented. This means that more steps are necessary to reach riskier firms. This fact is particularly interesting when considering that each firm is in theory unaware of others firms' ratings and in some cases even its own.

When considering the same quantities for incoming paths, results (not shown) are very similar, namely at short distances the low risk class is over-represented, while medium and high risk nodes are over-represented for longer distances.

A possible explanation for these observations could be that among the hubs of the systems (i.e the most connected nodes) firms with rating L (i.e the most creditworthy) constitute the vast majority. This holds true when considering both in-coming and out-going links, and including also the nodes with no rating. Moreover, they are in the denser core previously described, while many high risk firms have a few or no out-going links and they are peripheral in

[4] An alternative strategy to go beyond first order neighbours in the computation of assortativity has been recently proposed by (Arcagni et al., 2017).

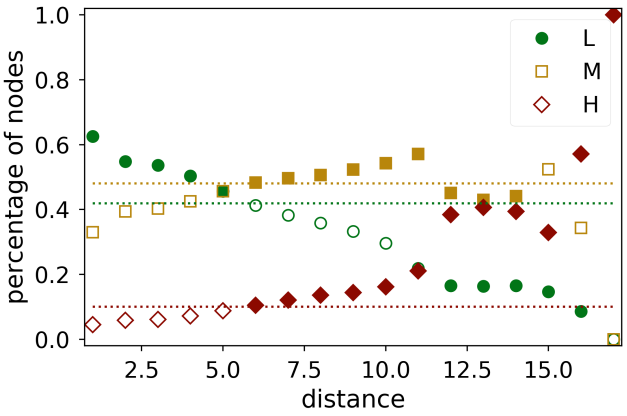


Figure 4.3: Distribution of ratings for nodes at distance k from a node with rating L. The dashed lines are the unconditional (null) distribution of ratings among nodes in the entire sample. A full marker indicates that the over or under representation with respect to the null distribution is statistically significant in the hyper-geometric test at 1% significance level with Bonferroni correction.

network. This asymmetry in the *position* in the network is observed also when considering the distribution of the closeness centrality(Newman, 2014) of the nodes, i.e the harmonic mean of the distances to all other nodes, conditioning on the risk class (not shown[5]).

4.3.3 Network organisation and risk

In this Section we study the relation between the organisation of the network at a more aggregate level and the distribution of risk. We are interested in two types of organisation of networks into groups. The first is the *modular* organisation: each module is composed by nodes, which are much more connected among themselves than with the rest of the network. In economic terms modules could represent, for example, firms operating in the same region or area, and the high density of the module reflects the fact that payments are more frequent with geographically close firms. We saw before that the network shows an assortative tendency with respect to risk, so we want to test if the homophily on risk can be observed beyond pairwise relationship.

[5]Results available upon requests

The second is a *hierarchical* organisation 1.2. This type of organisation could represent, for example, a supply chain and the flow of payments between the firms of a group and those in the group in the next rank class reflects the (opposite) flow of goods or services. This classification is important because a high risk concentration in low class nodes of a strongly hierarchical network can trigger a cascade of distress in the higher rank classes.

Finally, since the classification in modular or hierarchical organization are obtained by optimizing a specific objective function, namely maximizing modularity or minimizing agony (see below), we adopt a third method which instead infers the block structure by fitting a generative random graph model. More precisely, we employ the Stochastic Block Model (SBM) 1.3, which allows both for modular and for hierarchical structures.

Modularity and hierarchy are conceptually opposite as the first penalises connections towards other groups, which instead are encouraged in the latter (provided that they go from low rank to high rank nodes). SBM is more neutral in this sense, being in theory capable to capture both connection preferences, however it has the downside of being a parametric method.

For each metric, we proceed in the following way:

i. find the optimal partition according to the criterion;
ii. compute the distribution of ratings within each subset of the partition;
iii. test whether such *local* distribution is statistically different from the overall distribution of ratings by employing the hypergeometric test used in the previous Section and described in 4.B.3. In order to have a sample large enough to perform the test, we only consider subsets with at least 500 known ratings.

We showed so far that the structure of the payments network is very complex. Since our goal is to obtain information on the risk of the firms, it can be helpful to filter the network before performing communities detection, in order to keep the most relevant connections. Thus we focus on the subgraph of customers. The reasons for this choice are many. First, the percentage of nodes with rating active every month is quite low, around 20%, but it raises to 70% when considering only the customers (see Table 4.5 in 4.A.1 for a summary). This will help having a more informative local distribution of risk when considering subsets of nodes. Secondly, more than a half of the volume is transferred between customers (see Table 4.4 in 4.A.1), so even if a large fraction of transactions is dropped, we are mostly pruning weak connections, while keeping the strongest ones. Finally, as it has been shown in the previous Subsection

about assortativity, considering the entire network can be misleading, especially when looking at the connections without considering the weights, as it will be necessary for some metrics.

Modular structure

One of the standard methods for inferring a modular structure in a network is via modularity maximisation. This method divides nodes into subsets, called modules, such that nodes are well connected with other nodes in the same module and there is a smaller number of links with nodes in other modules. Given a partition P in modules C the modularity is

$$Q = \frac{1}{2m} \sum_{C \in P} \sum_{i,j \in C} \left(A_{ij} - \frac{k_i^{in} k_j^{out}}{2m} \right) \tag{4.2}$$

where A_{ij} is the (i, j) element of the adjacency matrix and k_i^{in} (k_i^{out}) is the in- (out-) degree of node i. The optimal partition is the one which maximizes modularity. Despite the associated optimisation problem is NP-Hard, fast and reliable heuristics for an approximate solution exist, and here the well known Louvain method (Blondel et al., 2008) is employed.

In each month we find that the optimal partition has around $2,000$ modules. These are really heterogeneous in size: for example, the 13 largest ones cover more than 95% of the nodes of the network. We perform the hypergeometric test of the null hypothesis of an homogeneous distribution of risk in each module with at least 500 known ratings. These are less than 1%, around 19 every month. (see Table 4.11 in 4.B.3 for more details). These are clearly very large modules but a significant number of them shows an over or under-expression of one or two risk classes.

For some specific module it is possible to draw statistical robust conclusions on its risk profile. The top panel of Fig. 4.4 shows the over- or under-representation for the largest modules for January. The seventh module, for example, has an over-representation of firms with low risk and an under-representation of the other two risk profiles, thus it represents a group of firms with small risk. On the contrary the eighth module has an over-representation of highly risky firms and under-representation of low risk firms, representing a possible warning for the bank.

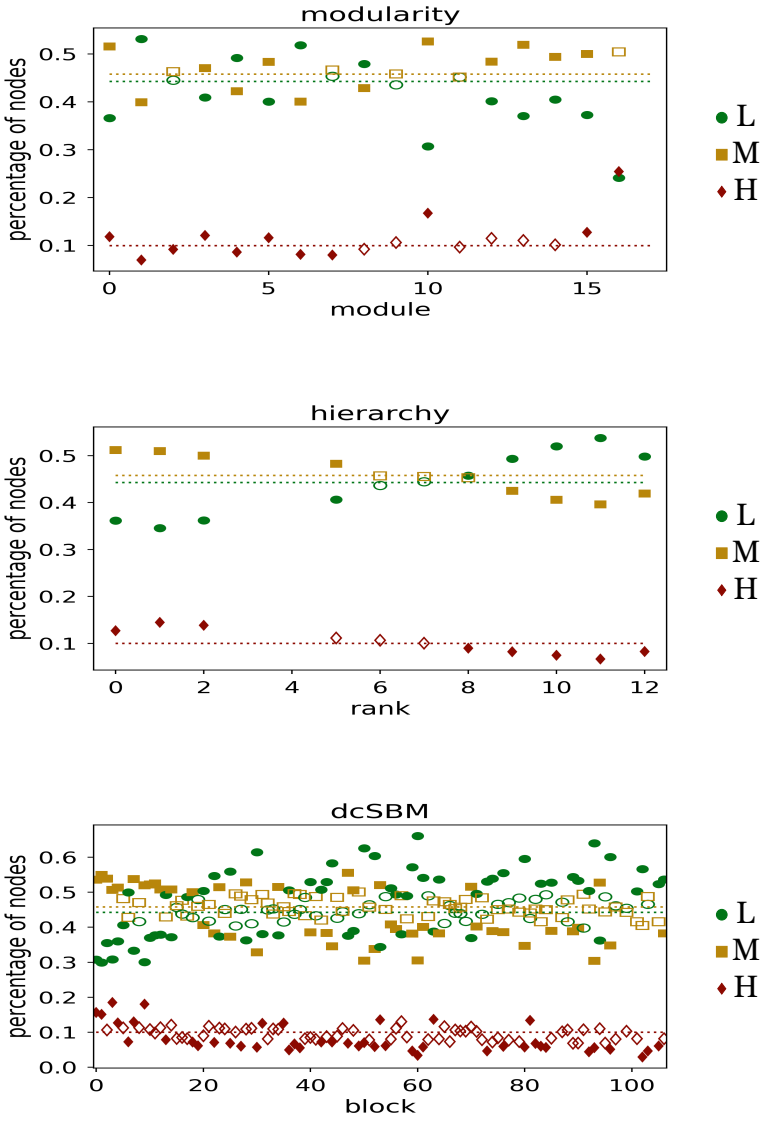


Figure 4.4: Distribution of ratings in the three partitions, modularity (top), hierarchy (middle), dcSBM (bottom). The dashed lines are the unconditional (null) distribution of ratings among nodes in the entire sample. A full marker indicates that the over (above the dashed line) or under (below the dashed line) representation with respect to the null distribution is statistically significant in the hypergeometric test at 1% significance level with Bonferroni correction.

Hierarchical organisation

We now consider explicitly the directed nature of the payment graph and the hierarchical organisation of the network. For more details on the method see Section 1.2 in Chapter 1.

We apply the hierarchy detection to the monthly networks of payments and the results are summarised in Table 4.12 of 4.B.3. First of all we notice that the number of inferred classes, roughly 18, is much lower than in the modular case. Moreover the size of the classes is much more homogeneous. The value of h is also quite stable, around 0.75, indicating a strong hierarchical structure, a remarkable result considering that we are studying only the customers network.

We now consider the distribution of risk in each class and we study the over or under-expression of certain levels of risk as a function of the rank of the class in the inferred hierarchy. The test rejects the null hypothesis of uniform risk distribution a considerable number of times (also compared with the total number of subsets in the partitions). As displayed in the bottom panel of Figure 4.4, low rank classes have an over-expression of high and medium risk firms, while middle and low rank classes (i.e. $r \in [8, 12]$) have an over expression of low risk firms and an under-expression of medium and high risk firms. More details on the test results are given in This empirical evidence may signal the presence of paths of risk propagation, since low rank firms, typically more risky, are payers of high rank firms, which are instead less risky.

Stochastic Block Model

Both methods considered above start from a cost function (modularity or agony) and look for the optimal partition minimizing or maximizing it. As we mention in section 1.3 of Chapter 1, Stochastic Block Model (SBM) inference is based on the parametrisation of the affinity matrix, whose element c_{ij} gives the probability that a node in block i is linked with a node in block j. It is interesting to note that depending on the structure of the affinity matrix one can obtain modular, hierarchical, or more complex structures.

Graphs from SBM have a Poisson distributed degree distribution, while we have shown that the payment networks have a power law degree distribution. For this reason, we use here a generalised model, the degree corrected SBM (dcSBM). In this model the link probabilities are adjusted for each node to take into account the degree distribution Karrer and Newman (2011).

When estimating a dcSBM on the payment networks we find on average

280 blocks, none of which comparable in size with the entire network. Indeed, around 200 blocks are necessary to cover at least 95% of nodes and around 115 are large enough to be tested. The inferred affinity matrix shows generally the dominance of modular structure, with most connection within the blocks, with the exception of some of the largest blocks (see Figure 4.5 for an heatmap of the affinity map, and a zoom of the largest 30 classes).

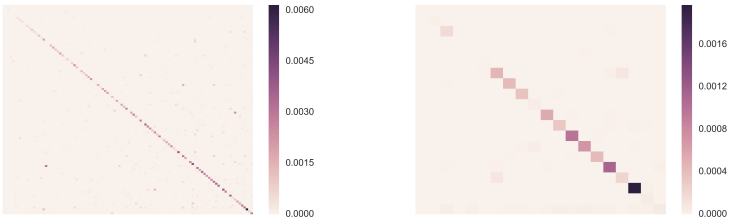


Figure 4.5: Heatmaps of the affinity matrix of January for tested classes (left) and zoom of the largest 30 (right). A darker colour indicates highest number of connections. Overall a modular structure seems to dominate, being the diagonal almost always darker than the rest, with few exception, especially among the largest classes.

From the test of the null hypothesis of homogeneous distribution of risk in each block (see Table 4.13 in Appendix 4.B.3) we observe a large number of rejections, indicating that many blocks have over and/or under expressions in the rating of the composing firms. The bottom panel of Fig. 4.4 shows the over and under expression for the largest blocks in a specific month. These might correspond to communities which are also homogeneous in the geographical location and or in the sector. Certainly a more in depth investigation is needed to understand better this important feature, which might signal closely related firms with similar rating.

4.3.4 Discussion

All the three investigated partitions give interesting insights on the relationship between risk and network structure. On one side, the percentage of rejected tests in the case of modularity partition is consistent with the observed assortativity of risk. It may be noticed that the preference for low risk business partners is not always a realistic option, because in some sectors business part-

ners are not replaceable for geographical reasons. To better assess this point, one possibility could be to include the comparison between modules and geographical location of firms, which is not available to us. On the other side, the hierarchical partition appears to follow the risk distribution slightly better and this is probably related to the peculiar conditional distribution of risk with respect to the distance described in Subsection 4.3.2. Indeed, given the fact the high risk nodes are over represented for longer distances, they should be located in extreme positions in the ranking, either at the top or at the bottom, and this is what is observed. It must be stressed that the first two methods presented, one does not exclude the other, as they give different and complementary standpoints for interpretation. In this sense, including the SBM fit is a way to check if one of the mechanisms behind the construction of the previous two partitions, predominates over the other. This check is necessary, because most methods to partition nodes of a network always give an output partition. The results here shows that a modular structure seems to dominate as the diagonal entry in most case is at least one order of magnitude greater than the others. Interestingly, there are exception among the biggest blocks meaning that modularity alone is not able to capture the complexity of the interactions among firms. In this sense a *multi-dimensional* perspective is needed, where the dimensions are the mechanisms that either favour or discourage the creation of business relationships.

4.4 Missing ratings prediction using payments network data

In the previous Sections we showed that network metrics can be informative of the risk of a firm. It is therefore natural to ask whether it is possible to predict the missing risk rating of a firm by using *only* information on network characteristics of the corresponding node, as well as risk rating of the neighbour firms. This problem is particularly relevant since we noticed that around 30% of the customers in the dataset do not have a rating and this percentage is even higher when the entire dataset is considered (see table 4.5 in 4.A.1).

Here we use network characteristics as predictors for the missing ratings into well known methods of machine learning for classification problem. The predictors we employ are the following:

i. in- and out-degree;

ii. weighted fraction of (in- and out-) neighbours with a given rating (H,M,L or NA)
iii. rank of the class in the hierarchy inferred by agony minimisation;
iv. membership in community inferred by modularity maximisation;
v. sum of in- and out-strength.

The fractions in (ii.) are computed considering the amount (weight) of each payment and are together a measure for rating assortativity, while (v.) is a proxy for the size. Data are preprocessed following (Friedman et al., 2001) so that variables are comparable in order of magnitude, as detailed in Appendix 4.C.1. These transformations result into a total of 25 predictors. The dataset is the one which includes only the customers, and we consider the monthly network for January. In order to assess the performance of the prediction, we train each model using 75% of the data, and the remaining 25% is used for testing.

4.4.1 1-step classification

We consider three methods for classification:

i. multinomial logistic;
ii. classification trees;
iii. neural networks.

See (Friedman et al., 2001) for a review of these methods. Table 4.2 shows the results of the test of each model, and the same metrics are also computed for the random classification (i.e the class is assigned sampling from the unconditional null distribution of risk in the train dataset) as a reference. The simple 1-step classifiers sizeably outperform the random assignment, both in terms of accuracy and of recall.

Table 4.2: Accuracy and recall for 1-step classifiers.

method	accuracy	recall		
		L	M	H
random	0.413	0.438	0.625	0.108
multinomial logistic	0.530	0.582	0.600	0.000
classification tree	0.530	0.539	0.635	0.023
neural network	0.535	0.510	0.67	0.019

However, as noticed in Section 4.3, the class H is under-represented in the sample, as it includes only around 10% of the firms with rating. This affects the ability of the classifiers to recover this class, as it can be seen by the value of recall for all the methods. This is undesirable, since the class H the most critical for the riskiness.

4.4.2 2-steps classification

To address this issue we proceed with a 2-step classification strategy for all the three methods. The intuition behind this strategy is to train a classifier more specialised in the recovery of one specific class at the first step, and then separate the remaining classes in the second step. In the first step we fix a risk class, say L, and we merge the other two classes into a fictitious class X. We fit a first instance of the chosen model on the modified database. In the second step, we train another instance of the model only on the two previously merged classes. This is repeated for all the three risk classes. In the case of class H being the one selected for step one, we apply SMOTE (Chawla et al., 2002) before training, a well-known algorithm for data rebalancing[6].

Once the models are trained, the prediction are obtained by iterating the following two steps for each risk class (see the schematic representation in Figure 4.6)

i. apply the first step classifier;
ii. if the entry is classified as X, apply the second step classifier.

The final prediction is the median of the predictions. In case of draw, more weight is given when the class is obtained from the first instance (as the classifier is more specialised).

Table 4.3 shows the results for each classifier, together with the value for the same metrics computed for the random classification[7]. In the case of classification trees and neural networks, different combinations for the hyper-parameters

[6]Using *SMOTE* in the 1-step classification would also be an option if the objective were to use the classifier as a first filter to detect possibly critical nodes. However, we found that the overall performance of the classifier is quite poor, especially when considering the cost of classifying as highly risky (H) a firm which is creditworthy (L).

[7]Also for the 2-steps method, the random classifier can be defined, and it gives results different from the previous case. The implementation is straightforward: the null distribution for the first step is obtained for each classifier, by taking into account the fictitious class, and at the second step by considering only the two classes previously merged.

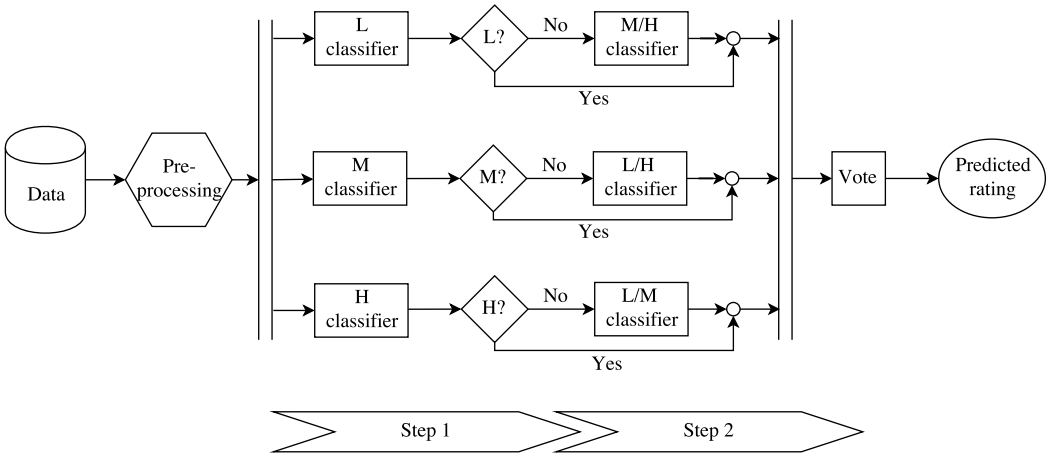


Figure 4.6: Schematic representation of the 2-steps classifier

have been tested (such as depth for the trees, and number and size of hidden layers for neural networks), here we present the results for the best choice for each model, and in 4.C.2 we explain the selecting procedure.

Table 4.3: Accuracy and recall for 2-steps classifiers.

method	accuracy	recall		
		L	M	H
random	0.366	0.368	0.391	0.249
multinomial logistic	0.477	0.553	0.452	0.253
classification tree	0.496	0.502	0.567	0.151
neural network	0.505	0.526	0.559	0.166

The three models behave quite similarly, with slightly better overall performance of neural networks, and the training times are comparable. It must be noted that, among the predictors only *network deduced* metrics have been included, while any data from the balance sheet, which is likely to represent the main source for the risk rating model, as well as the sector or geographic location, are excluded. When adding the economic sector, which is the only metadata available to us, as further predictor the prediction power only slightly improves to from 49% – 50% to around 52% of accuracy for both classification trees and neural networks. The natural benchmark models are the random classifiers, both 1-step and 2-steps, due to the total lack of data employed in the

proprietary rating model. We are able to outperform the first by 30% to 38%, and the latter by 15% to 22% in term of accuracy, and especially in the case of neural network, we are able to find a good compromise with recall for H.

4.5 Conclusions

In this Chapter we studied a unique dataset of payments between Italian firms via network methods.

We started with the study of the structure of the network of firms which highlights a complex structure: a small core of firms is responsible for most transactions. Moreover, the number of connections and the total volume exchanged by the firms display a power-law tail distribution. Both features are particularly interesting if one has in mind propagation phenomena on the network, as they can enhance the spread of distress, or positive feedbacks. Also relevant is the observed tendency of large, well-connected firms to be connected to small (in terms of exchanged volume), poorly connected firms. This can be the result of almost exclusive relationships between a big producer and its subsidiaries.

Secondly we considered the relation between the network structure and the distribution of risk. For single firms, we observed that low risk firms are more likely to have a high number of connections, and some of them acts as hubs for the entire network, being connected to thousands of other firms. When pairs of linked firms are considered, we observed the tendency to favour connections towards firms with the same risk level. This tendency can be observed also on a more aggregate level. Indeed, we found that also groups of firms which are more connected among them that with the rest of the network, have a local distribution of risk which is statistically different from the global one, meaning that some risk classes are over- or under- represented. Finally, we divided firms into a hierarchical organisation, in such a way to highlight the main direction along which money circulates and also in this case we found that many groups have a local distribution of risk statistically different from the global one. As high risk firms are over-represented at the beginning of the flow of money, this can be a source of distress for the entire system.

Finally, to show the potential of information that can mined from network metrics and communities, we presented a machine learning prediction of rating based only on these data. We propose a 2-steps strategy to which deals with under-representation in the dataset of the smallest but most risky class. We

test our strategy with three methods, namely multinomial logistic, classification trees and neural networks. Since predictors are all network-derived quantities, and no information from balance sheets or other meta-data are used, the random rating assignment is the natural benchmark. We find that all the three methods are able to outperform sizeably the benchmark, with slightly better results for neural networks.

Several extensions of the present work are possible. On one side, we employed a very simple, assumption-free, method to build the network from transactional data, but other choices are feasible, for example using a rolling window in order to give less importance to payments far in the past. This is related to problem of modelling the dynamic of the network, instead of considering snapshot as we did. On the other side we decided not to deal with the modelling for the mechanism for network formation, that can give a deeper economic insight on the system. Finally, a crucial point would be to introduce a model for risk assessment that includes network information. Our analysis clearly point out the amount of information on the riskiness that can extracted from taking network perspective, so embedding this in more general risk model, for example in the form of an adjustment of a Basel-compliant risk measure, appears to be a natural consequence.

Appendix 4.A Dataset and network metrics

4.A.1 The dataset

The dataset is built from transactional data of the payment platform of a major Italian bank for a total of 47M records. In table 4.4 the details of the exchanged volume by customer status is presented. Table 4.5 shows the distribution of rating across the firms, disaggregating them in terms of their customer status.

Table 4.4: Percentage of volume by customer status, the row indicates the status of the payer, the column the recipient

month	no	yes	ex	NA		no	yes	ex	NA	month
Jan	0.000	0.036	0.001	0.001	no	0.000	0.014	0.000	0.000	Feb
	0.009	0.604	0.030	0.110	yes	0.027	0.543	0.037	0.154	
	0.000	0.046	0.002	0.003	ex	0.000	0.037	0.000	0.002	
	0.000	0.149	0.002	0.006	NA	0.000	0.184	0.001	0.000	
Mar	0.000	0.015	0.000	0.000	no	0.000	0.018	0.000	0.000	Apr
	0.023	0.541	0.037	0.151	yes	0.023	0.525	0.033	0.155	
	0.000	0.036	0.000	0.002	ex	0.000	0.040	0.000	0.002	
	0.000	0.193	0.001	0.000	NA	0.000	0.199	0.003	0.000	
May	0.000	0.018	0.000	0.000	no	0.000	0.015	0.000	0.000	Jun
	0.023	0.542	0.035	0.144	yes	0.018	0.534	0.037	0.172	
	0.000	0.040	0.000	0.001	ex	0.000	0.033	0.000	0.001	
	0.000	0.194	0.001	0.000	NA	0.000	0.189	0.001	0.000	
Jul	0.000	0.014	0.000	0.000	no	0.000	0.014	0.000	0.000	Aug
	0.019	0.538	0.031	0.181	yes	0.018	0.591	0.029	0.140	
	0.000	0.031	0.000	0.002	ex	0.000	0.029	0.000	0.001	
	0.000	0.183	0.001	0.000	NA	0.000	0.172	0.005	0.000	
Sep	0.000	0.015	0.000	0.000	no	0.000	0.013	0.000	0.000	Oct
	0.019	0.599	0.027	0.131	yes	0.022	0.581	0.029	0.141	
	0.000	0.032	0.000	0.001	ex	0.000	0.037	0.000	0.001	
	0.000	0.175	0.001	0.000	NA	0.000	0.175	0.000	0.000	
Nov	0.000	0.015	0.000	0.000	no	0.000	0.014	0.000	0.000	Dec
	0.013	0.578	0.037	0.165	yes	0.012	0.578	0.036	0.194	
	0.000	0.031	0.000	0.001	ex	0.000	0.028	0.001	0.001	
	0.000	0.158	0.000	0.000	NA	0.000	0.137	0.000	0.000	

Table 4.5: Average monthly distribution of nodes by customer status and rating.

status	rating	count	%	%with rating
not customer incl. NA	L	2121	0.000	0.010
	M	4592	0.003	
	H	305	0.000	
	ND	676762	0.990	
customer	L	87801	0.305	0.702
	M	95893	0.333	
	H	18811	0.065	
	ND	85841	0.298	
former	L	3901	0.017	0.340
	M	7850	0.179	
	H	926	0.017	
	ND	41775	0.767	
total		1026577		0.217

4.A.2 Time aggregation

When defining a network from temporal data, choosing the time scale of analysis is crucial because it can affect deeply the topology. Shorter time scales (daily or weekly) emphasise peculiar behaviours as, for example, which supplier is paid first once liquidity is available. Longer time scales help giving a more stable picture of the supply chain structure among firms.

In order to give an intuition of different behaviours, two quantities can be considered. The first is the persistence of links and nodes, which is measured by counting the number of times a node or an edge appears in the networks for different time aggregations. From Figure 4.7 one can see that most of nodes are active only for few days, while a small core of firms is intensely active through the whole year. Secondly, the size of the networks, both in terms of number of nodes and links, for different time aggregations is shown in Figure 4.8. Interestingly, for daily aggregation, see Left panel, both quantities show a high periodicity, with a very high peak (a factor ~ 5 with respect to the other days) at the end of each month. This effect is evident also with weekly aggregation, see (central panel), but not in the monthly time scale. This last observation justifies the choice of monthly networks as focus of this analysis.

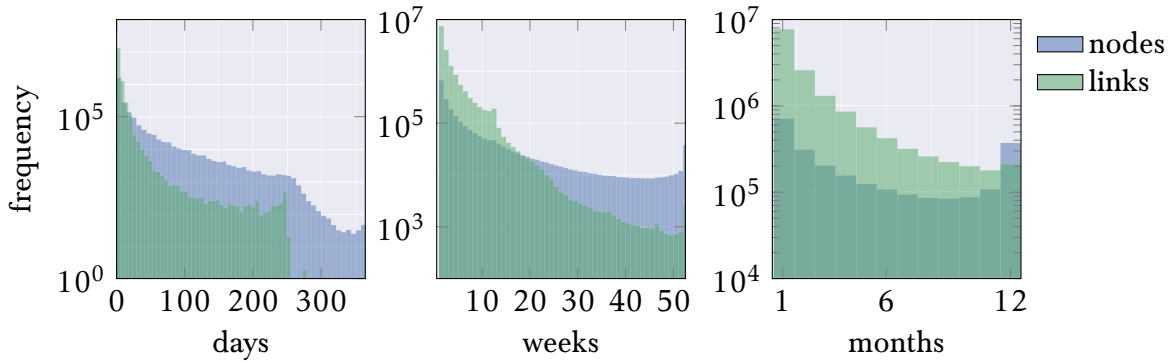


Figure 4.7: Histogram for the number of days, weeks, months of activity for nodes (blue) and existence for edges (green) for different time aggregations

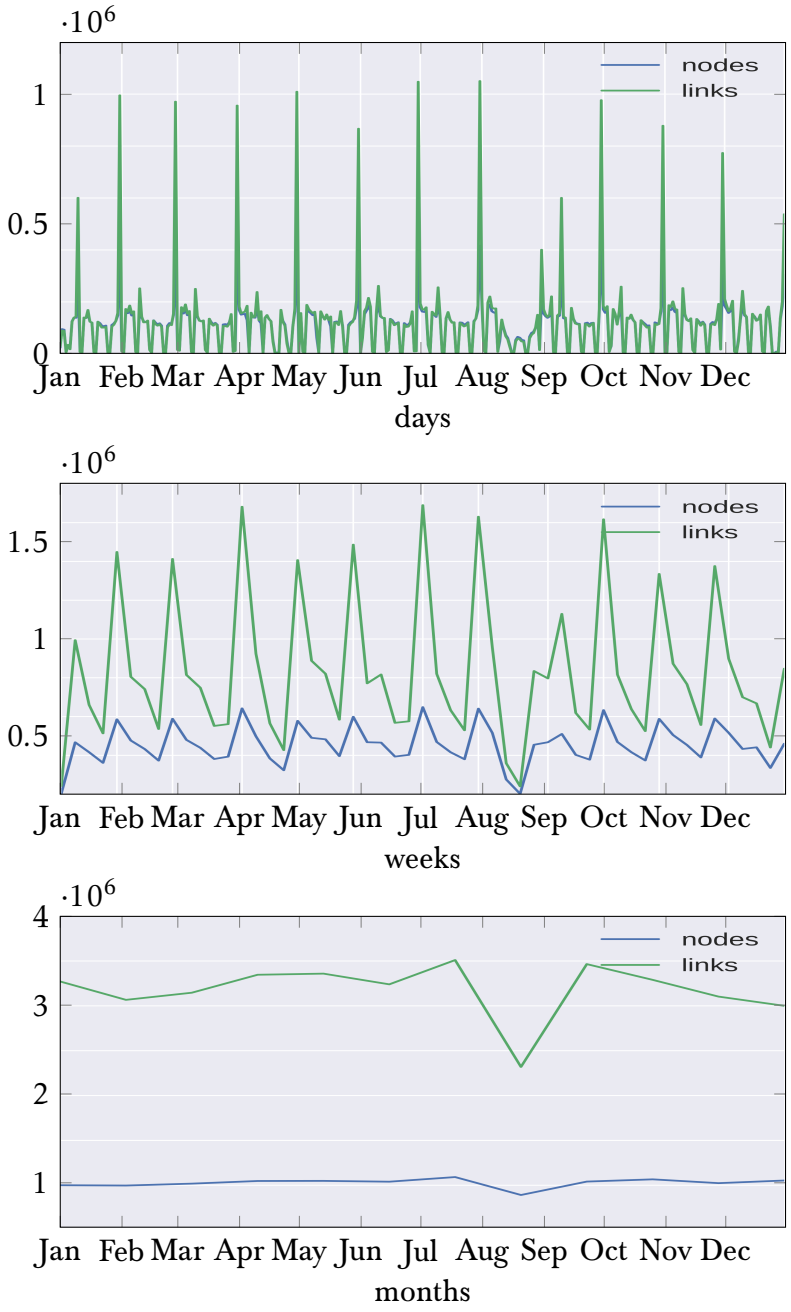


Figure 4.8: Number of nodes (blue) and links (green) for each day, week, and month. Only with longer time aggregations one is able to eliminate periodicity.

4.A.3 Network metrics

A network's component is a subsets of nodes such that there is path between any pair of nodes, either undirected (weakly connected components), or directed (strongly connected components). From the definition of the networks it is clear that there are no isolated nodes, since the smallest weak components include at least two nodes, namely a payer and a payee. As it is common for many other real networks, it is possible to identify a weak component, which is of the order of magnitude of the entire network. In our case this giant component (GC) includes on average 98% of the nodes. Considering instead the largest strongly connected component (SCC), it includes approximately 20% of the nodes but more than half of the links. As a consequence the density of the strongly connected component is an order of magnitude larger than the density of the whole network or of the weakly connected component. See Table 4.6 in for more details of these quantities.

In the standard definition of the bow-tie structure of a network, the nodes in the GC but outside the strongly connected component are divided between the in-component, the nodes from which links arrive in the strongly connected component, and the out-component, the nodes reachable from the SCC. Nodes in the in-component that have no incoming links, represent each month about one half of the active firms and their activity is sporadic.

Table 4.6: Percentage size ($\%n$) and density (ρ) of the largest weakly (GC) and strongly (SCC) connected components. The last column ($\%w$) contains the relative volume transferred among nodes in the SCC with respect to the total volume.

	GC		SCC		
month	$\%n$	ρ	$\%n$	ρ	$\%w$
Jan	0.989	$3.4 \cdot 10^{-6}$	0.232	$3.29 \cdot 10^{-5}$	0.75
Feb	0.989	$3.21 \cdot 10^{-6}$	0.237	$2.99 \cdot 10^{-5}$	0.69
Mar	0.980	$3.15 \cdot 10^{-6}$	0.235	$2.98 \cdot 10^{-5}$	0.70
Apr	0.980	$3.16 \cdot 10^{-6}$	0.231	$3.09 \cdot 10^{-5}$	0.67
May	0.981	$3.16 \cdot 10^{-6}$	0.232	$3.06 \cdot 10^{-5}$	0.69
Jun	0.980	$3.11 \cdot 10^{-6}$	0.230	$3.03 \cdot 10^{-5}$	0.69
Jul	0.982	$3.05 \cdot 10^{-6}$	0.237	$2.88 \cdot 10^{-5}$	0.70
Aug	0.970	$3.08 \cdot 10^{-6}$	0.204	$3.23 \cdot 10^{-5}$	0.69
Sep	0.981	$3.31 \cdot 10^{-6}$	0.233	$3.23 \cdot 10^{-5}$	0.73
Oct	0.981	$3.00 \cdot 10^{-6}$	0.237	$2.81 \cdot 10^{-5}$	0.68
Nov	0.979	$3.08 \cdot 10^{-6}$	0.227	$3.00 \cdot 10^{-5}$	0.65
Dec	0.979	$2.81 \cdot 10^{-6}$	0.228	$2.69 \cdot 10^{-5}$	0.67

Table 4.7: Results of power law fit of the degree and strength distribution for all the months obtained by using the algorithm described in (Clauset et al., 2009). The α parameter is the fitted exponent and the k_{min} and w_{min} parameter is the estimated minimum value after which the behaviour of the distribution is consistent with a power law tail. Since the volume of payments are scaled, the values of w_{min}s are not much informative, so for the strength $F(w_{min}) = 1 - EDCF(w_{min})$ is reported instead.

	degree				strength			
month	α^{in}	$k^{\text{in}}_{\text{min}}$	α^{out}	$k^{\text{out}}_{\text{min}}$	α^{in}	$F(w^{\text{in}}_{\text{min}})$	α^{out}	$F(w^{\text{out}}_{\text{min}})$
Jan	2.55	159	2.85	24	1.89	0.03	1.89	0.01
Feb	2.56	148	2.80	19	2.10	0.03	1.99	0.01
Mar	2.53	125	2.70	44	2.11	0.03	1.97	0.01
Apr	2.67	257	2.82	22	2.14	0.03	2.01	0.01
May	2.66	227	2.83	23	2.12	0.03	2.07	0.01
Jun	2.58	135	2.83	23	2.07	0.03	1.96	0.01
Jul	2.52	124	2.80	21	2.07	0.03	2.02	0.01
Aug	2.62	236	2.74	14	2.07	0.03	1.94	0.01
Sep	2.64	187	2.83	32	2.09	0.04	2.03	0.01
Oct	2.53	129	2.75	25	2.05	0.03	1.95	0.01
Nov	2.59	134	2.83	19	2.06	0.03	2.04	0.01
Dec	2.55	180	2.62	41	2.06	0.03	1.98	0.01

Table 4.8: Assortativity coefficient for degree and strength. The columns *with rating* refers to the subgraph of nodes with known rating. The columns *customers* refers to the subgraph of nodes with customer status *yes*.

attribute	degree			strength		
nodes	all	with rating	clients	all	with rating	clients
Jan	-0.035	-0.035	-0.046	-0.036	-0.031	-0.046
Feb	-0.027	-0.029	-0.036	-0.030	-0.031	-0.039
Mar	-0.025	-0.030	-0.038	-0.027	-0.027	-0.037
Apr	-0.027	-0.027	-0.037	-0.030	-0.029	-0.041
May	-0.026	-0.033	-0.039	-0.027	-0.026	-0.035
Jun	-0.025	-0.027	-0.032	-0.028	-0.030	-0.037
Jul	-0.025	-0.028	-0.036	-0.028	-0.028	-0.038
Aug	-0.027	-0.035	-0.040	-0.032	-0.034	-0.041
Sep	-0.024	-0.027	-0.030	-0.028	-0.030	-0.036
Oct	-0.028	-0.028	-0.037	-0.032	-0.033	-0.043
Nov	-0.023	-0.028	-0.031	-0.026	-0.028	-0.035
Dec	-0.027	-0.030	-0.034	-0.031	-0.035	-0.041

Appendix 4.B Risk distribution

4.B.1 Degree and risk

The multinomial logistic regression aims to model the probabilities for a classification problem with more than two outcomes. Here we treat the responses (L, M, H) as categorical and ordered. In practice this means to find parameters that best fit the model

$$\log\left(\frac{P(r \leq L)}{P(r > L)}\right) = a_L + b_L^1 X_1 ... + b_L^p X_p +$$

$$\log\left(\frac{P(r \leq M)}{P(r > M)}\right) = a_L + b_M^1 X_1 ... + b_M^p X_p +$$

. X_i are the predictors, a and $b_\cdot^i$ are the coefficients. We consider the cases $p = 1$, where the predictor is the degree $X_1 = k$, and the case $p = 2$ where also the size is used as predictor $X_2 = s$. In the following table 4.9, the b coefficients are show, together with an indication for the statistical significance.

Table 4.9: Coefficients for multinomial logistic regression. First two columns refers to the regression with the degree as only predictor. The last four columns refer to the regression with also the size as predictor. The superscript indicates the predictors: k for the degree, s for the size. The subscript indicates the risk rating. The stars indicate significance: one star if the p-value < 0.05, two stars if the p-value < 0.01

	b_L^k	b_M^k	b_L^k	b_L^s	b_M^k	b_M^s
Jan	0.258 **	0.352 **	0.261 **	-0.007	0.312 **	0.091 **
Feb	0.221 **	0.305 **	0.210 **	0.024 **	0.233 **	0.159 **
Mar	0.226 **	0.314 **	0.220 **	0.013	0.237 **	0.173 **
Apr	0.243 **	0.328 **	0.240 **	0.007	0.270 **	0.131 **
May	0.229 **	0.324 **	0.206 **	0.051 **	0.237 **	0.195 **
Jun	0.239 **	0.325 **	0.232 **	0.017 *	0.237 **	0.199 **
Jul	0.238 **	0.344 **	0.227 **	0.026 **	0.263 **	0.187 **
Aug	0.183 **	0.272 **	0.175 **	0.020 *	0.179 **	0.211 **
Sep	0.238 **	0.355 **	0.218 **	0.046	0.255 **	0.228 **
Oct	0.220 **	0.329 **	0.207 **	0.030 **	0.232 **	0.220 **
Nov	0.226 **	0.338 **	0.211 **	0.034 **	0.233 **	0.234 **
Dec	0.219 **	0.331 **	0.220 **	-0.002 *	0.231 **	0.227 **

4.B.2 Assortativity of risk

Table 4.10: Assortativity coefficient for risk rating. The columns *with rating* refers to the subgraph of nodes with known rating. The columns *customers* refers to the subgraph of nodes with customer status *yes*. In the last two columns, the metric for assortativity is modified in order to take into account weights, specifically e_{ij} is computed as the fraction of volume, not the number of edges (see main text for more details).

metric	standard			weighted		
nodes	all	with rating	clients	all	with rating	clients
Jan	-0.063	0.025	0.035	0.073	0.115	0.109
Feb	-0.066	0.026	0.038	0.106	0.181	0.188
Mar	-0.067	0.025	0.039	0.073	0.150	0.150
Apr	-0.067	0.026	0.036	0.069	0.154	0.156
May	-0.067	0.025	0.038	0.065	0.146	0.139
Jun	-0.068	0.026	0.039	0.060	0.150	0.128
Jul	-0.072	0.025	0.037	0.046	0.142	0.137
Aug	-0.078	0.025	0.040	0.078	0.149	0.224
Sep	-0.067	0.025	0.040	0.087	0.168	0.216
Oct	-0.076	0.024	0.037	0.080	0.151	0.213
Nov	-0.072	0.024	0.039	0.070	0.175	0.149
Dec	-0.082	0.024	0.040	0.037	0.199	0.151

To test if nodes show different preferences in connection between incoming and outgoing payments we define the quantities

$$\Delta_i^{(\text{in})}(X) = \frac{w_i^{(\text{in})}(X) - \tilde{a}_X \tilde{b}_{r(i)}}{1 - \tilde{a}_X \tilde{b}_{r(i)}}, \quad X \in \{L, M, H\}$$

$$\Delta_i^{(\text{out})}(X) = \frac{w_i^{(\text{out})}(X) - \tilde{a}_{r(i)} \tilde{b}_X}{1 - \tilde{a}_{r(i)} \tilde{b}_X}, \quad X \in \{L, M, H\}.$$

The notation is consistent with the definition in (4.1): $r(i)$ is the risk of node i; $\tilde{a}_X, \tilde{b}_X$ are the percentage volume from or to nodes with rating X for the whole network, $w_i^{(\text{out})}(X)$ ($w_i^{(\text{in})}(X)$) is the percentage of the volume from (to) node i to (from) nodes of rating X. Samples are obtained by grouping nodes by rating,

for a total of $18 (= (3 \text{ ratings})^2 \cdot 2 \text{ directions})$ distributions. For example, the distribution of excess percentage volume from L towards M is given by

$$\{\Delta_i(M)^{(\text{out})} \,|\, i \in L\} \sim F_L^{(\text{out})}(M).$$

Similarly, the excess percentage volume entering M from H is given by

$$\{\Delta_i(H)^{(\text{in})} \,|\, i \in M\} \sim F_M^{(\text{in})}(H).$$

Note that in general, $F_X^{(\text{in})}(Y) \neq F_Y^{(\text{in})}(X)$.

We perform two set of test. In the first case we fix one rating and we compare out- and in- excess percentage volume with respect to a certain rating. In all the cases the null hypothesis is rejected with very low p-values, however it is not straightforward to give an economic interpretation of the overall results: for all the rating, the excess percentage toward L is greater that the analogous for incoming volume, while the opposite holds for payments to and from H. In the second set of test we fix a rating and a direction (in or out), and we compare the excess percentage volume from (or to) all the ratings. Also in this case all the tests reject the null with very low p-values, so we are able to order the distributions and evaluate the preference in connection. For the outgoing volume, rating L is preferred to the more risky ones in all the case. Payments to nodes rated M follows in preference from nodes having risk M and H, but are last in order for nodes having rating L. For incoming payments, the situation is slightly different. Rating M is preferred by nodes rated M and H, and it is followed by L. While the preference is reversed for payments from nodes rated L.

4.B.3 Test for risk distribution within a community

The statistical test employed in the main text has the purpose to assess whether a given rating is under- or over- represented in a certain subset, obtained by one of the partitioning methods described in the paper. In general, this means to test if the distribution of ratings in a single subset is statistically different from the unconditional distribution obtained considering the entire sample. To do so, one computes the p-value representing the probability to observe a given number of ratings in each community under the null hypothesis of that ratings are distributed in the community as in the whole sample. As shown in (Tumminello et al., 2011) the probability under the null is the hyper-geometric

distribution. Moreover, since for each community multiple tests (one for each rating and community) are performed, a correction for the p-value for multiple hypothesis testing is used. In particular, the Bonferroni correction is chosen, i.e. fixed a threshold p_s for the p-value, the corrected threshold is given by $\frac{p_s}{N_r}$, where N_r is the number of tests. The threshold of is fixes at $p_s = 1\%$ before correction.

Specifically, given a partition $\{C_i\}_i$ the following quantities are computed

$$\begin{aligned} k_{x,i} &= \#\{\text{nodes in } C_i \text{ with rating } x\} \\ n_i &= \#\{\text{nodes in } C_i\} \\ K_x &= \#\{\text{nodes with rating } x\} \\ N' &= \#\{\text{nodes}\} \end{aligned}$$

and the p-value is given by

$$p = \begin{cases} \mathbb{P}(y > k_{x,i} & \frac{k_{x,i}}{n_i} > \frac{K_x}{N'}) \\ \mathbb{P}(y < k_{x,i} & \frac{k_{x,i}}{n_i} < \frac{K_x}{N'}) \end{cases}, \quad y \sim \text{hypergeom}\left(\frac{K_L}{N'}, \frac{K_M}{N'}, \frac{K_H}{N'}; N'\right).$$

Note that $\{K_x\}$ and N' are computed in the specific monthly network under consideration.

In the case of the distribution conditioned on the distance, the subsets are obtained by considering pairs of nodes. For example, the fraction of nodes with rating L at distance k from H is computed as

$$p_{HL}^{(k)} = \frac{|\{(i,j) : d(i,j) = k, i \in H, j \in L\}|}{|\{(i,j) : d(i,j) = k, i \in H\}|}.$$

The partitions resulting from the other methods are very different in terms of number and size of subsets, so to make tests comparable, only communities including at least 500 nodes with known rating. In the cases of modularity, subsets are ordered by descending size. Note that, since each month the set of active nodes and the labelling of subsets changes, one cannot easily compare the behaviour of a subsets across months.

Tables 4.11, 4.12, present a summary of the tests, recording for each month and risk class the number of times the null hypothesis has been rejected, separated in over- (+) and under- (−) representation. The last two columns contain the number classes respectively tested, and in total (nC).

Table 4.11: Summary for test results: modularity

	L		M		H				
	+	-	+	-	+	-	tested	$nC_{95\%}$	nC
Jan	4	9	8	4	6	5	17	13	1971
Feb	5	11	9	2	9	7	20	15	1900
Mar	5	13	7	2	8	4	20	14	2070
Apr	4	9	7	3	7	6	19	14	1902
May	3	9	8	2	7	5	18	13	1856
Jun	5	12	11	3	6	6	21	15	2148
Jul	6	8	10	5	5	3	17	12	1862
Aug	5	12	8	3	8	5	21	16	2608
Sep	3	9	9	2	4	4	16	12	1879
Oct	5	11	9	2	6	4	18	13	1922
Nov	5	9	7	4	5	4	17	11	2083
Dec	3	11	10	3	7	3	19	15	2323

Table 4.12: Summary for test results: hierarchy

	L		M		H					
	+	-	+	-	+	-	tested	$nC_{95\%}$	nC	h
Jan	5	5	5	4	3	5	12	11	18	0.75
Feb	5	5	4	4	4	5	12	11	17	0.74
Mar	4	4	4	4	4	4	12	11	18	0.74
Apr	4	4	3	4	3	6	12	10	15	0.75
May	6	4	4	4	3	6	12	13	18	0.74
Jun	5	3	3	4	4	6	12	11	17	0.75
Jul	4	3	3	4	3	5	12	11	16	0.74
Aug	6	4	3	5	3	5	14	11	20	0.78
Sep	5	4	3	4	4	6	12	10	17	0.74
Oct	5	4	3	4	3	5	12	11	17	0.73
Nov	5	3	4	4	3	5	12	10	18	0.75
Dec	4	3	3	4	3	7	12	12	19	0.75

Table 4.13: Summary for test results: dcSBM

	L		M		H					
	+	-	+	-	+	-	tested	$nC_{95\%}$	nC	h
Jan	5	5	5	4	3	5	12	11	18	0.75
Feb	5	5	4	4	4	5	12	11	17	0.74
Mar	4	4	4	4	4	4	12	11	18	0.74
Apr	4	4	3	4	3	6	12	10	15	0.75
May	6	4	4	4	3	6	12	13	18	0.74
Jun	5	3	3	4	4	6	12	11	17	0.75
Jul	4	3	3	4	3	5	12	11	16	0.74
Aug	6	4	3	5	3	5	14	11	20	0.78
Sep	5	4	3	4	4	6	12	10	17	0.74
Oct	5	4	3	4	3	5	12	11	17	0.73
Nov	5	3	4	4	3	5	12	10	18	0.75
Dec	4	3	3	4	3	7	12	12	19	0.75

Appendix 4.C Classification

4.C.1 Data pre-processing

It is well established (Friedman et al., 2001) that rescaling/ transforming data in order them to be $\in [0,1]$ or $\in [-1,1]$ or standardised, generally improves the performance of classification, especially when different predictors have very different scale. So, before training the models we perform data preprocessing, in particular:

i. for in- and out- degree we use quantile transformation of the logarithm of the degree. This choice is explained by the aforementioned power-law tail distribution of these quantities, and aim to avoid too scattered data;
ii. the predictors for assortativity are already $\in [0,1]$ so they do not need preprocessing;
iii. the distribution of nodes into hierarchy classes is standardised, i.e each rank is shifted and rescaled to have mean 0 ad variance 1;
iv. the module is the only categorical variable. The usual binary transformation would result into a new binary variable for each possible value. As we discussed before, the number of modules is very high but a small fraction of them contains almost all the nodes, so we only keep those that have more than 500 nodes and merge all the remaining into a residual class;
v. quantile transformation is applied also to the log-distribution of the size.

4.C.2 Models training and hyper-parameter optimisation

Models training is performed using already implemented packages: for multinomial logistic and classification trees *Scikit-learn* Python package (Pedregosa et al., 2011) has been employed, while for neural networks *Keras* Python package (Chollet et al., 2015) and *Tensorflow* (Abadi et al., 2015) have been used. However, during optimisation, the parameters that define the *architecture* of the model, the so called hyper-parameters, remain fixed. For this reason, a common practice is to train many models using different values for these hyper-parameters and compare performance according to the chosen metric(s). A thorough discussion on this topic is beyond the scope of this paper, we refer to (Bergstra and Bengio, 2012) and related literature for detailed information.

Here we apply a simple grid search for the hyper-parameters of interest.

This has been done for both 1-step and 2-steps classifiers. The metrics we employ take into account the domain specific interpretation of the risk classes. In particular we want to penalise more misclassification towards lower risk classes, i.e $M \to \bar{L}, H \to \bar{M}, H \to \bar{L}$[8], and towards *distant* classes, i.e $L \to \bar{H}, H \to \bar{L}$. For this reason, beside the standard accuracy and recall, we also consider weighted scores for accuracy ws_{acc}, recall ws_{rec}, precision ws_{pr}, which are function of the confusion matrix C. With the notation

$$C_{x,y} = |\{x \to \bar{y}\}|, \quad C_{\cdot,y} = \sum_x C_{x,y}, \qquad \forall x, y \in \{L, M, H\}$$

$$ws_{\text{acc}} = \frac{1}{C_{\cdot,\cdot}} \sum_{x,y \in \{L,M,H\}} C_{x,y} P^{\text{acc}}_{x,y}, \qquad P^{\text{acc}} = \begin{bmatrix} 1 & -0.25 & -0.5 \\ -0.75 & 1 & -0.25 \\ -1 & -0.75 & 1 \end{bmatrix}$$

$$ws_{\text{rec}} = \sum_{x,y \in \{L,M,H\}} \frac{C_{x,y}}{C_{x,\cdot}} P^{\text{rec}}_{x,y}, \qquad P^{\text{rec}} = \begin{bmatrix} 1 & -0.25 & -0.75 \\ -0.75 & 1 & -0.25 \\ -1 & -0.75 & 1.75 \end{bmatrix}$$

$$ws_{\text{pr}} = \sum_{x,y \in \{L,M,H\}} \frac{C_{x,y}}{C_{\cdot,y}} P^{\text{pr}}_{x,y}, \qquad P^{\text{pr}} = \begin{bmatrix} 1 & -0.25 & -0.75 \\ -0.75 & 1 & -0.25 \\ -1 & -0.75 & 1.75 \end{bmatrix}$$

For classification trees, the hyper-parameter of interest is the depth, i.e the maximum number of condition to be satisfied for classification (or the length of the longest path from root to leaves). A higher value for depth results in lower training error but may lead to over-fitting. We considered value of depth from 3 to 10. For the 1-step model, the tree with depth 6 resulted the best choice, while for the 2-steps, the best results have been attained with a depth of 9 for the first step tree and 5 for the second. For neural networks, the hyper-parameters of interest are the number and size of hidden layers. As before, increasing too much these values may lead to over-fitting. In order to avoid extremely high number of parameters when adding layers, we consistently reduce their size as their number increases (intuitively, the number of parameter grows as $\prod_i |l_i|$, where $|l_i|$ is the size of the ith layer). For example, in the case of 1 (hidden) layer

[8] X indicates the real class, while $\bar{X}$ indicates the predicted class.

the number of nodes is between 10 and 100, while for two layers, it goes from 5 each to 10 each. For the 1-step model the best results are obtained with 1 layer of 50 nodes, while for the 2-steps the best choice is 2 layers of 5 nodes each for the first step and 1 layer of 10 nodes for the second.

Conclusions

This thesis contributes to network literature in several directions.

In the first part we addressed the issues related to the detection of hierarchical structure in directed networks.

First, we introduced an ensemble of random graphs, termed the Ranked Stochastic Block Model, and we study how the metric agony, penalising links contrary to the hierarchy, is able to identify the planted ranking. Using symmetry arguments we have explored ranking alternative to the planted one and obtained from it by merging, splitting or inverting its classes. We have shown that when the hierarchy is not strong enough some of these alternative rankings of nodes have a value of the hierarchy larger than the planted one. This demonstrates that the metric has a resolution limit, being unable to detect small classes in large networks. In some cases we have strong numerical indications that the proposed alternative rankings, are actually close to the global optimal one. We showed that in these cases the iterated application of agony can lead to significant improvement of the hierarchy detection. We also proposed a bootstrap strategy to asses the statistical significance of the inferred partition.

Secondly, employing similar arguments of the previous case, we proved that by acting on one parameter of the metric agony, it is possible to loosen the resolution limit of hierarchy detection. Since the change of the parameter results in a problem which has non polynomial computational complexity and no exact algorithm exist in the literature, we propose an heuristic algorithm. The tests on synthetic networks give encouraging results on its performance.

To conclude the first part we showed that by exploiting the knowledge of the hierarchical structure of financial networks, it is possible to effectively monitor and control the build up of instability of the interbank network. This is achieved

by first choosing to employ the leading eigenvalue of the adjacency matrix as a model-free proxy for the level of instability of a system. The size

Finally in the second part we studied the interactions and the risk distribution of 2 million Italian firms, via the investigation of payments networks built from transactional data. Our contribution is threefold. On one side, the study of the structure of the network highlighted a complex interdependence between firms: the presence of a small denser core of well connected firms and the power-law tail distribution of the number of connections signal an architecture which may favour the spread of distress. The second and main contribution is the assessment of the correlation between the network structure and the distribution of risk. From our analysis, we can conclude that the risk level of a firm is related to its features and role in the network at different levels: from single firms to pairs, to modules. Also, the hierarchical organisation showed once more that many levels of the hierarchy have a local distribution of risk statistically different from the global one. As high risk firms are over-represented at the beginning of the flow of money, this can be a source of distress for the entire system. Finally, we showed that network metrics and communities can be successfully used to predict the missing ratings with machine learning models. We proposed a simple 2-steps strategy to compromise between overall accuracy and recall on the smallest but most risky class.

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Acharya, V., Engle, R., and Richardson, M. (2012). Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review*, 102(3):59–64.

Acharya, V., Pedersen, L., Philippon, T., and Richardson, M. (2009). Regulating systemic risk. *Restoring financial stability: How to repair a failed system*, pages 283–304.

Acharya, V. V., Pedersen, L., Philippon, T., and Richardson, M. (2013). *Taxing Systemic Risk*, page 226–246. Cambridge University Press.

Acharya, V. V. and Steffen, S. (2013). *Analyzing systemic risk of the European banking sector*, pages 247–282. Cambridge University Press.

Achlioptas, D., D'souza, R. M., and Spencer, J. (2009). Explosive percolation in random networks. *Science*, 323(5920):1453–1455.

Adams, M., Galbiati, M., and Giansante, S. (2010). Liquidity costs and tiering in large-value payment systems. *Bank of England Working Paper No. 399*.

Allen, F. and Babus, A. (2009). Networks in finance. *The network challenge*, pages 367–382.

Altman, E. I., Marco, G., and Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the italian experience). *Journal of Banking & Finance*, 18(3):505–529.

Altman, E. I. and Saunders, A. (1997). Credit risk measurement: Developments over the last 20 years. *Journal of Banking & Finance*, 21(11):1721 – 1742.

Arcagni, A., Grassi, R., Stefani, S., and Torriero, A. (2017). Higher order assortativity in complex networks. *European Journal of Operational Research*, 262(2):708 – 719.

Bardoscia, M., Battiston, S., Caccioli, F., and Caldarelli, G. (2017). Pathways towards instability in financial networks. *Nature communications*, 8:14416.

Bargigli, L., di Iasio, G., Infante, L., Lillo, F., and Pierobon, F. (2015). The multiplex structure of interbank networks. *Quantitative Finance*, 15(4):673–691.

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.

Battiston, S., Caldarelli, G., May, R. M., Roukny, T., and Stiglitz, J. E. (2016). The price of complexity in financial networks. *Proceedings of the National Academy of Sciences*, 113(36):10031–10036.

Berg, T. and Koziol, P. (2017). An analysis of the consistency of banks' internal ratings. *Journal of Banking & Finance*, 78:27 – 41.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305.

Bickel, P. J. and Chen, A. (2009). A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073.

Biggs, N. (1993). *Algebraic graph theory*. Cambridge university press.

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics*, 104(3):535–559.

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.

Boginski, V., Butenko, S., and Pardalos, P. M. (2005). Statistical analysis of financial networks. *Computational Statistics & Data Analysis*, 48(2):431–443.

Bollobás, B. (2013). *Modern graph theory*, volume 184. Springer Science & Business Media.

Bonanno, G., Caldarelli, G., Lillo, F., Micciche, S., Vandewalle, N., and Mantegna, R. N. (2004). Networks of equities in financial markets. *The European Physical Journal B*, 38(2):363–371.

Bondy, J. A., Murty, U. S. R., et al. (1976). *Graph theory with applications*, volume 290. Citeseer.

Boss, M., Elsinger, H., Summer, M., and Thurner 4, S. (2004). Network topology of the interbank market. *Quantitative Finance*, 4(6):677–684.

Botta, F. and del Genio, C. I. (2016). Finding network communities using modularity density. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(12):123402.

Bucalossi, A. and Scalia, A. (2016). Leverage ratio, central bank operations and repo market.

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.

Cantono, S. and Solomon, S. (2010). When the collective acts on its components: economic crisis autocatalytic percolation. *New Journal of Physics*, 12(7):075038.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of Artificial -intelligence Research*, 16:321–357.

Chollet, F. et al. (2015). Keras.

Chung, F. R. (1997). *Spectral graph theory*. Number 92. American Mathematical Soc.

Clauset, A., Moore, C., and Newman, M. E. J. (2007). Structural Inference of Hierarchies in Networks. *Statistical Network Analysis: Models, Issues, and New Directions. Lecture Notes in Computer Science, Vol 4503*, pages 1–13.

Clauset, A., Moore, C., and Newman, M. E. J. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.

Cont, R. (2013). *Networks*, page 283–286. Cambridge University Press.

Corominas-Murtra, B., Goñi, J., Solé, R. V., and Rodríguez-Caso, C. (2013). On the origins of hierarchy in complex networks. *Proceedings of the National Academy of Sciences*, 110(33):13316–13321.

Coscia, M. (2018). Using arborescences to estimate hierarchicalness in directed complex networks. *PloS one*, 13(1):e0190825–e0190825.

Crouhy, M., Galai, D., and Mark, R. (2000). A comparative analysis of current credit risk models. *Journal of Banking & Finance*, 24(1):59–117.

Crouhy, M., Galai, D., and Mark, R. (2001). Prototype risk rating system. *Journal of Banking & Finance*, 25(1):47–95.

Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems:1695.

Cucuringu, M. (2016). Sync-rank: Robust ranking, constrained ranking and rank aggregation via eigenvector and sdp synchronization. *IEEE Transactions on Network Science and Engineering*, 3(1):58–79.

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, 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.

Delpini, D., Battiston, S., Caldarelli, G., and Riccaboni, M. (2018). The network of us mutual fund investments: Diversification, similarity and fragility throughout the global financial crisis. *arXiv preprint arXiv:1801.02205*.

Drehmann, M. and Tarashev, N. (2013). Measuring the systemic importance of interconnected banks. *Journal of Financial Intermediation*, 22(4):586–607.

Eades, P., Lin, X., and Smyth, W. F. (1993). A fast and effective heuristic for the feedback arc set problem. *Information Processing Letters*, 47(6):319–323.

Estrella, A. and on Banking Supervision, B. C. (2000). *Credit Ratings and Complementary Sources of Credit Quality Information*. Working paper: Basel Committee on Banking Supervision. Bank for International Settlements.

Euler, L. (1736). Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 8:128–140.

Fama, E. F. and French, K. R. (2002). Testing trade-off and pecking order predictions about dividends and debt. *Review of Financial Studies*, 15(1):1–33.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486:75–174.

Fortunato, S. and Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41.

Frank, M. Z. and Goyal, V. K. (2003). Testing the pecking order theory of capital structure. *Journal of Financial Economics*, 67(2):217–248.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in Statistics New York.

Fujiwara, Y., Aoyama, H., Ikeda, Y., Iyetomi, H., and Souma, W. (2009). Structure and temporal change of the credit network between banks and large firms in japan. *arXiv preprint arXiv:0901.2377*.

Gai, P., Haldane, A., and Kapadia, S. (2011). Complexity, concentration and contagion. *Journal of Monetary Economics*, 58(5):453 – 470. Carnegie-Rochester Conference on public policy: Normalizing Central Bank Practice in Light of the credit Turmoi, 12–13 November 2010.

Gai, P. and Kapadia, S. (2010). Contagion in financial networks. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*.

Garlaschelli, D. and Loffredo, M. I. (2005). Structure and evolution of the world trade network. *Physica A: Statistical Mechanics and its Applications*, 355(1):138–144.

Gauthier, C., Lehar, A., and Souissi, M. (2012). Macroprudential capital requirements and systemic risk. *journal of Financial Intermediation*, 21(4):594–618.

Georg, C.-P. (2013). The effect of the interbank network structure on contagion and common shocks. *Journal of Banking & Finance*, 37(7):2216–2228.

Girvan, M. and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826.

Glattfelder, J. B. and Battiston, S. (2009). Backbone of complex networks of corporations: The flow of control. *Physical Review E*, 80(3):036104.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Grunert, J., Norden, L., and Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance*, 29(2):509–531.

Gupte, M., Shankar, P., Li, J., Muthukrishnan, S., and Iftode, L. (2011). Finding hierarchy in directed online social networks. In *Proceedings of the 20th International Conference on World Wide Web*, pages 557–566. ACM.

Haldane, A. G. (2013). Rethinking the financial network. In *Fragile stabilität–stabile fragilität*, pages 243–278. Springer.

Haldane, A. G. and May, R. M. (2011). Systemic risk in banking ecosystems. *Nature*, 469(7330):351.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137.

Huang, W.-Q., Zhuang, X.-T., and Yao, S. (2009). A network analysis of the chinese stock market. *Physica A: Statistical Mechanics and its Applications*, 388(14):2956–2964.

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.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1):193–218.

Huremovic, K. and Vega-Redondo, F. (2016). Production networks.

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.

Johnson, S., Domínguez-García, V., Donetti, L., and Muñoz, M. A. (2014). Trophic coherence determines food-web stability. *Proceedings of the National Academy of Sciences*, 111(50):17923–17928.

Johnson, S. and Jones, N. S. (2017). Looplessness in networks is linked to trophic coherence. *Proceedings of the National Academy of Sciences*, 114(22):5618–5623.

Karrer, B. and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107.

Kim, H.-J., Lee, Y., Kahng, B., and Kim, I.-m. (2002). Weighted scale-free network in financial correlations. *Journal of the Physical Society of Japan*, 71(9):2133–2136.

Kogut, B. and Walker, G. (2001). The small world of germany and the durability of national networks. *American Sociological Review*, 66(3):317–335.

Krackhardt, D. (1994). Graph theoretical dimensions of informal organizations. *Computational Organization Theory*, 89(112):123–140.

Kumpula, J. M., Saramäki, J., Kaski, K., and Kertész, J. (2007). Limited resolution in complex network community detection with potts model approach. *The European Physical Journal B*, 56(1):41–45.

Langville, A. N. and Meyer, C. D. (2012). *Who's# 1?: the science of rating and ranking*. Princeton University Press.

Lee, Y.-C. (2007). Application of support vector machines to corporate credit rating prediction. *Expert Systems with Applications*, 33(1):67 – 74.

Lehmann, E. and Neuberger, D. (2001). Do lending relationships matter?: Evidence from bank survey data in germany. *Journal of Economic Behavior & Organization*, 45(4):339 – 359.

Leskovec, J. and Krevl, A. (2014). SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`.

Letizia, E., Barucca, P., and Lillo, F. (2018). Resolution of ranking hierarchies in directed networks. *PloS One*, 13(2):1–25.

Letizia, E. and Lillo, F. (2017). Corporate payments networks and credit risk rating. *arXiv preprint arXiv:1711.07677*.

Levy-Carciente, S., Kenett, D. Y., Avakian, A., Stanley, H. E., and Havlin, S. (2015). Dynamical macroprudential stress testing using network theory. *Journal of Banking & Finance*, 59:164–181.

Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.-O., Han, J.-D. J., Chesneau, A., Hao, T., Goldberg, D. S., Li, N., Martinez, M., Rual, J.-F., Lamesch, P., Xu, L., Tewari, M., Wong, S. L., Zhang, L. V., Berriz, G. F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H. W., Elewa, A., Baumgartner, B., Rose, D. J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S. E., Saxton, W. M., Strome, S., van den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K. C., Harper, J. W., Cusick, M. E., Roth, F. P., Hill, D. E., and Vidal, M. (2004). A map of the interactome network of the metazoan c. elegans. *Science*, 303(5657):540–543.

Li, W., Kenett, D. Y., Yamasaki, K., Stanley, H. E., and Havlin, S. (2014). Ranking the economic importance of countries and industries. *arXiv preprint arXiv:1408.0443*.

Lohmann, C. and Ohliger, T. (2017). Nonlinear relationships and their effect on the bankruptcy prediction. *Schmalenbach Business Review*, 18(3):261–287.

Lopez, J. A. and Saidenberg, M. R. (2000). Evaluating credit risk models. *Journal of Banking & Finance*, 24(1):151 – 165.

Maiya, A. S. and Berger-Wolf, T. Y. (2009). Inferring the maximum likelihood hierarchy in social networks. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, pages 245–250. IEEE.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Aannals of Mathematical Statistics*, pages 50–60.

Mantegna, R. N. and Stanley, H. E. (1999). *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press.

Mayer, C. and Sinai, T. (2003). Network effects, congestion externalities, and air traffic delays: Or why not all delays are evil. *American Economic Review*, 93(4):1194–1215.

Mode, C. J. (1971). *Multitype branching processes: theory and applications*, volume 34. American Elsevier Pub. Co.

Mones, E. (2013). Hierarchy in directed random networks. *Physical Review E*, 87(2):022817.

Nadakuditi, R. R. and Newman, M. E. J. (2012). Graph spectra and the detectability of community structure in networks. *Physical Review Letters*, 108(18):1–5.

Nepusz, T. and Vicsek, T. (2013). Hierarchical self-organization of non-cooperating individuals. *PloS One*, 8(12):e81449.

Newman, M. E. (2002). Assortative mixing in networks. *Physical Review Letters*, 89(20):208701.

Newman, M. E. J. (2012). Communities, modules and large-scale structure in networks. *Nature Physics*, 8(1):25–31.

Newman, M. E. J. (2014). *Networks: An introduction*. Oxford University.

Newman, M. E. J. (2016). Community detection in networks: Modularity optimization and maximum likelihood are equivalent. *arXiv preprint arXiv:1606.02319*.

Nguyen, H. and Zheng, R. (2014). A data-driven study of influences in twitter communities. In *2014 IEEE International Conference on Communications (ICC)*, pages 3938–3944. IEEE.

Ohnishi, T., Takayasu, H., and Takayasu, M. (2009). Hubs and authorities on japanese inter-firm network: Characterization of nodes in very large directed networks. *Progress of Theoretical Physics Supplement*, 179:157–166.

on Banking Supervision, B. C. (2000). *Range of Practice in Banks' Internal Ratings Systems*. Basel Committee on Banking Supervision.

on Banking Supervision, B. C. (2003). *The New Basel Capital Accord. Consultative Document,*. Bank for International Settlements.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Parnes, D. (2012). Approximating default probabilities with soft information. *The Journal of Credit Risk*, 8(1):3.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Peixoto, T. P. (2014a). The graph-tool python library. *figshare*.

Peixoto, T. P. (2014b). Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X*, 4:011047.

Poledna, S. and Thurner, S. (2016). Elimination of systemic risk in financial networks by means of a systemic risk transaction tax. *Quantitative Finance*, 16(10):1599–1613.

Romei, A., Ruggieri, S., and Turini, F. (2015). The layered structure of company share networks. In *IEEE Data Science and Advanced Analytics , DSAA-2015*, pages 1–10. IEEE.

Rørdam, K. B., Bech, M. L., et al. (2009). The topology of danish interbank money flows. *Banks and Bank Systems*, 4:48–65.

Salthe, S. N. (2010). *Evolving hierarchical systems*. Columbia University Press.

Salwinski, L. and Eisenberg, D. (2004). In silico simulation of biological network dynamics. *Nature biotechnology*, 22(8):1017.

Santos, J. A. C. (2001). Bank capital regulation in contemporary banking theory: A review of the literature. *Financial Markets, Institutions & Instruments*, 10(2):41–84.

Serrano, M. A. and Boguná, M. (2003). Topology of the world trade web. *Physical Review E*, 68(1):015101.

Shetty, J. and Adibi, J. (2005). Discovering Important Nodes through Graph Entropy: The Case of Enron Email Database. *Proceedings of the 3rd International Workshop on Link Discovery*, pages 74–81.

Simon, H. A. (1991). The architecture of complexity. *Facets of Systems Science*, 106(6):457–476.

Slater, P. (1961). Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48(3/4):303–312.

Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin Math. Univ. Moscou*, 2(2).

Soramäki, K., Bech, M. L., Arnold, J., Glass, R. J., and Beyeler, W. E. (2007). The topology of interbank payment flows. *Physica A: Statistical Mechanics and its Applications*, 379(1):317–333.

Souma, W., Fujiwara, Y., and Aoyama, H. (2006). Change of ownership networks in Japan. In *Practical Fruits of Econophysics*, pages 307–311. Springer.

Tatti, N. (2017). Tiers for peers: a practical algorithm for discovering hierarchy in weighted networks. *Data Mining and Knowledge Discovery*, 31(3):702–738.

Tibély, G., Pollner, P., Vicsek, T., and Palla, G. (2013). Extracting tag hierarchies. *PloS One*, 8(12):e84133.

Treacy, W. F. and Carey, M. (2000). Credit risk rating systems at large us banks. *Journal of Banking & Finance*, 24(1):167–201.

Trusina, A., Maslov, S., Minnhagen, P., and Sneppen, K. (2004). Hierarchy measures in complex networks. *Physical Review Letters*, 92(17):178702.

Tumminello, M., Aste, T., Di Matteo, T., and Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*, 102(30):10421–10426.

Tumminello, M., Miccichè, S., Lillo, F., Varho, J., Piilo, J., and Mantegna, R. N. (2011). Community characterization of heterogeneous complex systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(01):P01019.

Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability*, 7(3):111–125.

Vitali, S., Glattfelder, J. B., and Battiston, S. (2011). The network of global corporate control. *PloS One*, 6(10):e25995.

Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.

Watanabe, H., Takayasu, H., and Takayasu, M. (2012). Biased diffusion on the japanese inter-firm trading network: estimation of sales from the network structure. *New Journal of Physics*, 14(4):043034.

Wellman, B. (1983). Network analysis: Some basic principles. *Sociological Theory*, pages 155–200.

Wilson, R. L. and Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems*, 11(5):545–557.

Wolski, M. and van de Leur, M. (2016). Interbank loans, collateral and modern monetary policy. *Journal of Economic Dynamics and Control*, 73:388–416.

Zachary, W. (1977). Ww zachary, j. anthropol. res. 33, 452 (1977). *J. Anthropol. Res.*, 33:452.

Zhang, F., Tadikamalla, P. R., and Shang, J. (2016). Corporate credit-risk evaluation system: Integrating explicit and implicit financial performances. *International Journal of Production Economics*, 177:77 – 100.

Zhang, P. and Moore, C. (2014). Scalable detection of statistically significant communities and hierarchies, using message passing for modularity. *Proceedings of the National Academy of Sciences*, 111(51):18144–18149.