



Master in Artificial Intelligence (UPC-URV-UB)

Master of Science Thesis

**Social Network Analysis and the
illusion of gender neutral
organisations**

Atia Cortés Martínez

Advisor/s: Arturo Tejeda M.Sc.

September 2012

Declaration of Authorship

I, Atia Cortés Martínez certify that the thesis entitled *Social Network Analysis and the illusion of gender neutral organisations* I have presented for examination for the Degree of Master on Artificial Intelligence of the Technical University of Catalonia is solely my own work, developed under the supervision of my advisor. Where I have quoted from the work of others, the source is always given.

I declare that this authorisation does not, to the best of my knowledge, infringe the rights of any third party.

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without the prior written consent of the author.

I would like to dedicate this thesis to my worldwide distributed
family

Acknowledgements

And I would like to acknowledge to the people that gave me support during this thesis period and that without them I would felt lost more than once:

- Arturo Tejeda, my advisor during the last period of the thesis, for always encouraging me and being so positive
- Ramon Sangesa, for introducing me into the world of social networks and their study
- My friends and my master colleagues for the continuous support
- My family and my partner, just for being always there and stand me.

Contents

1	Introduction	1
1.1	Main goals	4
1.2	Plan of the work	5
2	Social Network Analysis	7
2.1	Background information	7
2.2	Connections	8
2.3	Node distribution	10
2.3.1	Degree Centrality	10
2.3.2	Closeness centrality	11
2.3.3	Betweenness centrality	12
2.3.4	Eigenvectors centrality	12
2.3.5	Structural holes	13
2.4	Community structure	14
2.4.1	Dyads and triads	15
2.4.2	Hierarchical clustering	16
2.4.3	Girvan-Newman algorithm	18
2.4.4	Clique based methods	19
2.4.5	Quality measures	20
2.5	Complex networks	21
3	State of the art	25

CONTENTS

4	Methodology and data	33
4.1	Data	33
4.1.1	Background	34
4.1.2	Dataset information	36
4.2	Methodology	36
5	Results	39
5.1	Structuring the data	39
5.2	Data Analysis	40
5.3	Analyzing influence over time	51
6	Conclusions and Future work	65
6.1	Conclusions	65
6.2	Future work	67
A	Appendix A	69
B	Appendix B	75
	References	83

List of Figures

1.1	Facebook VS Twitter. User data 2010	2
1.2	Social media platforms by total number of users	3
2.1	Isomorphism classes with MAN labelling	17
2.2	Example of a network	20
2.3	Random networks VS scale-free power-law networks	22
2.4	Birth of a scale-free networks	23
3.1	Moreno's earliest <i>sociogram</i>	26
3.2	Moreno's sociogram of fourth grade friendship	28
4.1	Average proportion of women on boards and chairwomen.	35
5.1	Evolution of the annual average gender representation	40
5.2	Evolution of the annual average of the number of edges and triads	41
5.3	Evolution of the annual average of the number of communities . .	42
5.4	Evolution of cumulative clustering coefficient	43
5.5	Evolution of Node Degree and Diameter of the Network	45
5.6	Cumulative node degree, year 2002	46
5.7	Cumulative node degree, year 2005	47
5.8	Cumulative node degree, year 2008	49
5.9	Cumulative node degree, year 2010	50
5.10	Evolution of Top 25 rankings for Betweenness and PageRank . . .	54
5.11	Relations among directors represented by betweenness centrality. January-February 2003	55
5.12	Pagerank JanFeb2003	57

LIST OF FIGURES

5.13	Relations among directors represented by betweenness centrality. March-April 2005	58
5.14	Relations among directors represented by betweenness centrality. June-July 2005	59
5.15	Relations among directors represented by PageRank centrality. June- July 2005	60
5.16	Relations among directors represented by betweenness centrality. April-May 2009	62
5.17	Relations among directors represented by PageRank centrality. April- May 2009	64
A.1	CDF 2003	70
A.2	CDF 2005	71
A.3	CDF 2007	72
A.4	CDF 2009	73
A.5	ICDF 2011	74
B.1	Number of communities by month	76
B.2	Effective diameter by months	77
B.3	Number of open and closed triads by month.	78

List of Tables

5.1	Ranking of the top 25 nodes for centrality.	52
-----	---	----

LIST OF TABLES

Chapter 1

Introduction

When we talk about social networks, many people will directly think about social media. Few years ago, our social environment was limited to our school or university friends, our colleagues at work, or people that we met in other activities like gyms, associations, university exchanges, etc. Nowadays, we have the Internet and, through it, access to any kind of information at any moment.

The Internet allows us to know every novelty in the world at live almost online, to buy shares in New York or Tokyo from home and, more related to this thesis, to interact with everyone instantaneously, no matter the distance. To illustrate this, let us give some figures: the radio needed 38 years to reach 50 million users; TV needed 13 years and finally Internet only 4 years. Facebook added over 200 million users in less than one year. We still spread information or *rumours* by the world of mouth, but now this information moves faster and further. We have many social tools, through which we can share any kind of information, share our opinions, interests, pictures and any kind of our daily activities. Twitter¹, Facebook² are currently the most popular social networks, used everyday by millions of users around the world (900 and 500 millions respectively)³. While Twitter acts more like a RSS reader, where you follow other users posts (limited to 140 characters) according to some similar interest and have the possibility to forward (retweet) others messages to your list of followers, Facebook is more about

¹<http://Twitter.com>

²<http://www.facebook.com>

³ http://www.diffen.com/difference/Facebook_vs_Twitter

1. INTRODUCTION

contacts between friends, sharing photos or events and chatting with people you know. The relations in Facebook are of mutual friendship, while in Twitter relations are not necessarily reciprocal: you may be followed by thousands of users you do not know (this is, for example, the case of celebrities).

FACEBOOK VERSUS TWITTER 2010

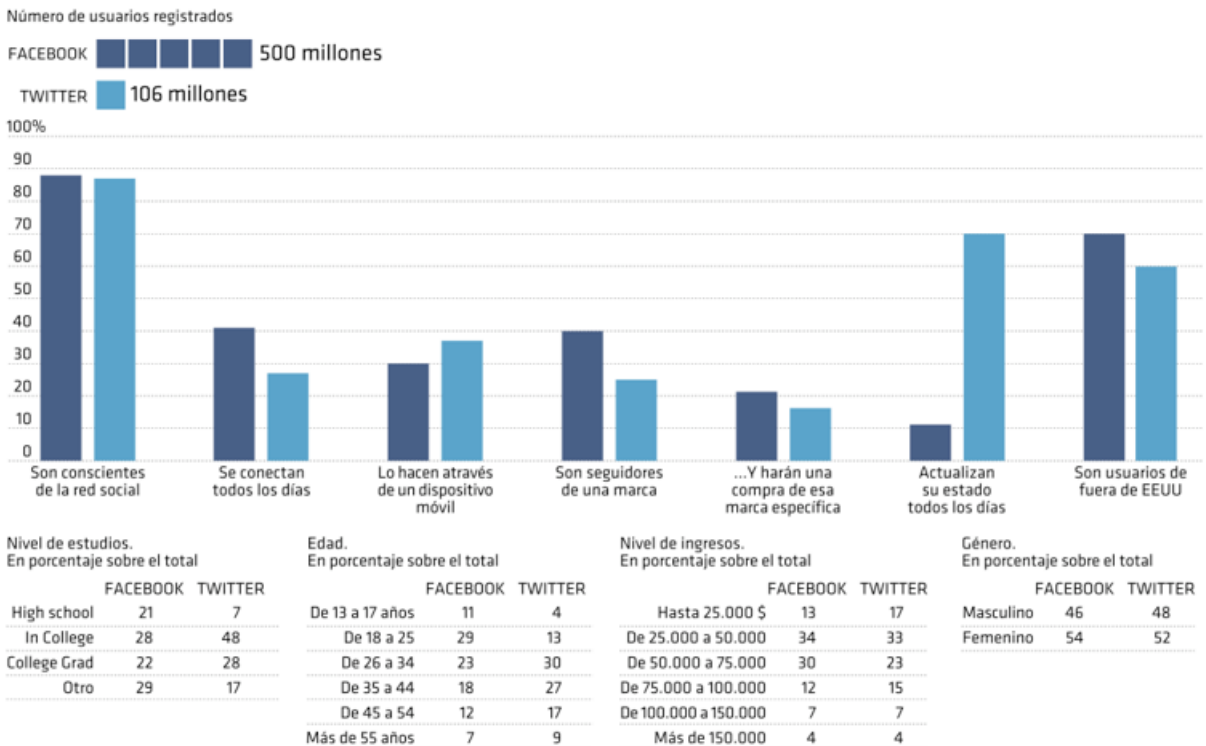


Figure 1.1: Facebook VS Twitter. User data 2010

There are other more specialized social networks, like LinkedIn for professional contacts; YouTube and Vimeo for video sharing; Instagram and Flickr for pictures, and here again we can find many categories of pictures; Foodspotting for sharing restaurants and foods; Endomondo, a sports community based on free real-time GPS tracking of running. The appearance of blog platforms like Blogspot or Wordpress allows everyone to express their opinions about something in particular and spread it to the world. Google, and later Wikipedia, have

changed our methods of looking for information, leaving in the background encyclopaedias and libraries. We give all the information to build a complete profile of ourselves that will be extended every day.

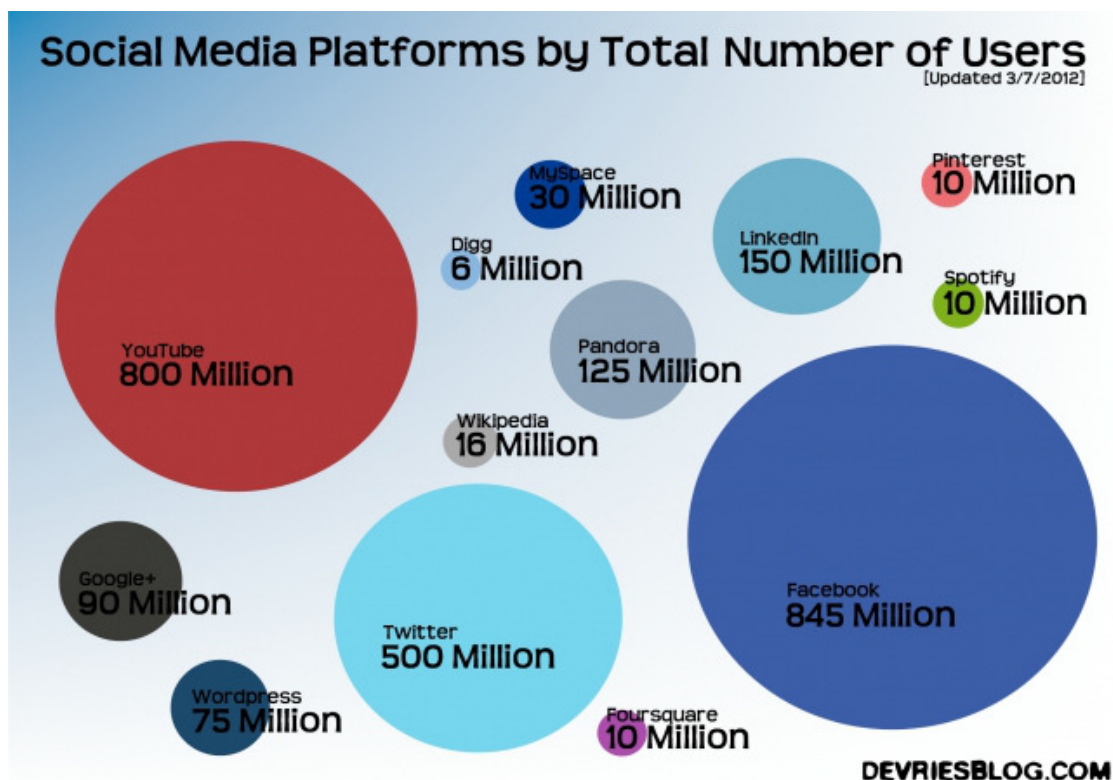


Figure 1.2: Social media platforms by total number of users

All these tools have an impact in our lives in these days. First, we are able to interact with a larger number of people than few go, even if we do not know them; even if they live in other countries or time zone. This means that our social network increases as we interact in the Internet. The more information you share, the more social tools you use, and then the more people you will be able to reach. Social media networking is an online service, or a platform, where you get the opportunity to broaden your network by meeting new people from all over the globe.

Second, the time a piece of information needs to go from one side of the Earth to another is virtually zero. Information can be spread easier thanks to

1. INTRODUCTION

the Internet, but why does a new video, article or idea spread faster than others? It is thanks to its social impact, on *how* many people talks about or share in their social media that information will become a *viral*. This term comes from the study of epidemiology, based on the behaviour of a biological virus: something that will be easily replicated and spread around an organism. While with traditional methods of communication such as a newspaper or magazine you are simply given the information, social media interacts with you while giving you the information. Before the appearance of the Internet and the boom of social media, we could only receive information through traditional media and consume it; now we are able to send, or provide, information to others.

Third, marketing business, now called Internet marketing, have learned how to get benefits of this situation. Thanks to all the information we provide, they can create more specialized content for customers. For example, social media users can now follow brands on their networks to be updated of new products. Many brands are focusing in creating social media campaigns so that fans engage to them and share it with their contacts. This provides a direct access to everyone who shares or likes the brands activity, so it creates an increasing accessible target market that was not possible before. If we search for the worlds top 20 largest brands, 25% of the results are links to user-generated content. It is not about whether brands and organizations do social media, it is about how well they do it (37).

1.1 Main goals

The questions addressed by this thesis are:

1. Who decides which information is more relevant than another?
2. How does the viral begin?
3. Is it a strategy or just chance?
4. By taking the structure of a network, can we detect the individuals in a network able to spread information and reach most of the rest of the population?

This thesis uses tools and measures emerging social network analysis to study whether organisations are gender neutral. This work is inspired in previous work by (41),(40). We will be using the same data as these authors as Scandinavian countries and, specially, Norway are quite advanced in implementing *equality policies*. Equality strategies have been seen as potential ways to counteract the strong patterns of occupational sex segregation, both from governments, policy-makers as well as researchers (40). We will focus only in the data referred to the corporate Boards of Directors¹ (BODs). The complete data description of the data is given in §4.1.

In particular, we do believe that social network analysis will be a useful to help to answer some relevant questions as:

1. Which will be the influence of women in society after the implementation of gender equality policies?
2. Do relations among BODs members will change?
3. Are mandatory gender quotas on corporate boards *good* policies?

M. Teigen has addressed these topics from the social point of view (45).

1.2 Plan of the work

This thesis document will be divided in 6 chapters. Chapter 2 provides some basic concepts related to social networks and social network analysis, basically those that will be used in this work. This chapter is intended to ease the reading of the rest of the document for non-experts.

Chapter 3 is a state-of-the-art about methods for analysing the structure of complex networks. Chapter 4 describes the methodology and data used in this thesis. Results of our experimentation are explained in Chapter 5.

Finally, in chapter 6 contains our conclusions and comments about future work.

¹Organizations, whether for-profit, or not-for-profit, usually have a Board of Directors. We can think of this board as a network that belongs to the organization. All members are linked if they sit on an organization's board together.

1. INTRODUCTION

Chapter 2

Social Network Analysis

2.1 Background information

Social networks are useful in social science to study relationships between individuals, groups, organizations or entire societies . The study of these structures use methods called social network analysis (SNA) and covers a set of interdisciplinary academic field emerged from social psychology, sociology, statistics and graph theory, but is nowadays applied in other fields like computer science, communication and information, economics or biology. There is an increasing interest in studying topological features of the network,as the structures we may find in recent real networks are non-trivial. The main idea here is that the topology of a network can affect the kind of interactions between units, making them more or less efficient when communicating.

For a complete analysis, it is important to study the structure of the network as a whole, but also at individual level. This is usually done applying graph theory to SNA, due to its representational capacity and simplicity. Basically, the graph consists of nodes n) and connections, or links, (1) which connect the nodes. In graph representation they correspond to vertices and edges respectively. In social networks, the representation by graphs is also called *sociogram*, where the nodes are the actors or events, and the lines of connection establish the set of relationships in a two dimensional drawing. The relationship between two nodes can be either non-directional (reciprocal, like marriages or contracts between companies), or non-directional, which means that the relation is not

2. SOCIAL NETWORK ANALYSIS

necessarily mutual. In this case a connection goes from an actor (origin) and ends at another (destination). An example of non-directional connections is a network representing sales and purchases between different companies. Graphs provide some tools to analyse and visualise some issues of the network. The size of real networks are, however, increasingly larger, with many actors and connections turning such studies impossible. To resolve this problem, SNA use the matrices developed by sociometrics (called sociomatrices) to complement graph theory, establishing a mathematical basis for analyses of social network. An example is the symmetry matrix which represents in a matrix the relations among nodes. In the case of non-directional networks, this matrix will rarely be symmetric.

The use of graphs and sociomatrices is necessary in order to create models, or simplified representations, of networks of relationships. However, it is not enough to represent the whole of the characteristics and attributes of the network, as we only have by now information about whether a relation exists or not. To complete the analysis we need to go deeper and study the structure of the network from the individual point of view and as a whole. There are many metrics for analysing social networks in the literature. We can mainly separate them in three groups that will be described below.

2.2 Connections

Since actors and the connections among them define networks, it is useful to start examining some properties related to the proportion between actors and connections. First, we focus on the network as a whole, considering the number of actors, the number of connections that are possible, and the number of connections that are actually present. Differences on the size of the network and how well connected are the actors are critical for extracting conclusions: it is very different to analyse a small network, where probably all actors know each other, that a larger network where we can find very well connected groups but also isolated groups or individuals. These differences will make the network more dense, cohesive or complex.

From the point of view of the actors, it is interesting to examine if they are receivers or senders of information (or maybe both). The number of ties that

actors have are keys for determining how much their embeddedness in the network constraints their behaviour, and the range of opportunities, influence and power that they have.

In order to have a first impression of how the diameter of the network, and how well connected are its nodes we can analyse the network in terms of density, reachability and distance (21). These three measures are, for example, very representative in the case of the small-world networks as we mentioned before.

The **density** of a binary network is the proportion of direct ties in a network relative to the total number. The density of a network may give us some information about the speed at which information diffuses among the nodes, and the extent to which actors have high levels of social capital and/or social constraint.

The **reachability** of an actor A by another B is given if we can trace any set of connections between them, no matter the number of steps we need. It simply tells us whether two actors are connected or not by way of either a direct or indirect pathways of any length. In the case of directed graphs, it is possible that A can reach B but B is not able to reach A. We can thus see if there are some actors in the networks that are more isolated, forming a division of the network. Density and reachability deal with adjacency among actors, but do not provide information on how embedded are people in networks. To capture this aspect, one main approach is to examine the distance that an actor is from others.

Distance is the number of ties required to connect to given nodes. But sometimes it may be interesting to find out how many ways there are to connect between two nodes at a given distance. Multiple paths may indicate a stronger connection between two actors than a single connection.

One particular definition of the distance between actors in a network is the geodesic distance. For both directed and undirected data, the geodesic distance is the number of relations in the shortest possible path from one actor to another. It is often considered the optimal or most efficient connection between two actors.

However, sometimes it is not about how quick a piece of information spreads but how often do you receive it. It is the case of rumour spreading, where the veracity of the information will depend on how many times the information arrives to a certain node from different sources, and not how soon he hears it. In this case, we need to take into account all the connections among actors.

2. SOCIAL NETWORK ANALYSIS

With these basic concepts on connections and distances, we can now talk about the second group of social network analysis metrics. We will focus the distribution of the actors in a network.

2.3 Node distribution

One of the main objectives of SNA is to determine how important a node is, or how much power and influence does a node have over the rest of the network. Many sociologists agree that power and influence are a fundamental property of social structure (29), but there is less agreement about what their definitions, and how we can describe and analyse its causes and consequences. Although they seem similar concepts, they are often confused. Power is the ability to force someone to behave in a particular way by controlling his outcomes. Social influence, however, is the process by which individuals make real changes to their feelings and behaviors as a result of interaction with others who are perceived to be similar, desirable, or expert (38). The kind and variety of information is directly related to a node's position within the network. It will also affect on the time needed to spread or receive the information. One of the main approaches of SNA to measure the influence and power of a node is the study of centrality, which focuses on determining whether a node is in advantaged or disadvantaged position with respect to the structure of the network.

2.3.1 Degree Centrality

The node degree represents how many connections a node has. In terms of centrality, nodes with more ties to others are in advantaged position for catching whatever is flowing to the network (like a virus, a rumour or some kind of information). This autonomy makes them less dependent than any other node of the network, and hence more powerful. In the case of directed networks, we need to distinguish between in-degree and out-degree centrality. The former one refers to the number of ties directed to the node and the latter one to the number of ties that the node direct to others. Nodes with a high out-degree are actors who are able to exchange information with many others, but also they are able to aware

others of their point of view. Usually, in-degree centrality is related to popularity and out-degree centrality is related to influence or gregariousness.

Counting the number of in and out ties we can say if an actor is more or less central, that is the centrality as an attribute of individual actors as a consequence of their position. But we can also consider a network as a whole that may have a group of central actors, which means that we can see how centralized the graph as a whole is, or how unequal is the distribution of centrality. Linton Freeman first proposed the general definition of centralization in 1979 (13).

Bonacich proposed a modification of the degree centrality approach (3). The original idea of degree centrality argues that actors with more connection are more powerful. Bonacich goes a step further saying that two actors having the same high degree does not necessarily makes both actors equally important. Bonacich's argument states that being connected to others who are also well connected makes you a central node, but not a powerful one. However, being connected to others that are not well connected (i.e. That are more isolated) makes one powerful, as these other nodes will depend on you, whereas well connected nodes are not as they can reach the information through many other ways. Bonacich proposed that both centrality and power were a function of the connections of the actors in one's neighbourhood.

2.3.2 Closeness centrality

Degree centrality might be criticized because it only takes into account the immediate ties that an actor has, or the ties of the actor's neighbours, rather than indirect ties to all others. One actor might be tied to a large number of others, but those others might be rather disconnected from the network as a whole. In a case like this, the actor could be quite central, but only in a local neighbourhood.

Closeness centrality approach emphasizes the distance of an actor to all others in the network by focusing on the distance from each actor to all others. In graph theory, the distance between a pair of nodes is defined by the length of their shortest path (13). The far-ness of a node is defined as the sum of its distances to all other nodes, and its closeness is defined as the inverse of far-ness. Thus, the more central a node is, the lower its distance to all other nodes.

2. SOCIAL NETWORK ANALYSIS

When a node is able to reach other nodes at shorter path lengths, or is more reachable by other actors at shorter path lengths, it is said to have a favoured position. Here power comes from acting as a *reference point*, as these kinds of nodes are closer to more nodes than any other in the network. Closeness can be regarded as a measure of studying how fast will a piece of information spread from the *reference point* node to all other nodes.

2.3.3 Betweenness centrality

Closeness centrality measures how quickly a node can reach all other nodes in the network, and so how easily can it access to whatever it is flowing through the network. However, it does not provide information on how much control a node has over what flows in the network, *i.e.* How often is this node on the path between other nodes.

Freeman introduced betweenness centrality as a measure for quantifying the control of a human on the communication between other humans in a social network. Some actors in a network exhibit a mediating role between other actors, which can be beneficial for them in terms of power. Burt emphasizes this idea of the powerful third-party by measuring the brokerage role of a certain actor employing betweenness centrality (5).

Consider actor A wants to reach actor B in order to influence him by sending information or make a deal to exchange some resources, but he needs an intermediary as it does not have direct access to B. All the people who lies between A and B have power with respect to A, but if A has other channels to reach B, then A is less dependent and thus more powerful while the others loose some power. Betweenness centrality of a certain node (its actor centrality) will be given by the proportion of times it is between other nodes for sending information and the number of falls in pathways between other nodes.

2.3.4 Eigenvectors centrality

The third centrality measure described by Freeman is the eigenvector centrality, which aims to find the the most central actors (in terms of smallest farness), but wihtin the global structure of the network, paying less attention to local

patterns of node distribution (13). The approach beyond this measure is the factor analysis, by which eigenvalues are defined: it identifies *dimensions* of the distances among actors. The location of each actor with respect to each dimension is called an *eigenvalue*, and the collection of such values is called the *eigenvector*. Usually, the first dimension captures the *global* aspects of distances among actors; second and further dimensions capture more specific and local sub-structures.

The idea is that even if a node influences just one other node, who subsequently influences many other nodes (who themselves influence still more others), then the first node in that chain is *highly* influential. Eigenvector centrality can be considered as a recursive version of the degree centrality: the node's centrality is proportional to the sum of centralities of those it has ties to. Eigenvector centrality is often related as a measure to define a node's popularity within a network. Google's PageRank is a variant of the eigenvector centrality measure.

2.3.5 Structural holes

Ronald Burt coined and popularized the concept of structural holes to refer to some important aspects of the positional advantage/disadvantage of the actors in a network, that results from how they are embedded in neighbourhoods. Burt says that people in a same environment interact by homophily, which means that relations are more likely between people who share socially significant attributes, such as age, education, or gender. From a structural point of view, people will interact in clusters, with dense relationships inside, which promote trust within teams and productive collaboration.

Burt states, however, that if we apply this to cohesive social networks, like offices or industries, we will observe that people will tend to form clusters, or groups, within the organization, reducing their relations to those clusters, and they will finally act and think the same. In the long run, this homogeneity will dead-end creativity. This does not mean that people from a cluster are unaware of the existence of people from other clusters. They just focus on their own activities, not attending others activities: they move in different flows of information. The physical space between clusters is what Burt defines as structural

2. SOCIAL NETWORK ANALYSIS

holes. Some of the individuals of a cluster may have, however, an external contact in the organization that could bring him to other new contacts, leading to an advantageous position with respect to the rest of the cluster.

The concept of structural holes describes how individuals who span two different clusters or groups can become powerful by brokering the relationships and information flow across the clusters. Those individuals are called brokers, and they build bridges across the whole creating a relationship for which there is no effective indirect connection through third parties. According to Burt, brokers will have some advantages:

1. Access to a wider diversity of information;
2. Early access to that information;
3. Control over diffusion

Actors in this position have the capacity to influence over other nodes that are not connected between them when deciding to pass a piece of information, or participating as a mediator (broker) for a negotiation or exchange of ideas. This will end to the creation of a leader of opinion, responsible for spread new ideas and behaviours. Managers who span structural holes often move quickly up the corporate ladder.

It should seem that the more you try to enlarge your social network, the more people you will have to deal with and the more information you will have to process: the better you are connected, the more valuable is your social capital. However this can end to eliminate the creativity benefits, as everyone will be more connected and there will not be new ideas to discover. The appearance of redundant ties among actors will decrease their power over the rest of the nodes in a neighbourhood.

2.4 Community structure

The metrics describes previously aimed to examine connected individuals and distances between them. In the study of complex networks, where topological

features of the resulting graph are not simple to analyse, we need to find some patterns of connection between the elements of the network from a point of view of social structure.

First analyses on this field proved that there are some groups of individuals in a network, which are better connected between them than with the rest of the population. If the nodes of the network can be easily grouped into sets of nodes, such that each set of nodes is densely connected internally, then we can say that they belong to a community structure.

From a sociological point of view, individuals interact by homophily, which makes humans tend to be attracted by other with social similarities (like age, gender, mutual friends or common hobbies). This would lead to non-overlapping communities (also called local clusters), with nodes having dense internal connections and sparse connections between groups, which was the initial aim of community finding. However, this may not be applicable in many cases, as in complex networks nodes can have many connections with other nodes that are also connected; the more connections they have, the more dependent they can be of the structure. In this case, the groups of nodes can be potentially overlapping. Being able to identify sub-structures within a network can provide insight into how network function and topology affect each other. There are several methods of community finding. We will briefly describe some of them that are related with this thesis.

2.4.1 Dyads and triads

The known smallest social structure in which an individual can be embedded in a dyad, that is, a pair of actors. For binary ties there are two possibilities for each pair in the population: either they have a tie or they do not. The density measure describe above classifies the whole population in terms of dyadic structures. In the case of directed relations, there are three kinds of dyads: no tie, one likes the other but not vice-versa, or both like the other. Taking into account the amount of reciprocated ties in a network may tell us about the degree of cohesion, trust and social capital present in the network.

2. SOCIAL NETWORK ANALYSIS

The smallest social structure that has true character of *society* is triad, that is, any triple A , B , C of actors (29). A triad basically consists of the union of three dyads. A principal interest in the study of triads is the phenomenon of transitivity. A triad is transitive if there exists a tie between A and B , B and C , and A and C . For example, if ties between A - B and B - C exist, but no between A - C , then the triad is intransitive. In this last case, actor B may serve to play a role as an intermediary of the relation between A and C , helping in solidifying an alliance or mediating a conflict. We can say that B has a power from his brokerage position.

Holland and Leinhardt first proposed the now standard MAN notation for triads: mutual (M), asymmetric (A) and null (N) dyads in each triad (23). They define a triad census as the combination of all the relations across all the possible triples. This census can give a good sense of the extent to which a population is characterized by isolation, couples only, structural holes or clusters. With undirected data there are four possible types of triadic relations (no tie, one tie, two ties, or all three ties). With directed data, there are 16 possible types of relations among three actors according to the MAN notation. These kinds of relations exhibit hierarchy, equality and the formation of exclusive groups, see figure 2.1 (12).

Triads have been widely used for analysing the structure of social networks, as fundamental forms of social relationships can be extracted from them. In particular, they may be useful for studying transitivity, clustering and structural balance of a network. A related measure to triads is the clustering coefficient, which examines the local neighbourhood of a node. The clustering coefficient is a ratio N / M , where N is the number of edges between the neighbors of n , and M is the maximum number of edges that could possibly exist between the neighbors of n . The clustering coefficient of a node is always a number between 0 and 1.

2.4.2 Hierarchical clustering

A cluster is a collection of individuals with a dense friendship pattern internally and a sparse friendship pattern externally. There are several kind of cluster algorithms, returning each one a cluster model with particular properties. For the case

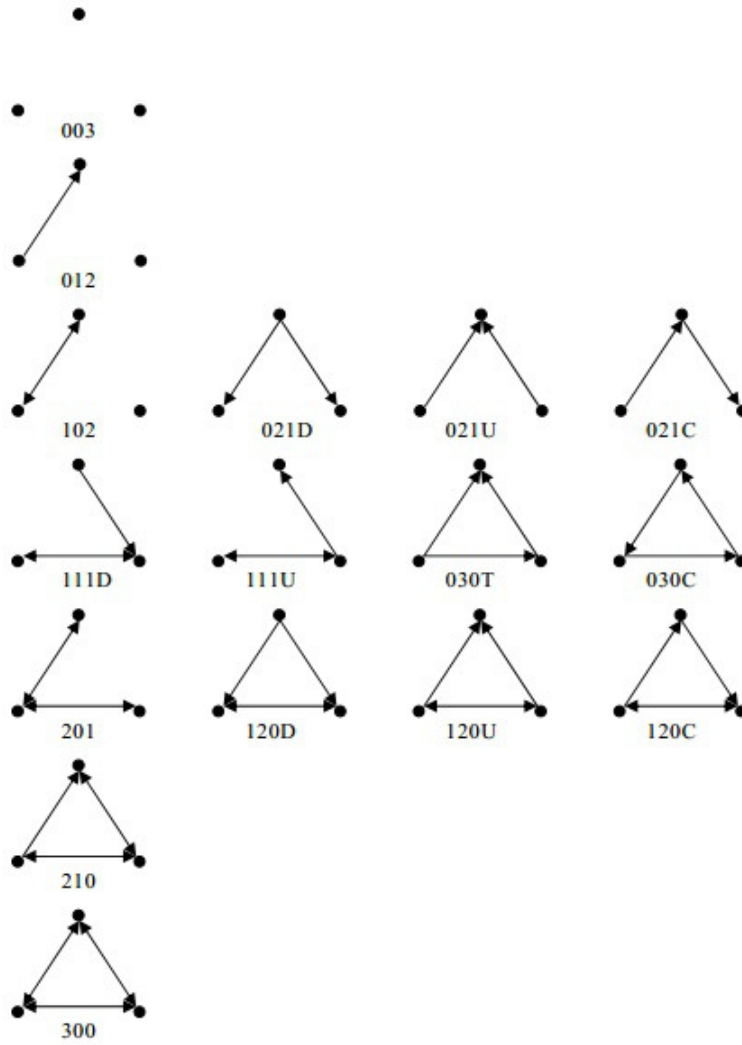


Figure 2.1: Isomorphism classes with MAN labelling

2. SOCIAL NETWORK ANALYSIS

of social network analysis, researchers use connectivity models like hierarchical clustering.

Hierarchical clustering groups nodes according to some similarity measures (29). These measures are usually related with the approximate equivalence, which focus on topology of the network (a measure of distance between pairs, like Euclidean or Hamming distance) or structural equivalence, where two nodes are said to be equivalent if they have the same set of neighbours (17). Hierarchical clustering can be either bottom-up (agglomerative) or top-down (divisive).

Hierarchical Agglomerative Clustering (HAC) method is based on assigning a weight for every edge and placing these edges into an initially empty network, starting from edges with strong weights and progressing towards the weakest ones. The edges with the greatest weights within the community are the most central ones. Typically HAC is represented by a dendrogram; by moving up from the bottom layer to the top node, a dendrogram allows us to reconstruct the history of merges that resulted in the depicted clustering. As a disadvantage, this method results slow for large networks and presents an inability to classify in a community a node, which is connected to the network with only one edge.

2.4.3 Girvan-Newman algorithm

This algorithm is based on Hierarchical Divisive Clustering, and works on the opposite way of HAC: at first step, all nodes belong to the same cluster and it recursively split them into clusters until we have one cluster per node. The Girvan-Newman algorithm focuses on these edges that are least central, the edges that are more *between* the communities. The communities are detected by progressively removing edges from the original graph, which are identified by adapting the graph-theoretic measure *node betweenness* of Freeman to *edge betweenness* (19). The edge betweenness of an edge is defined as the number of shortest paths between pairs of vertices that run along it. As a disadvantage, it runs slowly, making it impracticable for networks of more than a few thousand nodes.

2.4.4 Clique based methods

Cliques are sub-graphs in which every node is connected to every other node in the clique. As a node can be part of more than one clique, a node can then be member of more than one community giving an overlapping community structure. Even if recent approaches to community detection in networks are based on finding cliques and studying the overlaps, other refuses to use it for this latter reason (47).

One approach is to use **maximal cliques**, which is, given a minimum size of nodes, find the cliques, which are not the sub-graph of any other clique. The union of these cliques then defines a sub-graph whose components (disconnected parts) form communities.

The alternative approach is to use cliques of fixed size k , k -cliques. The overlaps of these cliques can be used to define a k -regular hyper graph, also called Clique graph. Applying any of the community detection methods to the clique graph would assign each clique to a community. This can be used to determine community membership of nodes in the cliques, and detect which nodes belong to more than one community.

The **clique percolation** method (CPM) defines communities as percolation clusters of k -cliques. The CPM finds all the k -cliques in a network. Two k -cliques are considered adjacent if they share $k-1$ nodes. A community here is defined as the maximal union of k -cliques that can be reached from each other through a series of adjacent k -cliques. The definition above is also local: if a certain sub-graph fulfills the criteria to be considered as a community, then it will remain a community independent of what happens to another part of the network. CPM results a useful tool to identify cohesive groups, as it takes the overlap as a starting point to identify cliques. By relaxing clique membership in favour of clique adjacency, hence capturing group overlaps; CPM achieves greater sociological realism that allows for closer approximation to the notion of community than does the concept of a sociometric clique (47)

The CPM method allows detecting which are the nodes that can be really important to keep a community communicated. If we take the figure as an example, we can see that node 1 has a higher node degree that the rest of the nodes in the

2. SOCIAL NETWORK ANALYSIS

network. At the same time, he is close to many of the nodes, and he is placed between many of the paths of the network. However, the nodes connected to node 1 are also connected between them forming a chain. If we remove node 1 from the network, the nodes connected to node 1 will experience a variation in their centrality values, but will be able to keep connected to the network. It results obvious that communication will be less efficient than before, especially for node 2 that is placed at the end of the chain, but no node will remain isolated.

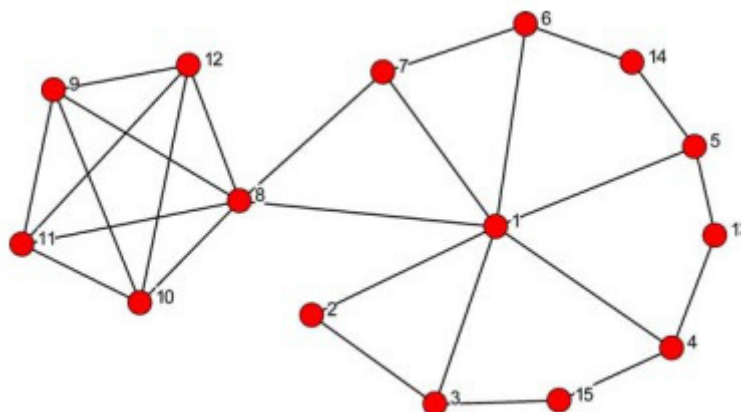


Figure 2.2: Example of a network

Let us focus on node 8 of figure 2.2: it has also a high node degree but is less central than node 1. However, it is in a more advantaged position than node 1 as if we remove it, the network will be fragmented into two isolated communities. This example belongs to (4).

2.4.5 Quality measures

Once we have detected the communities within the network, it is interesting to evaluate how good is the partition. In general terms, we seek for clusters with dense intra-connections and sparse inter-connections. There are several ways to calculate the goodness of a cluster, but the two most popular are modularity and conductance.

Modularity is a benefit function that measures the quality of a particular division of the network into modules or communities. For a given partition of a network into clusters, modularity measures the number of within community edges, relative to a null model of a random graph with the same degree distribution. Even if this method is very used for optimization in community finding, it has been shown that it suffers a resolution limit and, therefore, it is unable to detect small communities (20).

Another commonly used measure is conductance. It can be thought of as the ratio between the number of edges inside the cluster and the number of edges leaving the cluster.

2.5 Complex networks

Classical models of networks, called random networks or random graphs, share the assumption that the connections between units occur at a random process. A random graph is obtained by starting with a set of n vertices and adding edges between them arbitrarily with a probability p . Different random graph models produce different probability distributions on graphs. Usually, generated networks had a majority of nodes with similar number of connections following a Poisson distribution. One of the most studied methods for generating random graphs is the Erdős-Rényi model.

With studies now focused on real-world networks, such as computer networks or social networks, and the access to huge network data resources, researchers start turning interest on the study of complex networks. The main difference is that connections between nodes in a complex network do not occur at random nor are regular, but have instead a non-trivial topology given by the kind of connections among nodes.

One of the most well studied classes of complex network is the small-world network by analogy with the small-world phenomenon (also known as *six degrees of separation*) (48). A small-world network is based on the notion that there are only 6 degrees of separation between any two people in the world, and thus, it does not take many hops to get from one node to another in such networks. In many large networks (like, for example, the Internet) the average geodesic

2. SOCIAL NETWORK ANALYSIS

distance between a pair of nodes is relatively short. The *six degrees* phenomenon is an example of this: most of the nodes, even in large networks, may be fairly close to one another. Moreover, the distance between pairs of nodes in a complex network is often shorter in average than the distance between pairs in random graphs.

While trying to model the World Wild Web network expecting to find a random graph topology in 1999, Barabási, Jeong and Rekka noticed that the majority of the nodes of the network were very low connected, and, by contrast, there were some nodes of very extreme connectivity (also called hubs) (39). In mathematical terms, this is a power-law distribution (see figure 2.3). A network is considered a complex network if the degree distribution follows the power-law distribution.

The Barabási-Albert model explains the power-law degree distribution of networks by considering two main features in the algorithm: growth and preferential attachment. Growth because at each time step the number of nodes increases in the network. Preferential attachment refers to the fact that new nodes tend to connect to nodes with large node degree, which in the end tends to a power-law distribution. Networks presenting these characteristics are named *scale-free* networks.

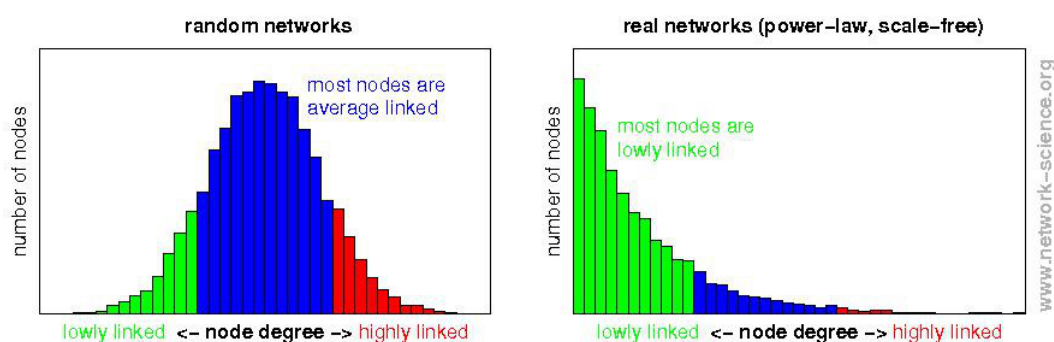


Figure 2.3: Random networks VS scale-free power-law networks

The figure 2.3 shows a comparison between random and scale-free networks. As we said before, we can observe that in random networks most nodes have a medium node degree. On the other hand, real networks often show a skewed node degree distribution in which most nodes have only few links but, by contrast, there

2.5 Complex networks

exist some nodes, which are extremely linked. An example of this behaviour is the Internet, which had only handful routers three decades ago and has gradually grown up to millions. New routers will link to those that were already part of the network, giving the opportunity to older nodes to acquire new contacts. Figure 2.4 shows the evolution of a scale-free network following growth and preferential attachment; we can observe how these two mechanisms lead to the appearance of hubs.

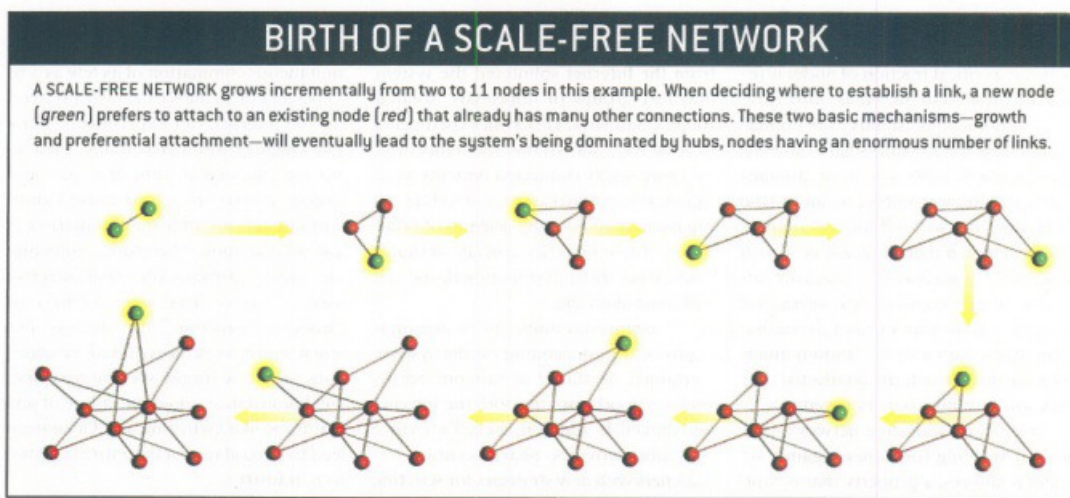


Figure 2.4: Birth of a scale-free networks

Scale-free topology has been widely used in several fields of investigation, from communication networks like the World Wide Web, to biological such as human brain (10), human sexual relations (28) or the study of virus spreading (43) (in this case both biological and computational). Other popular examples are the study on the time evolution of scientific collaboration networks by Newman (34) and the structure of co-authorship of scientist networks by Barabási (1).

2. SOCIAL NETWORK ANALYSIS

Chapter 3

State of the art

Social networks are prevalent in our society. The study of social network structures started almost one century ago, with the first studies from people working in educational and development psychology (14), which is still a current field of research (27). Jacob Moreno did one of the first studies related with SNA during the 1930s. He coined the term *sociometry*, a quantitative method for analysing interpersonal emotive relationships within a group. His methods have been used to identify informal leaders, social rankings and isolated individuals. Thus, sociometry is a tool to measure the degree of relatedness among people. Sociometry is based on the fact that people make choices in interpersonal relationships. Whenever people gather, they make choices where to sit or stand; choices about who is perceived as friend and who is not, who is central to the group, who is rejected, who is isolated.

Moreno studied the interpersonal relationships as structures where a person is represented as points and relationships between them are drawn as connecting lines: this is what we know today as a *sociogram*. One of his earliest images is shown in Figure 3.1 (31). He characterized that image as showing

a group where two dominating individuals are strongly united both directly and indirectly through other individuals

Moreno viewed that picture as a display of both cohesiveness (*strongly united*) and social roles (*dominating individuals*).

3. STATE OF THE ART

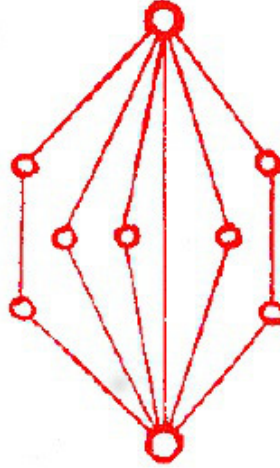


Figure 3.1: Moreno's earliest *sociogram*

Moreno defined five terms, or factors, to deal with the quantitative evaluation of an individual's role in a group or community (33):

- the *tele* factor, or the distance between two people that feel attraction or rejection between them.
- the *spontaneity*, or how would an human being respond to a new situation, starting from the moment of its birth. This degree of *spontaneity* will mark the level of creativity of each individual (32)
- the *social atom*, or the sum total of relationships created by the feelings of like and dislike.
- the *group formations* or *coteries*, or the patterns of attraction and rejection relationships formed in a geographic environment (like a community, a school or an institution). Studying those patterns we can find isolated groups, forming cliques, or isolated individuals, but we can also identify individuals that dynamise interlinks between different social atoms.
- the *psychological networks*, or how these interlinks brings to chain formations, converting every individual of the chain a potential influencer in the rest of the group.

One of his earliest works studied the friendship relations among fourth grade students. He used triangles to represent boys and circles for girls. He also used directed lines with arrowheads to show which child was the *chooser* and which one the *chosen*. He aimed to demonstrate that variations in the location of points could be useful to stress important structural patterns of the data (15). The interesting thing is that Moreno decided to separate geographically the circles and triangles, expecting that most of the relationships would be same-gender (as it is normal in this age). The result is shown in Figure 3.2, where we can perfectly appreciate the gender division and the group and subgroup formations, but we can also observe a couple of boys and girls which are better connected than the rest (in fact, among them there is the only friendship relation among different genders). On the top-right, we can also observe isolated girls that connect among them but not with the rest of the classroom.

Moreno's problem is that he did not develop any procedure to systematically locate points in images, but he did it every time specifically for the problem. With the large size networks that we deal nowadays this would be impossible. However his first approach to social network analyses is considered as a base in the literature.

We can appreciate that Moreno already wanted to separate the networks in some kinds of groups. This has been one of the major interests in the SNA field, as it gives an idea of how the network is structured.

In §2.4 we have mentioned some algorithms that have been developed in order to detect community structures.

The most widely known and used is the Girvan-Newman algorithm as it is quite efficient in large size networks (19). Much work has been developed also, trying to improve Girvan-Newman. Duch and Arenas proposed an algorithm to detect communities by applying an extremely optimization to the modularity measure. However, Pujol *et al* showed that it failed for very large networks (36). Moreover they propose an algorithm, called PBD, using hierarchical clustering that outfits Newman-Girvan and Duch-Arenas results (9), even in very large networks (reducing the number of clusters 500 times). In this thesis we execute the PBD algorithm for the dataset used in order to have a first idea of the number of clusters in the network. Unfortunately, the results are only numerical, making

3. STATE OF THE ART

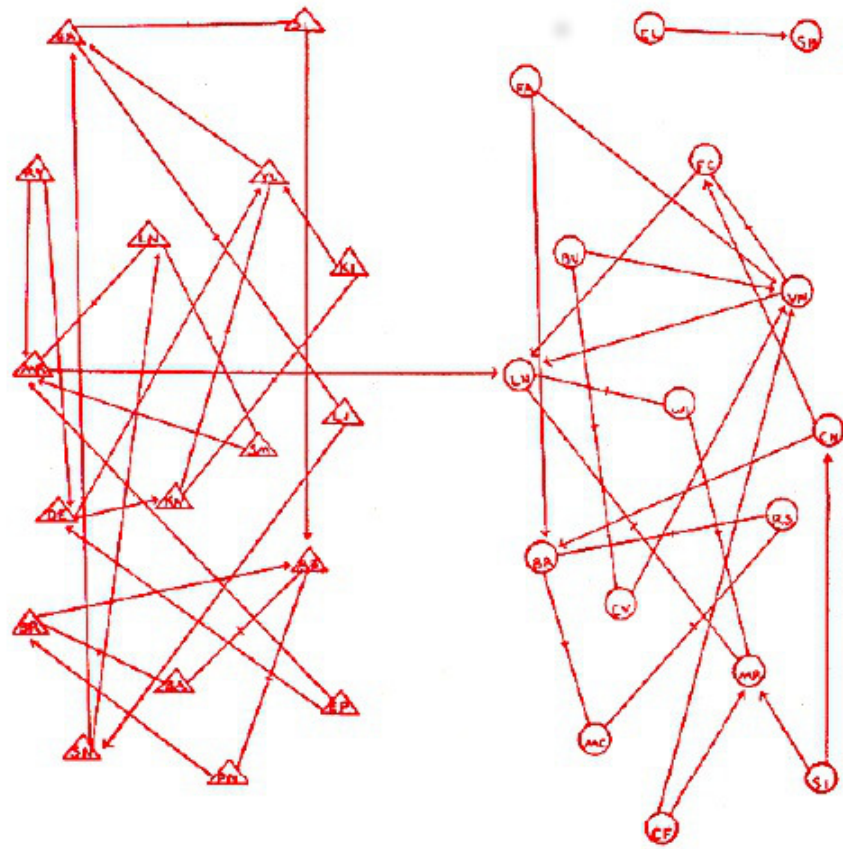


Figure 3.2: Moreno's sociogram of fourth grade friendship

difficult to work with them for a large network. Another main problem is that PBD focuses on non-overlapping communities, which is not applicable to all networks.

As we said before, it is not only about dividing the network into communities, we then have to evaluate them using criteria to qualify communities in a network. Modularity and conductance are the most popular functions, although scholars have proposed several modifications or adaptations depending on the desired result (30).

However, focusing only on the general structure of the network does not provide enough information to detect which may be the most influential nodes. Detecting communities allows us to *tag* nodes into clusters, and observe the shape of inter and intra-connections. Now we need to combine it with the information extracted from graph-theoretic measures, in order to analyse how well-positioned or isolated is a node with respect to the network. Scholars have taken this analysis to many different areas of research, and depending, on the field, they have focused the study in a manner or another. For example, in the study of *viral* or rumour spreading the goal is the efficiency in broadcasting an information; the redundant information factor is not a deal in this case, because the more you receive the information from different sources, the more viral the information is considered.

When analysing collaboration or entrepreneurial networks, scholars focus on the source of the information: which actors are able to reach others in a exclusive manner, in order to have the influence to negotiate (or collaborate) with others. Actors in this position are said to have a higher social capital. From a sociological point of view, we can directly relate an individual's social capital to the *closure* of the community, which is mainly given by trust and reputation. The analyse may be complemented some information about sociological, educational or economical background in order to establish an evolution over time of this individual within the network (7). This is usually given in closed communities, such as neighbourhoods or, the previously cited collaboration networks. In the case of entrepreneurial networks, this idea conflicts with Burt's idea of brokerage (§2.3.5) but at the same time they are interdependent (6).

In counter-part of this idea of spanning relationships over the network, which tends to represent a conflict, Stark and Vedres propose the concept of structural

3. STATE OF THE ART

folds, which benefits from the trust given in closure and the creativity given by brokerage: they try to identify actors that are not only able to spread valuable information or negotiate with it, but that may also be able to generate knowledge (47). The study focus on a temporal analysis of the evolution of members in boards of directors in Hungary, and to detect the actors in the structural folds they use the Clique Percolation Method described in §2.3.3. However, they observe that this intercohesion between groups (or boards) are disruptive, that is, groups break down more often if one or more of their members takes on multiple affiliations. As we said before, if an actor located in the overlap between two communities (the structural fold) is removed, then two isolated groups are formed.

Other studies have been carried out relating boards of directors from different approaches. In general, they analyse the social capital and the influence of members of boards according to economic benefits, and some historical factors like the permanence in a same board or affiliation to other boards, both in the past or simultaneously in the present (44).

From a structural point of view, many hypothesis have been proposed regarding a board of director's composition. In those studies, board size (46), the existence of outsiders and the proportion of outsiders on the board were frequently hypothesized to have a direct affect on firm performance (8). However, no consensus has arose from these works. Gilley *et al* develop and approach by mixing both boards' composition and historical factors of each member of the board to define influence and try to correlate it with firm's performance (18).

Boards of directors have been usually male dominant, but due to important changes in come countries' legislations involving gender equality representation, women are now more present these boards. A recent interest in last years works have focused on the effect of the introduction of women as board members from several viewpoints. One of the first studies relating the consequences of equality in gender representation is carried out by Seierstad and Opshal on the Norwegian legislation (see (§4.1)) taking a statistical approach. This work has been a reference for a set of studies (40),(41),(45), mostly from a sociological, economical or political viewpoint to analyse the evolution over time of gender quotas.

Finally, Hawarden and Marslan present an approach using social network tools to locate women directors in boards and examine the persistence of director networks over time to determine whether gender related differences apart from size contribute to the apparent resistance to change (22).

3. STATE OF THE ART

Chapter 4

Methodology and data

4.1 Data

The data used in this work consist on a list of 384 public limited companies in Norway that are available online through the Norwegian Business register on August 5, 2009, and over 5000 directors affiliated to these companies. Authors chose these companies, as they are the ones bound by the gender representation law (41). This law tries to ensure equality on the representation of women on Public Limited Companies BODs and, in general, all Boards of Directors in order to increase the influence of women in society. But, as the authors state, increasing the number of women does not directly derive in having more influence. Thus, equality should be understood at an influence level and not only at a representation level.

The BODs is central to corporate governance, it is the prime decision making body. An important feature of such boards is that they are often connected to each other by a shared director. Such network connectivity has important economic consequences. Research studies of company managers and directors look at multiple directorship holders just as *interlockers*; people who create linkages between corporations (11). This makes BODs, in our case, a very appealing object of study.

4. METHODOLOGY AND DATA

4.1.1 Background

Historically, there has always been a major representation of men in BODs. In the specific case of Norway, even if the proportion of women with studies on the tertiary sector is higher than for men, there was great number of male-dominated companies. Experts argument that introducing women on boards of directors would have an organisational advantage, as it implies introducing new perspectives, work styles, attitudes, interests, *etc.* All these factors will contribute to new thinking and ways of solving problems, which could result in higher productivity and a better working environment. The authors expect that the number of women in boards of directors will continue to rise, even after the end of the implementation period of the gender representation law, increasing above the minimums.

On the other hand, the authors notice that there is a minimal presence of women as chair of the board (41). Together with the CEO of a company, they are the two most important persons in a company (which, again, does not necessarily derive into influence). These two roles are not affected by the gender law, however the authors expect that with the introduction of women into boards of directors, they will progressively have the opportunity to occupy chairs of boards.

Participants from the BODs group are separated into three levels. Two of the levels include directors of Public Limited Companies BODs. Level one includes women in senior managerial positions that are directors of company boards, executive directors. Some of the participants also have previous experiences from non-executive BODs. Level two includes directors (non-executive) on Public Limited Companies BODs as well as having senior responsibility. At level three, the participants are directors (non-executive) of more than one BOD. This thesis includes directors who are members of two, three, four, five and more than eight Public Limited Companies BODs. The reason why level three includes both women being members of two BODs and women being members of eight or more BODs is to ensure the womens anonymity; as very few women belong to these categories, in order not to reveal identities, the group is broad. The Scandinavian model for BODs is characterised by a one-board, two-tier system. This means

that a single BOD exists, which in turn, is 125 composed of shareholder representatives and employee representatives. In addition, a second tier exists where a managing director or CEO is delegated the day-to-day running of the organisation. The boards duties, and sources of influence, include: defining the companys purpose and broad objectives, selecting, appointing, supporting, and evaluating the chief executive; providing advice; making ties with other organisations; financial stewardship; and monitoring and evaluating performance. To ensure that the senior management and the board do not overlap, the chief executive officer cannot be the chair of the board in Norway. Moreover, the employees are responsible for electing one third of the board members in firms that have more than 50 employees (40).

Figure 4.1 shows that there is a increase in the number of women in boards of directors, starting in 2005 (which is the year where the legislation started). This behaviour continues until 2008, where the implementation period ends. We can observe that the proportion of women representation stays stable at 40%, which was the minimum required. Thus, the companies accomplished with the law, but did not go further. Surprisingly enough, even if the proportion of women in boards of directors increases, the number of women occupying the chair of the board does not change during all the years registered.

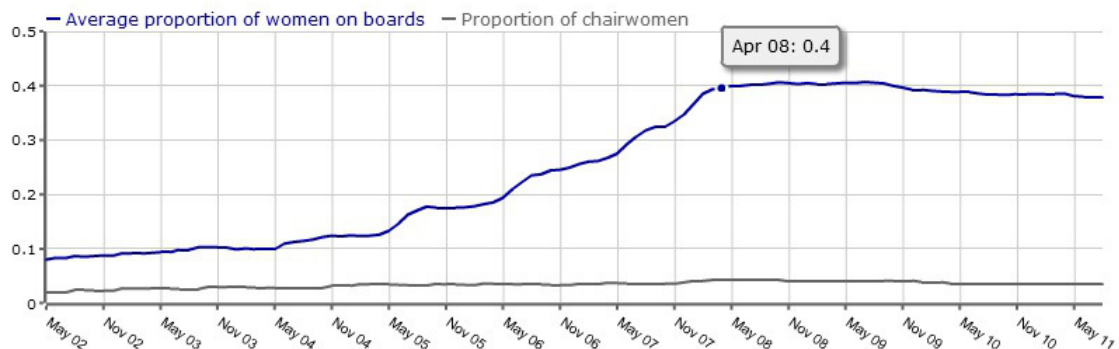


Figure 4.1: Average proportion of women on boards and chairwomen.

4. METHODOLOGY AND DATA

4.1.2 Dataset information

This dataset and all the information about the study are available on the Internet¹ (41),(40). The website provides a file containing basic information about the companies, such as an identifier, the organisation number, the full name of the company and the postcode and city registered. There is also a file containing information about the directors, such as an identifier, the name and the gender. This people is extracted from the boards of directors of the companies, excluding employee representatives, as the Norwegian legislation does not affect them in the same way.

In addition, we have a monthly report from August 2002 to August 2011 of all the affiliations to companies, and all the relations between directors. Authors anticipate that it is possible that a person can be part of more than one board of directors in a same month. For this work, it will be very interesting to analyse the social structure of the boards in order to detect some correlation among months to identify the most central nodes, and the most influential nodes. As we do not have more information about the population included in this study (like age or university studies), nor information about the companies benefits, we will focus on a structural analysis of the network, taking only into account the relations among directors and the changes on affiliations from a chronological point of view.

4.2 Methodology

In §2, a review of the most popular SNA tools and software has been made. Pujol *et al* (36), or the Girvan-Newman (19) algorithm focus on splitting the networks into communities, but do not provide any information about the actors as individuals. Moreover, they focus on separating the network into non-overlapping communities, which is not always applicable or useful. Centrality measures are one of the most important and widely used for analysing social networks, however we cannot detect really influential actors by only taking into account these measures. Clique Percolation is very useful in identifying the set of key nodes

¹<http://www.boardsandgender.com/data.php>

which, if removed, may lead to the isolation of some sub-groups, but we need to combine it with other tools in order to analyse the network distribution.

In this thesis we will combine some of these tools and compare them in order to select the set of nodes that play a really key role in the social network: thus the ones that may have a higher influence on the rest of nodes and may lead to a quick or efficient spread of information.

We will perform an statistical study of the dataset, which is divided in 111 files (from May 2002 to July 2011). We will extract the number of companies involved each month and trace for new incorporation in the boards of directors, or any possible change on the boards. We will also register the number of women participating every month.

We will then start the network analyse using the analysis tool **SNAP** (42), which is a general purpose network analysis and graph mining library developed by Stanford University. **SNAP** provides many different SNA metrics, both in a visual and text format. We will first focus on results extracted from community detection and the state of the network or **netstat** . This latter includes information about the diameter of the network, average clustering or in and out degrees. This will provide a first overview of the structure of the network as a whole, and its evolution during the examination period.

We will then focus on a second set of measures related with node centrality. **SNAP** provides information of several SNA metric related with the node, including betweenness, closeness and page rank, which will be the metrics used in this thesis. Again, for every month included in this study, and every nodes participating in a given moth, we can extract the values of each of the measures mentioned above. As one of the objectives of this thesis is to detect the most influential nodes, we will rank the nodes for each metric in order to get a list of the 25 most important nodes. For each metric, this list is obtained by taking the first 25 nodes with highest metric value for every month and then merge all the resulting lists in one. We will first consider the global ranking by the 25 nodes that are repeated more times in the global list, i.e. those who have a largest temporal participation in the network. The objective is to find a relation between the tree final lists (the top 25 rankings for betweenness, closeness and page rank) in order to check for

4. METHODOLOGY AND DATA

repetitions of nodes among lists. On the other hand, we also want to extract information about the gender representation in these lists.

This first analysis provides some quantitative data of the important nodes' centrality, in a general manner but does not provide any information about the evolution over time of these rankings. For this, we use Kendall's coefficient of concordance of ranks (24). The idea behind this algorithm is that n subjects are ranked (1 to n) by each of the rankers, and the statistics evaluates how much the rankers agree with each other. The obtained result is presented through a degree of similarity between -1,1 (being -1 not similar and 1 equals), between the two sets of ranks. We will visualize the evolution of this coefficient measures over time in order to detect possible sharp changes, which may be translated to changes of positions between nodes in the ranks, or new incorporations to the lists.

Unfortunately, Kendall does not provide such information, so we will try to reflect in graphically. This will be done through the use of Gephi (16), (2), an open source platform for visualizing and exploring all kinds or complex systems. This tool has a limitation with the network size, as very large networks are hard to be analysed only through graphic data; however its friendly user interface allows to take an overview of the network and apply some SNA metrics in order to detect important nodes in a visual manner.

By merging the results of these different platforms and taking into account the whole period of time considered by Opshal and Seierstad plus two more years (41), we aim to obtain: The evolution in the structure of the network, from the point of view of the relationships among directors. How does affect the incoming of women to the general structure and to the relations among directors and, *how* much are reflected these incoming of women into the ranking of most influential directors of the network.

Chapter 5

Results

5.1 Structuring the data

In this thesis we analysed the temporal evolution of gender representation in BODs according to the Norwegian gender representation law, based on the previous study carried out by Seierstad and Opshal (41). Their main objective was to determine whether the incorporation of women into BODs, and thus a more equal gender representation, would affect the network in terms of relations and influence. They looked at three parameters in addition to representation of women on corporate boards: the sex of the chair person, the emergence of and sex of prominent directors and director's social capital. As explained in 4.1.1, prominent directors are considered as those directors affiliated to more than one board at the same time.

In this work, we will go a step further by trying, not only to get conclusions about the relation between influence and gender representation, but also considering which may be the most influential nodes over the time. The work is divided in two parts: we will first analyse the evolution of the structure of the network over time from taking different variables as measures. Then we will define a ranking of the 25 most influential directors according to centrality measures.

For the first part, we have collected, for every month, the proportion of gender representation and some information about the structure of the network: number of nodes and edges, number of communities as well as the number of open and closed triads, the average clustering, the diameter of the network and the node

5. RESULTS

degree. Although the network is represented by a directed graph, we have checked that the relations are symmetrical, so in this case we do not need to differentiate between in and out degree.

The previous study showed already that the female representation increase, especially since the introduction of the gender representation law (see figure 5.1). However, there was an already existing network of contacts, according to some interests or business background. We may assume that directors that relate among them have already establish a degree of *trust*, which is higher among members of same boards due to the closure of the group, but may be also relevant among members of different memberships as it offers suitable situations for negotiations.

5.2 Data Analysis

Our first hypothesis is that the introduction of women into boards will disrupt this confidence, as they are seen as intruders in a society already established.

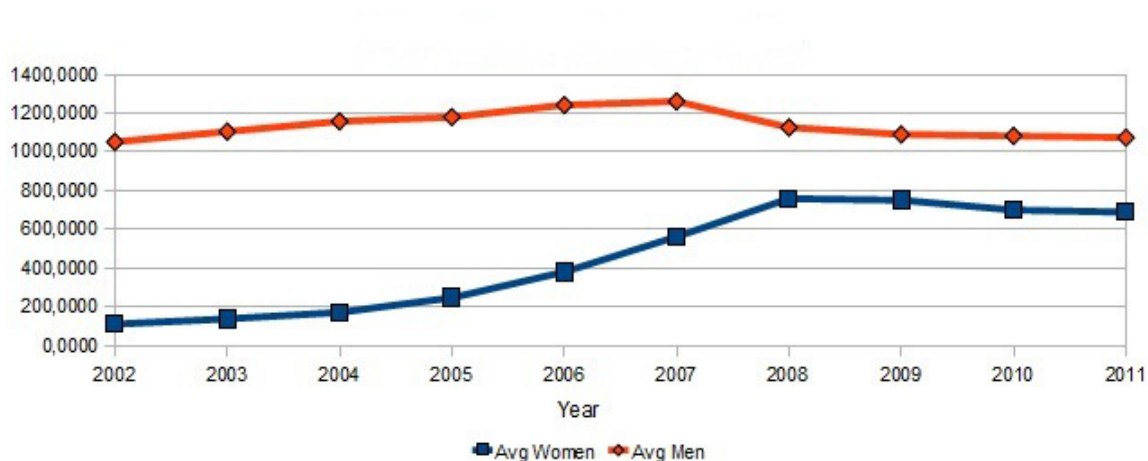


Figure 5.1: Evolution of the annual average gender representation

This is reflected in the relation among open and closed triads compared with the evolution of gender representation. As we said in §2.3.1, a closed triad is the representation of a triplet of actors that are fully connected (*i.e.*, that form a clique) and establish *strong* ties. The presence of closed triads can be related to

Burt's concept of closure of network (see §2.3.5) All other combination of relations among this triplet is considered as an open triad which leads to weak ties (but also to brokerage roles).



Figure 5.2: Evolution of the annual average of the number of edges and triads

For the sake of displaying the data, we show the annual average of all the metrics in order to facilitate the comparison between variables. In the Appendixes A and B, the complete graphs by months are given. If we contrast figure 5.1 with figure 5.2, we can establish some similarities: the number of women on BODs start increases faster from 2005 to 2008, which coincides with the period of the law enforcement, as it was expected. We can also appreciate that men on boards begin to lose representation in 2007. We assume that introducing women into boards probably implied the displacement of men already participating on those boards, so this behaviour is consequently expected, too.

However, if we take a look at figure 5.2, we can appreciate that the shape of open triads evolution is strongly related to female representation, and the same happens between closed triads and male representation. The number of closed triads measures the closure in a network, which is inherently related to trust and reciprocity (35). We can say that the increasing presence of women on boards interrupts in the relationships among already present members in the network, leading to a more open network. We will discuss later whether this loss of trust

5. RESULTS

is either positive nor negative in BODs environment from an economic and social point of view.

This is supported by other characteristics of the network's structure. In figure 5.3 we observe and increase in the number of communities, which starts at 119 number of clusters for 1038 nodes, so clusters are formed by clusters of size 10 to 100 nodes approximately. This average agrees with the idea of Lescovec *et al*, which states that best communities are relatively small with size up to 100 nodes; communities with higher sizes tend to get worse community quality (26). Until 2005, the network is progressively growing with the incorporation of both men and women until 2005 (with a major presence of male representation). Again, the plot represents an annual average but we can appreciate that the maximal number of communities is given after the beginning of the law enforcement. Directors regroup in smaller clusters in order to keep their strong connections inside the community.

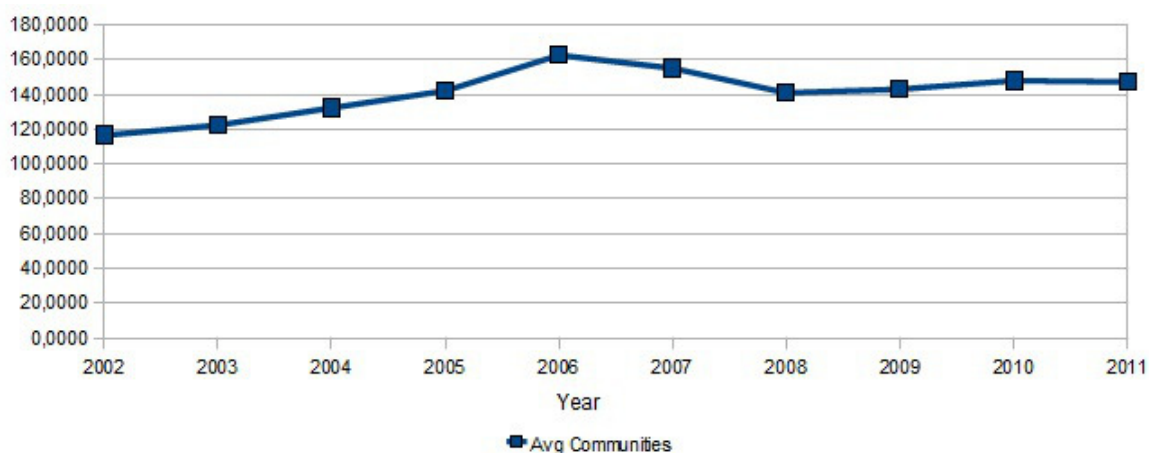


Figure 5.3: Evolution of the annual average of the number of communities

However, as a consequence of the law enforcement, we see how in 2006 the number of communities increases abruptly, up to 162 communities. Again, the plot represents an annual average but we can appreciate that the maximal number of communities is given after the beginning of the law enforcement. Directors regroup in smaller clusters in order to keep their strong connections inside the

community at the beginning of the law enforcement. However, the number of communities decreases again and get more or less stabilized from 2008. The proportion of number of clusters in relation with the number of nodes is maintained, although the number of nodes inside the cluster increases slightly. This reinforces the observation that the incorporation of women has lead to more open groups.

We have also studied the evolution on the average clustering coefficient. Surprisingly, it remains during all the months quite constant and with high values (above 0.9). Figure 5.4 shows the cumulative clustering coefficient (CCF) for years 2002, 2004, 2006, 2008 and 2010. The first two snapshots correspond to a very closed network, where everyone has many connections and corresponds with the period of time where there are more closed than open triads (see figure 5.1).

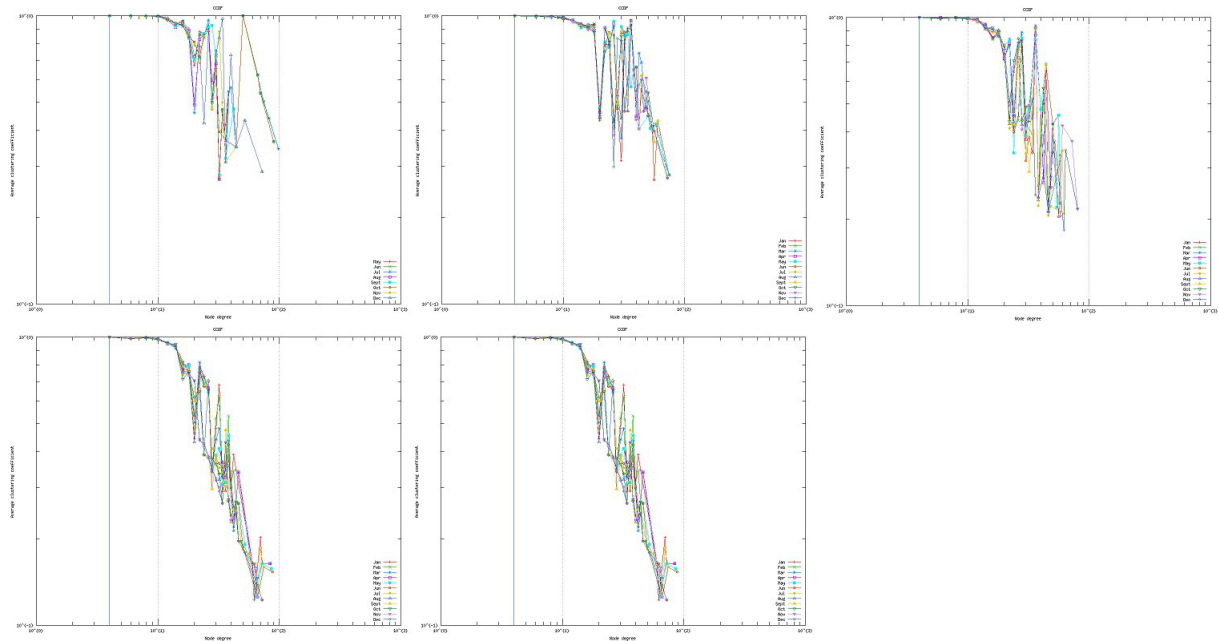


Figure 5.4: Evolution of cumulative clustering coefficient

From 2006, the CCF present a more exponential shape, tending to reduce the clustering coefficient for nodes with high degrees. However, the plots present a high oscillation between months: this may be because of the constant addition of

5. RESULTS

nodes and edges every month, so CCF will confirm the behaviour of the network but can not be considered as a significant metric for this analysis.

By now, we already know how are nodes distributed into clusters, but we need to determine how are the connections among these clusters. For this, we analyze two fundamental parameters which are the node degree and the diameter of the network. Although node degree is a centrality measure, we have included it in this previous part of the study as it does not provide information about the node in relation with the network, but only its in and out ties, so we consider it is not significant enough to state influence.

Almost all real-networks evolve over time by the addition or removal of nodes and edges. Most of the recent models of network evolution capture two phenomena:

- Constant average degree assumption: the average node degree remains constant over time.
- Slowly growing diameter assumption: The diameter is a slowly growing function of the network size, as in small world graphs (48).

Leskovec *et al* , however show that this is not the case for networks studied from a temporal viewpoint, instead of typical static one (25). They prove that these assumptions are not applicable for evolutive networks as the average node degree increases and the diameter of the network tends to decrease (what they call shrinking diameters). We will take a look on our network over time to see if these assumptions are given or not. We measure the diameter of the network by taking the effective diameter of the graph, that is the minimum number of hops needed to connect any pair of nodes of the network (or, in other words, the shortest paths from a given node to reach any other). Figure 5.5 represents the annual average of both the effective diameter of the network and the node degree. As we said before, relations are symmetrical, so we do not need to separate between in and out relations.

We can observe at first sight that the assumptions of Leskovec *et al* are not given in this network. In the case of the diameter of the network, we obtain a very irregular plot. Although the range of values is not very wide, we cannot

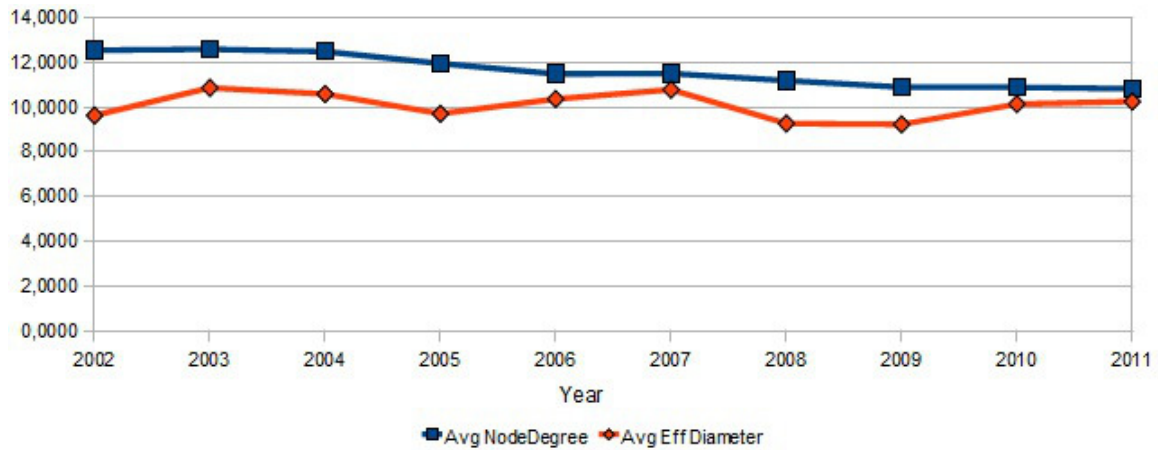


Figure 5.5: Evolution of Node Degree and Diameter of the Network

appreciate large periods of growing diameter, neither a shrinking point. Observing the monthly obtained graph in Appendix A, we detect abrupt changes of the diameter in the first years and a tendency on smoothing the variations among months. In this case, it is possible that in a long term there might be a shrinking diameter, although it is not possible to determine it with the current data.

On the other hand, the average node degree tends to decrease, although it does in a very slightly manner, but constant. If we take a look at the monthly graph of node degree in Appendix A, we could even consider that the node degree remains practically constants, as no abrupt changes are observed. For a deeper analysis, we will also study the cumulative node degree for every year. We will take as examples the plots obtained for the years 2002 (see figure 5.6), 2005 (see figure 5.7), 2008 (see figure 5.8) and 2010 (see figure 5.9). The rest of years are grouped in Appendix A). In general terms, we can observe that both graphs follow a power-law distribution where most of the nodes have few relations, and a small set of nodes monopolize the major part of connections.

In general terms, we can observe that both graphs follow a power-law distribution where most of the nodes have few relations, and a small set of nodes monopolize the major part of connections

Although the shapes of the three plots are similar, some differences can be

5. RESULTS

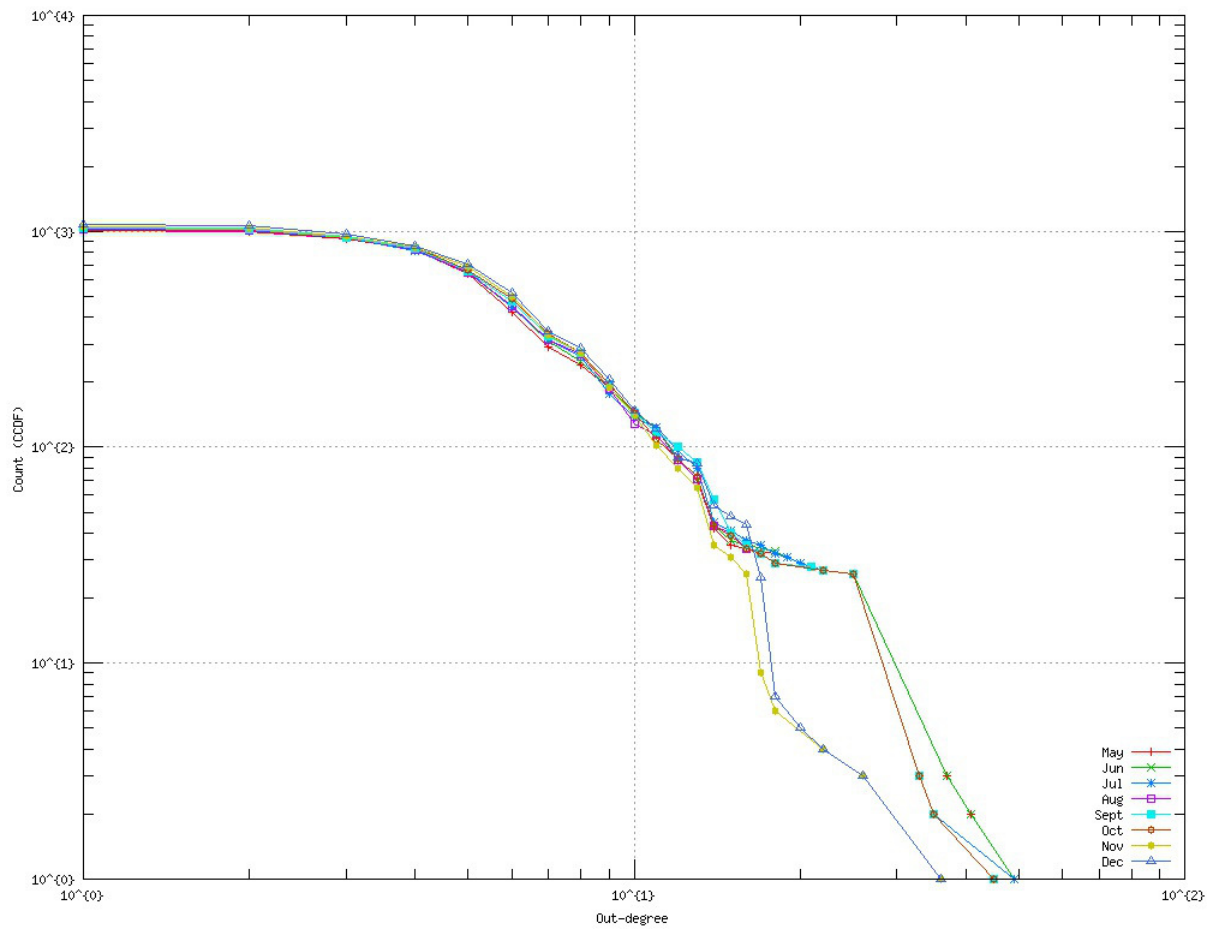


Figure 5.6: Cumulative node degree, year 2002

5.2 Data Analysis

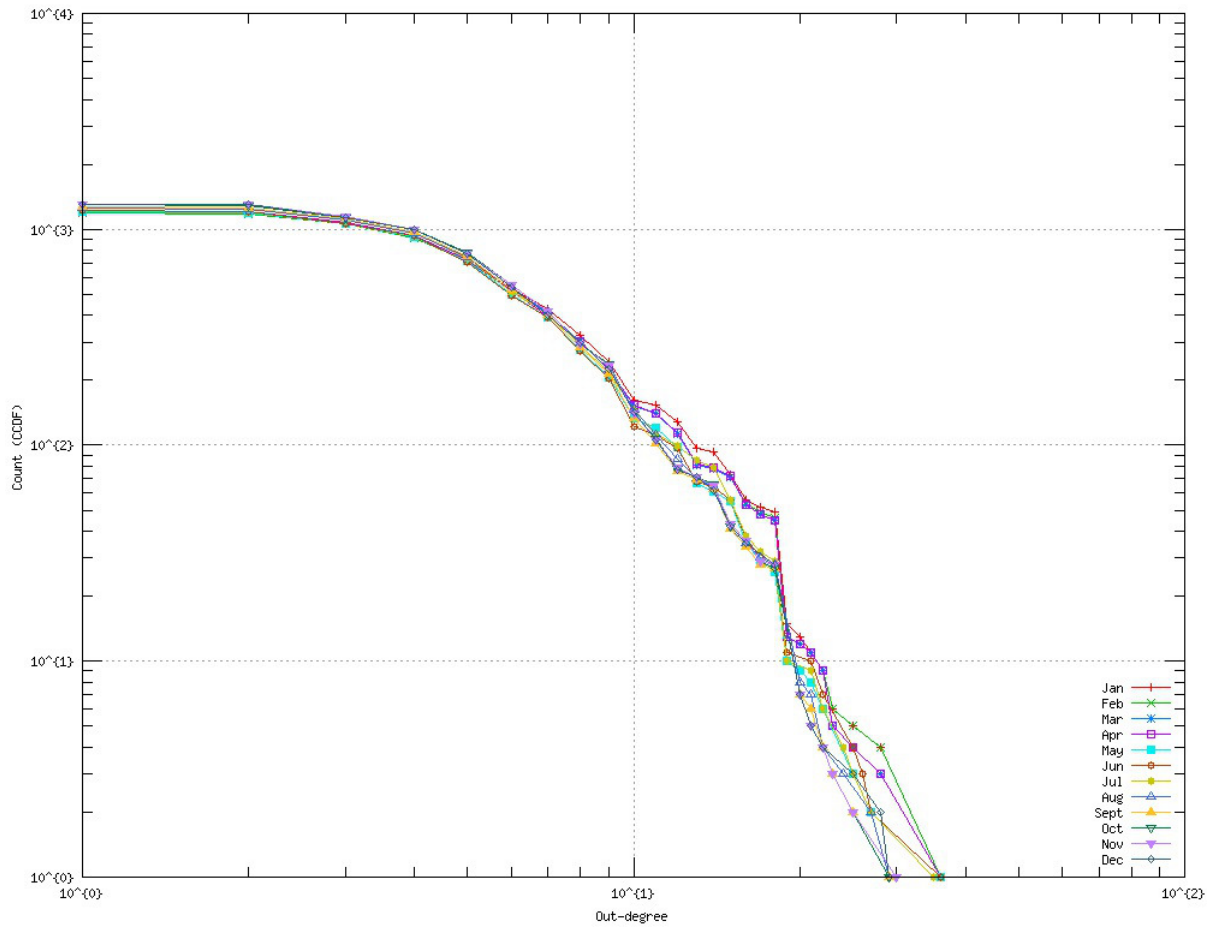


Figure 5.7: Cumulative node degree, year 2005

5. RESULTS

highlighted in head and tail of the curves. We first observe that as the years pass, the number of nodes having few connections increases and the number of nodes having many connections decreases. This behaviour is in concordance with the one described in figure 5.5. Lets us focus, however, on the evolution of the tails, that is the number of nodes having more connections (or, as we mentioned in §2.5, the hubs). In 2002, there is an important bifurcation between the months before summer and those after (see figure 5.6). In the latter the number of hubs decreases in comparison with the former ones. If we take year 2003 from Appendix A, we observe that the tail is grouped around the node degree value of the post-summer months of 2002, so we already appreciate the decreasing behaviour of figure 5.5.

Figure 5.7 represents year 2005, where the gender representation law starts. The bifurcation disappears and the ensemble of months are grouped in all moment of the curve, except for the hubs. Again we can separate between before-summer months (where the node degree is higher) and after-summer months where a slightly decrease is appreciate.

As we mentioned before, the entrance of women on boards lead to the removal of men, so it may seem normal that hubs loss of node degree is linked to the incorporation of new members on boards. Figure 5.8 represents year 2008, when the enforcement law ends and we see how the degree distribution is homogeneous among months. Although the hubs get a little increment in the number of connections, the average of node degree is still slightly decreasing. We can also say that the number of nodes with few connections increases, but not in a significant manner. Finally, we take year 2010, trying to detect some consequences of the equality in gender representation. What we observe is a stable behaviour among months, with few changes among them. This may be due to the fact the women are already established in the BODs network and lead to hubs with less connections and also less nodes *isolated*, see figure 5.9.

From a structural point of view we have analyzed the evolution of the BODs network and relationships over the time. At a first sight it may seem that the massive incorporation of women may disrupted the topology of the network established by the previous male-dominated representation. This is showed by a displacement of hubs and their number of connections, the appearance of more

5.2 Data Analysis

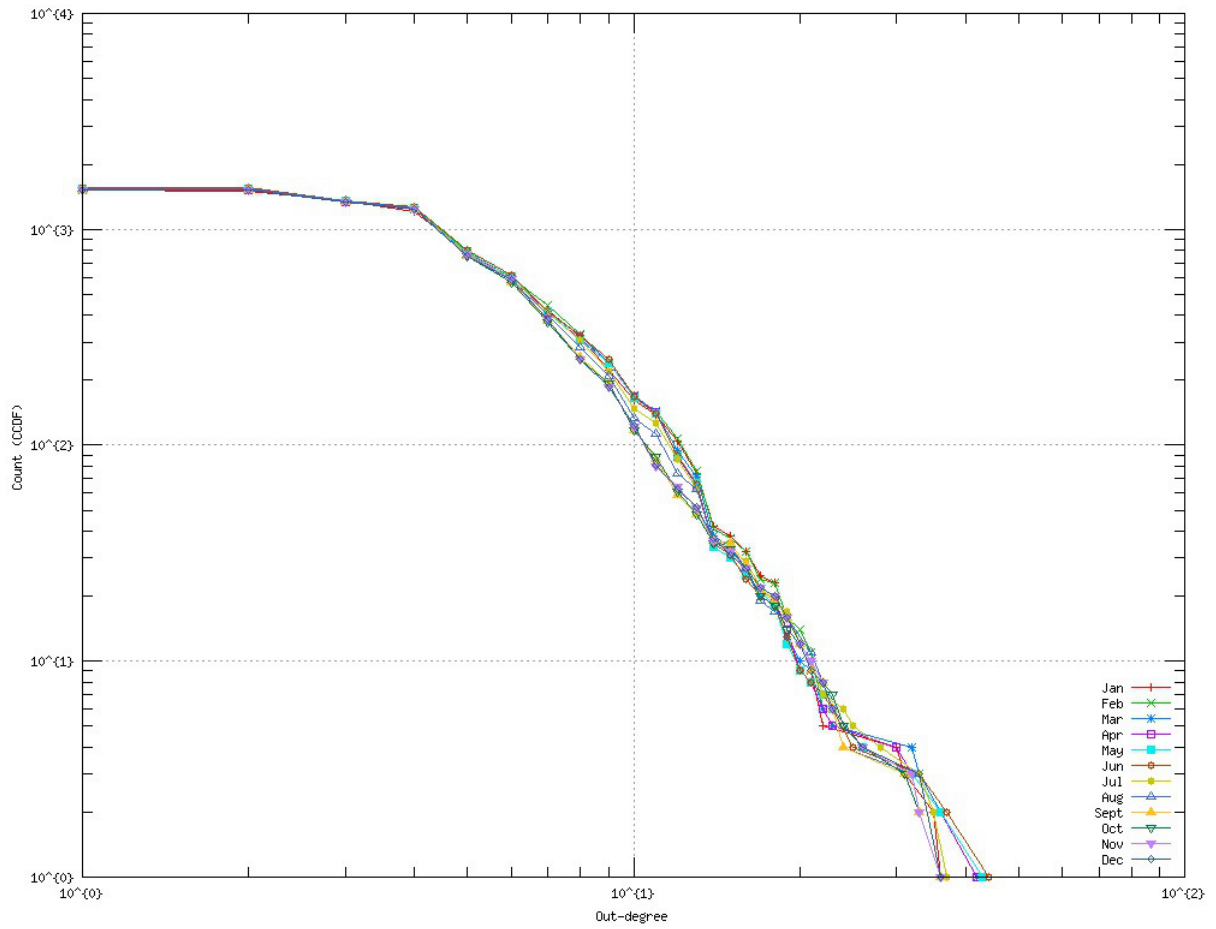


Figure 5.8: Cumulative node degree, year 2008

5. RESULTS

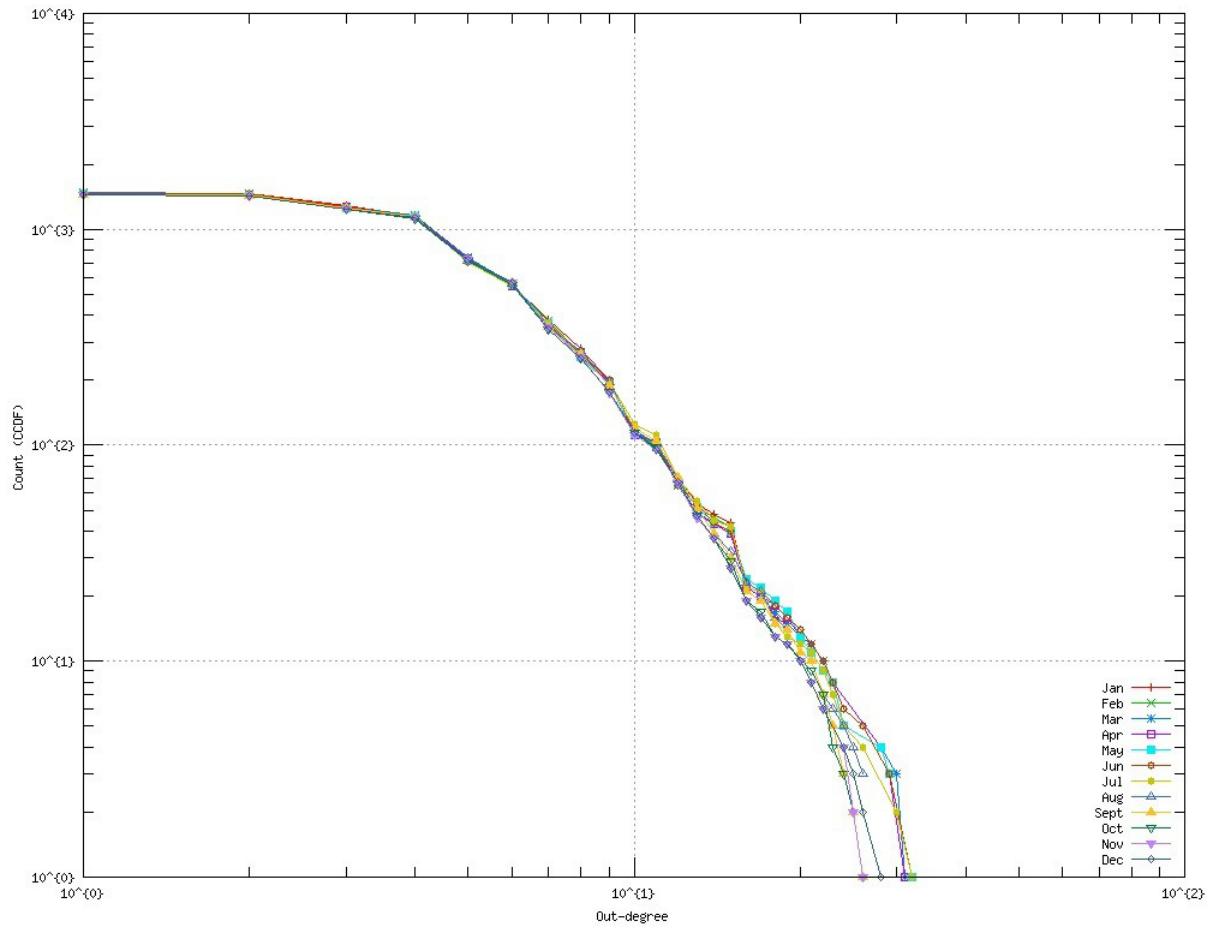


Figure 5.9: Cumulative node degree, year 2010

open triads which may be translated to a loss of trust in the relations among directors and the initial instability of the network diameter. As old board members do not know the new directors, and the incorporation of this latter is constantly growing, the paths to reach other members is continuously varying over time.

We can, however observe a tendency to shrinking the diameter, once the enforcement law ends and the network is more stable in the sense of modifications in the number of nodes and edges. In the second part of this study we are going to analyze which are the most influencer nodes over the time and observe the consequences of gender representation equality.

5.3 Analyzing influence over time

In §2 we have described some SNA metrics based on a node's centrality within the network which help to explain the influence of a node according to its position in the network. For this part of the analysis we will focus on betweenness, closeness and page rank metrics of nodes in order to detect the most influential nodes over time. For doing this, we take for every month the 25 nodes with better performance for each metric. We will study the evolution of these metrics over time, which allows to compare the similarity of lists and detect months with important changes. On the other hand we will get a global top 25 on most influential nodes for each measure. These rankings are represented in the following table 5.1, where we can see the id's of the directors, their gender (men are represented with 1, women with 2) and the number of times they appear on the rankings (number of months).

The first thing think we decided was to not take into account closeness values as a metric for measuring influence. The reason is that there are too many nodes with the same number of repetitions, so no significant information can be extracted from here. In fact, we only considered the first 24 nodes in the global ranking, as there were up to seven nodes tied in position 25. The only think we can say from this metric is that it has a vast majority of male representation. This can be explained by the years of experience of men on boards: as they have older relations with other members on boards, it is assumable to think that they will be closer to a large set of people than women that just arrive. When

5. RESULTS

Betweenness			PageRank			Closeness		
IDNode	gender	NOC	IDNode	gender	NOC	IDNode	gender	NOC
3621	2	80	4186	1	81	3358	1	21
931	1	63	1357	1	79	639	2	20
714	2	62	931	1	77	1978	1	18
2051	2	57	2898	2	71	2616	1	18
3648	2	57	3777	2	65	3741	2	18
2268	2	52	3827	1	61	3350	1	18
3777	2	51	714	2	59	2235	1	17
472	2	50	3621	2	57	4330	1	17
3137	1	47	5134	2	56	2198	2	16
3488	2	46	2051	2	55	5032	1	16
5134	2	45	3033	1	51	110	1	15
3503	2	44	2721	1	48	341	1	15
3410	2	42	3369	1	47	4503	2	15
3369	1	41	4252	1	46	16	1	14
3036	1	40	3499	2	44	142	2	14
4186	1	38	3648	2	44	974	1	14
226	2	37	3950	2	43	1027	1	14
2643	1	37	2268	2	43	1088	1	14
4880	1	36	1140	1	41	1983	1	14
1357	1	35	1452	2	40	2533	1	14
1452	2	35	226	2	39	3302	1	14
4894	1	35	5342	2	38	4103	1	14
5712	2	35	4880	1	38	5006	2	14
2898	2	34	4183	1	37	5223	1	13
5342	2	33	4894	1	35	-	-	-

Table 5.1: Ranking of the top 25 nodes for centrality.

5.3 Analyzing influence over time

considering nodes for an efficient spread of information it will be interesting to go a step further in the analysis of closeness centrality, as even if they are not the most important nodes, they may be key roles when spreading new ideas.

Let us focus now on the other two metrics. Betweenness is a measure of how short are the chains that connects a person p with the rest of the network, or in other words, how many shortest paths between two nodes pass through p . *PageRank* is the directed weighted degree centrality of a node. Both metrics are recursive: the people with highest betweenness or **PageRank** are likely to be connected to other people with high centrality values. Thus, it is not surprising to find that the top 25 of most important nodes for both metrics share 17 of the 25 directors, although they do not share positions in the global ranking. The first thing we notice is the huge representation of women in both rankings, especially for betweenness measure (up to 8 women in the top 10, 6 for the **PageRank** ranking).

We can say that, in general terms, women are located in better positions on the network, participating in many paths between pairs of directors. The gender representation law aimed for equality on gender representation, and in a quantitative way it did, but we can observe that there is no equality on the representation of important nodes. This agrees with the conclusions described in §5.2: *the network formed before the law enforcement was based on trust among strong relations of board members*. But women who joined following the implementation of the law from 2005 started to create new connections that in the end lead to gaining better positions in the network. This observation also supports the findings of Seierstad and Opsahl (41) which observed that women gained social capital during the years, surpassing the social capital of men (when in reality they expected that equality in gender representation would imply equality in social capital) (41).

It is interesting now to detect when were given the most significant changes in the top 25 rankings and related it with the law enforcement.

We compared pairs of months in order to quantify the similarity between them, or in other words, to quantify the number of modifications applied between two months (which can be addition and removal of nodes, or changes in the ranking position). Values between 0.8 and 1 are considered as quite similar, and values below 0.4 are considered important changes. At a first sight, it

5. RESULTS

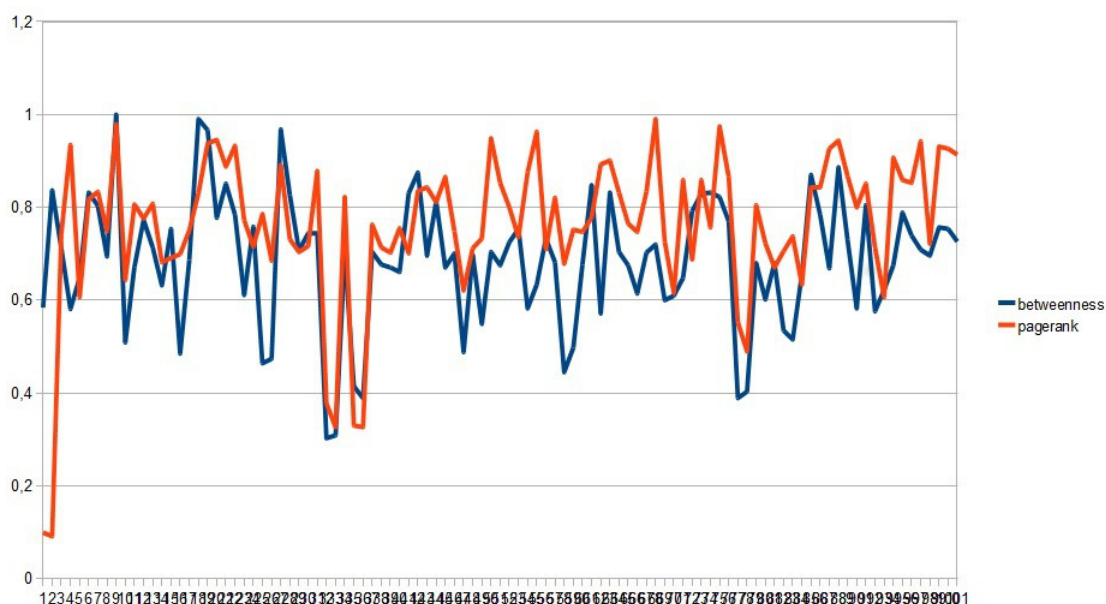


Figure 5.10: Evolution of Top 25 rankings for **Betweenness** and **PageRank**

seems that both metrics have a similar shape, although if we go deeper we can observe more oscillations in the betweenness rankings. This may be explained as a dependence of betweenness metric on the addition and removal of nodes every month. However, **PageRank** is related to the connections to important nodes, and we could say that these nodes do not experience many changes in the rankings over time. We will focus on four moments of the evolution, being those months those that present sharper changes with respect to the previous month. We will use the Gephi platform to get, not only a visual representation of the network, but also to extract which are the nodes that are moving into rankings.

The first moment analyzed is between January and February 2003. It is not surprising that the male representation dominates in both betweenness and **PageRank** measures. However, two women are present between the nodes with higher **PageRank** (one of them is node 3621, which is situated in the first position of the **PageRank** ranking by number of occurrences). We believe that the sharp change, specially for **betweenness** is due to the fact that the most influential nodes in January are not present in February and in counter-part, a woman get

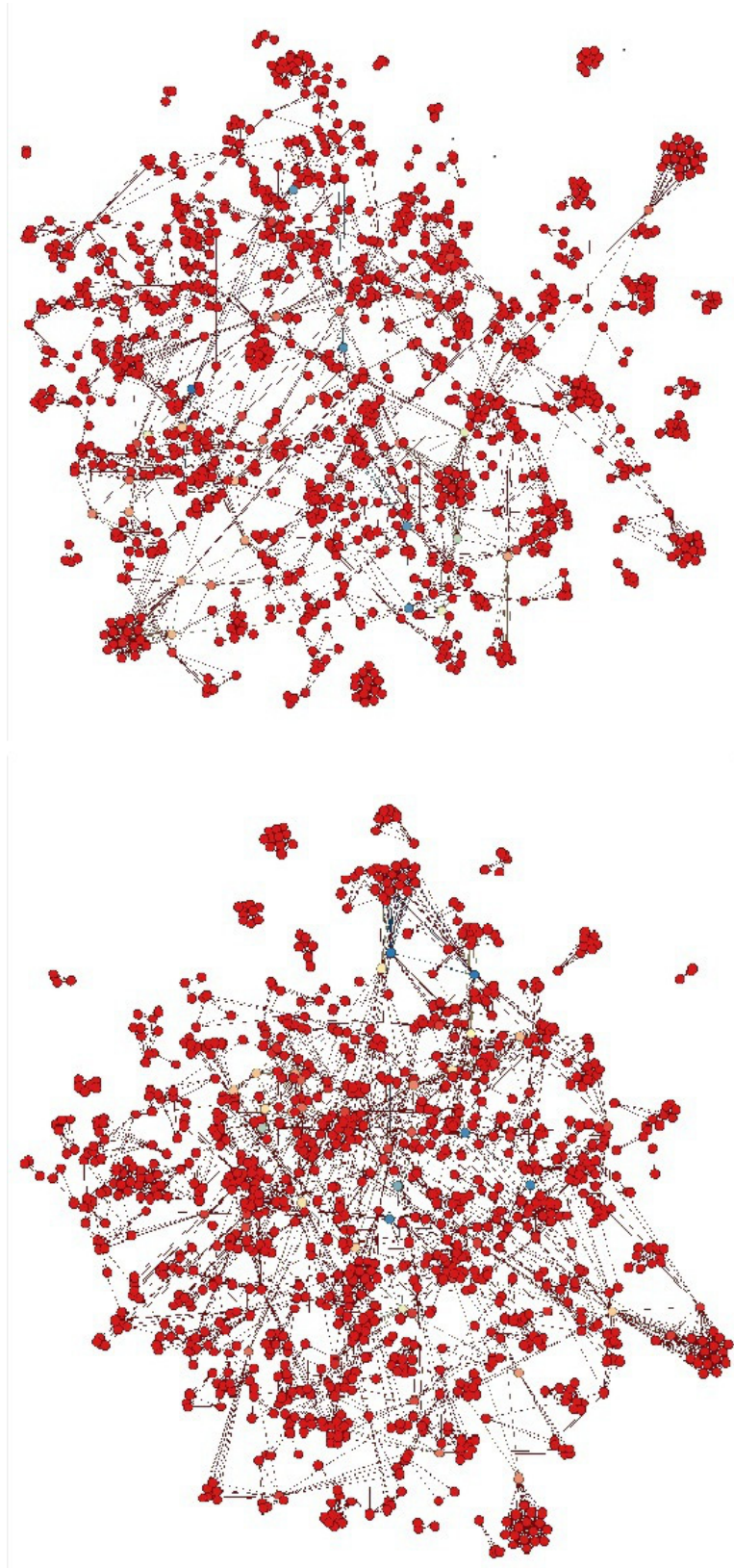


Figure 5.11: Relations among directors represented by betweenness centrality.
January-February 2003

5. RESULTS

in a top position of the ranking. By observing the graph distribution in figure 5.11, we can say that nodes in February are more sparsely distributed than in January, although some groups remain strongly intra-connected and we some broker nodes appear¹.

The same behaviour is observed when comparing the graphs for **PageRank**: the distribution is sparser, although the most important nodes are still connecting to nodes strongly connected to their clusters' nodes, see figure 5.12.

The next period analyzed takes place between March and April 2005, that is four months after the beginning of the gender representation law. Some of the nodes that were present in the previous period (and presumably before the study started) are still present in the top positions of both rankings. However, we now observe a more equal representation of genders in **betweenness** ranking. As we said before, it is normal that **PageRank** do not present changes immediately as incoming nodes will look forward to connect to important nodes.

The first thing that stands out in this image is the appearance of isolated little groups in the periphery of the network. This may agree with the observation made in §5.1 that the incoming of women lead to the decrease in the number of closed triads versus the increasing of open triads. Communities that used to be very strongly connected may have been affected by the removal of some of its members, bringing them to the peripheral zone. However, we can observe how these isolated groups start regrouping among them from one month to the other. On the other hand, this time, it is not surprising a major representation of women between the most influential nodes, as they have increased in number. We do not show the comparison between these two months according to **PageRank** values as no notable changes are given.

The following abrupt change in figure 5.10 is given between June and July 2005. Before even visualizing the graph one expects that the main change is in the increase of women representation in the top positions of the rankings, as the companies still have to reach the minimum required by law, although there should

¹The nodes are represented by their **betweenness** or **PageRank** value, going from higher values (strong blue) to lower values (strong red). Values in the middle are yellow, any other value has a degrading color (lighter blue or lighter red)

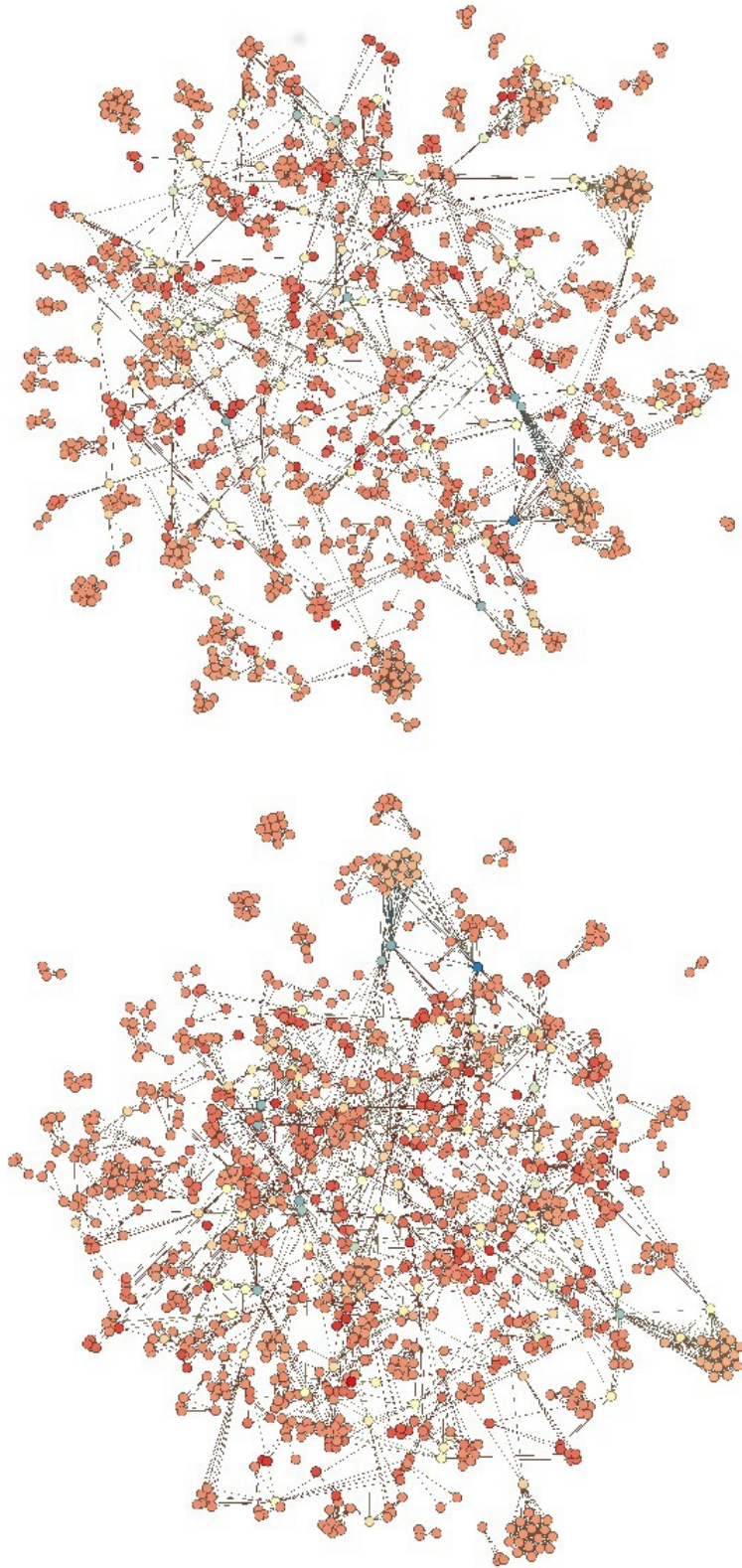


Figure 5.12: PageRank JanFeb2003

5. RESULTS

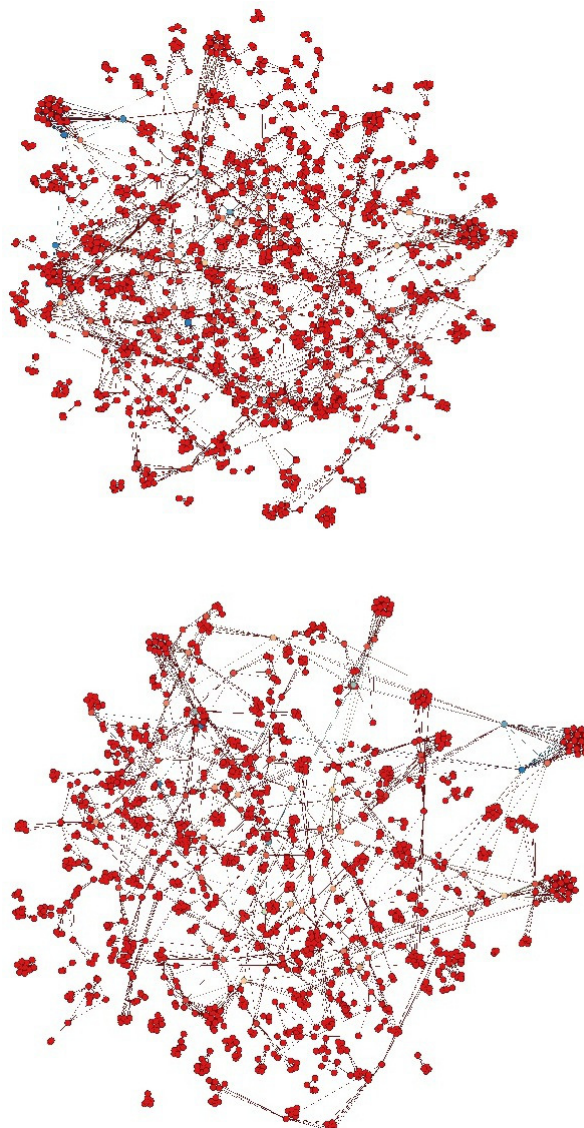


Figure 5.13: Relations among directors represented by **betweenness** centrality.
March-April 2005

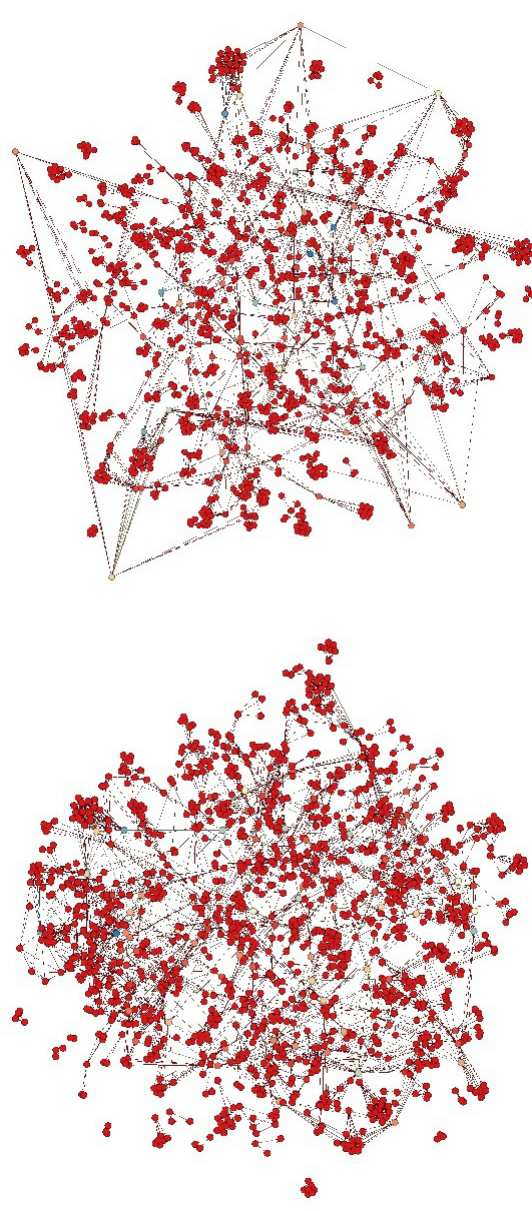


Figure 5.14: Relations among directors represented by **betweenness** centrality. June-July 2005

5. RESULTS

be an equilibrium at some point. Figure 5.14 represents the graph distribution following the values for betweenness.

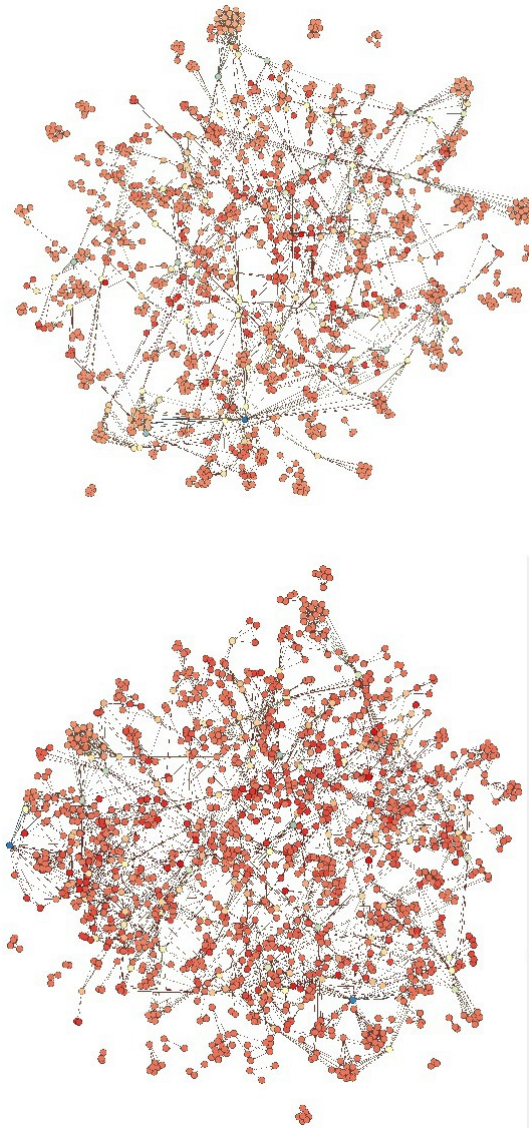


Figure 5.15: Relations among directors represented by PageRank centrality. June-July 2005

We can observe from one month to the other that there are more nodes with better values of betweenness, that is, directors being more central within the rest

5.3 Analyzing influence over time

of the network. Taking a look on which are the nodes on top positions for the betweenness ranking, we observe that while women predominated it in June, there is an gender equilibrium in July. Moreover, nodes with better rank page values in June appear at the top positions in July. This nodes are usually connected to other nodes with high betweenness and page rank (due to the recursion explained before), but also to other nodes which are less central but are better connected to closed groups. Nodes in this situation have more power, as they can play a broker role between closed groups with few inter-connections, but will also benefit them in the sense that many information will have to pass by them, so they will have the opportunity to be the first to catch new ideas, but also will have the power to either spread it or not. If we take a look at the distribution according to PageRank figure 5.15, we can observe a slightly change in the red from dark to lighter (*i.e.* there is and outperforming in PageRank values). We can assume that the responsible of this evolution are women taking better positions in the network as months pass by.

The last period analyzed corresponds to the months of April and May 2009, once the law enforcement has ended. We know from Seierstad and Opsahl that the representation of women has not rose up to the minimum required by law (40%) and that, despite this, the number of prominent women directors is higher than for men. Figure 5.16 show that people in better positions for both **betweenness** and **PageRank** are women, and still increasing in May 2009, which agrees with the previous study. We also confirm our hypothesis that these important women have lead to the disruption of the network, creating local communities of small size and several communities with strong intra-connections but completely isolated from the rest.

The red scale has also gone into dark in May 2009 and we find less important nodes (dark blues), which agrees with the observation that some directors (mainly women) have monopolized the majority of the connections. In fact, if we observe the evolution of the color scale over the years, blue scale tends to lighter and red scale tends to darker, leading to very few really important nodes, see figures 5.16 and 5.17. The opposite behaviour is given for page rank, where nodes tend over the years to go to a more intermediate color scale, because even if nodes are affected by the increase of nodes and of alternative paths to reach other nodes,

5. RESULTS

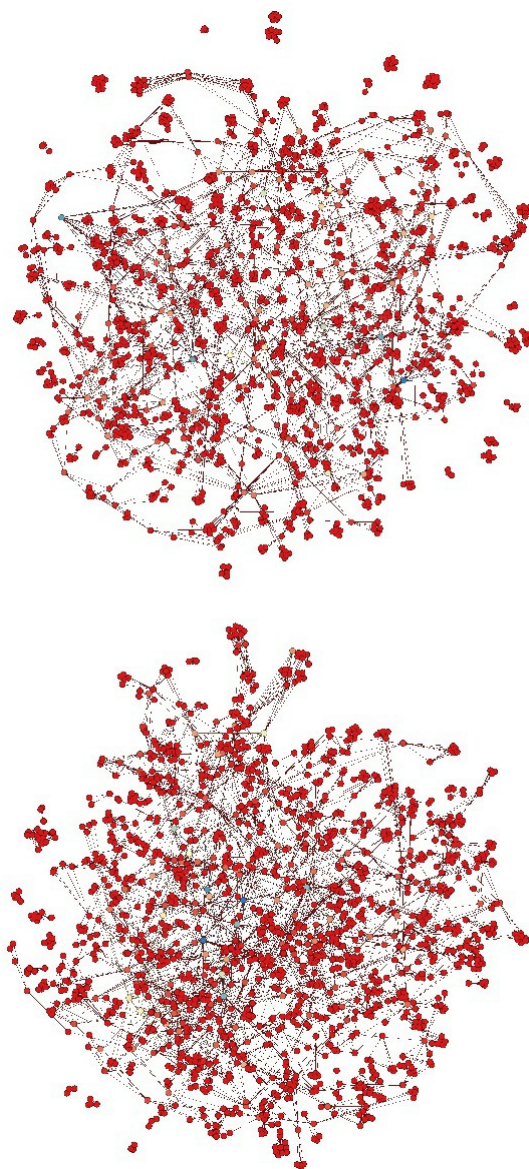


Figure 5.16: Relations among directors represented by **betweenness** centrality.
April-May 2009

5.3 Analyzing influence over time

nodes maintain their connections to important nodes, maintaining their centrality of PageRank .

5. RESULTS

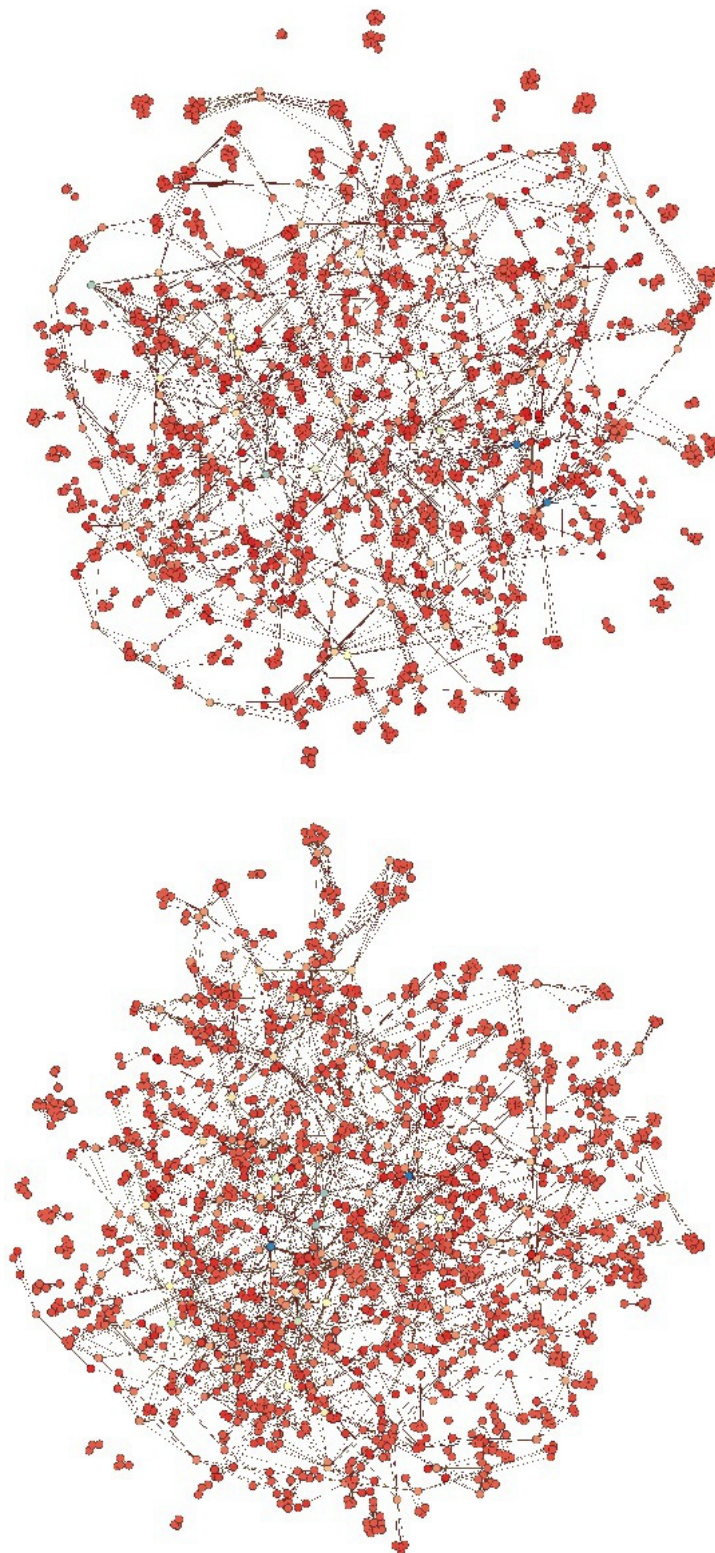


Figure 5.17: Relations among directors represented by PageRank centrality. April-May 2009

Chapter 6

Conclusions and Future work

6.1 Conclusions

From this thesis some conclusions can be deduced. Findings reveal that the idea of Norway's equality is still more of an aspiration than reality as gender inequality regimes are present in politics, academia, and Boards of Directors, but they take different forms. That was already known but we made an analysis of the same data to discover whether we can justify (challenge) them by using the SNA tools

1. Our analytical findings reveal that the idea of Norway's equality is still more of an aspiration than reality as gender inequality regime is still present boards of directors that we studied. This confirms C. Seierstad's findings based on sociological and economical facts (40). From the SNA point of view, we found that:
 - Network analysis contributes to understand the dynamics of the network by studying the addition or removal of nodes and edges, which may lead to the creation or not of cliques in the relations, and the evolution of the diameter of the network. Our findings indicate that the incorporation of women into boards as a consequence of the gender representation law has led to a rising number of open triads and the opposite behaviour on closed triads. BODs are driven by closed communities, where the reciprocal trust between members is very important as it is the basis for a negotiation. Plus, directors are motivated

6. CONCLUSIONS AND FUTURE WORK

to respect trust as their reputation can the direct consequence is a loss of reputation and may be directly related to their social capital. The constant addition of nodes and edges representing women has lead to a disrupt in the previous structure of the network.

- Instead of maintaining the already formed clusters (communities) and accepting women into them, we found that the number of communities increases according to the number of nodes. We can assume that, in average, the number of members inside communities remains more or less the same over time. We can also say that, in long terms, this has disadvantage some groups as they have been displaced to the peripheral zone of the network, an even isolated from it.
 - These two points are supported by the observation on the evolution of the cumulative clustering function. At the beginning of the study all the nodes were strongly connected among them, leading to densely connected clusters, but also presenting many inter-connections. As years pass, and with the loss of closed triads, the clustering coefficient decreases for nodes with higher node degree.
2. The objective of the gender representation law, is not only to focused on a quantitative viewpoint, but also on a qualitative. This is why our second part of the analysis uses SNA tools to study the quality of the relationships among board members as a consequence of the law enforcement. From this analysis we found that:
- The study of a ranking of important nodes taken from the point of view of the number of occurrences is not representative enough by its own. However, it has been useful to prove that, even if the equality in gender representation is quite recent, women have a predominant representation on the top positions of the rankings. This means that for larger period of times, those women have been located in advantageous position within the network, allowing them to be key roles.

- This first study has also proved a significant relation between the most important nodes according to two different centrality measures: betweenness and page rank.
- Considering changes in top rankings between months has given us the possibility to go deeper in the consequences of equality representation. First, we have confirmed that most of the nodes that are present in the global rankings presented in the previous point play a key role in the changes during months. Second, we have been able to detect in four determining periods the modifications between months, considering a modification as the addition and removal of nodes in the top positions, or nodes switching positions in the rankings. Third, we could describe an evolution on the topology of the network thanks to the visualization of the relationships among directors using Gephi. We confirm the hypothesis that women have sparsed the network, creating more hubs (thus, influential nodes), but also isolating some communities. They are also every time more present in those hubs positions, creating relations to other well positioned nodes (even if they have worse centrality, they permit to access to more nodes by shortest paths), but leading to a displacement in centrality terms for most of the nodes in the network

In brief, results show that, even if there is an equality in the gender representation, consequences on the social network do not show gender equality. We have a more sparse network, with small communities and a loss of trust from a structural point of view; we have more prominent women, and moreover, the number of prominent directors has been doubled between 2002 and 2009. This trend has been defined in Norway as the *Golden Skirts* phenomenon. Although the gender representation may be seen as non successful from the equality point of view, it may give the opportunity to establish a new women role model.

6.2 Future work

In order go a step deeper in the analysis of the dynamics of this network, it would be interesting to expand the study to the area of the evolution of communities

6. CONCLUSIONS AND FUTURE WORK

from a gender point of view. This would include a monthly report about the aggregations or removal of nodes, and a proportion of women and men representation in those communities. The idea is to have a report of which are the most stable communities (that is, the ones that accept less incorporations, or none at all) and the most evolving communities, so the ones that have experienced more changes, putting especial attention in those communities that have aggregated women over time. We can study the stability of communities from a gender view point over time.

It would be very interesting to relate these findings with the concepts of Burt's structural holes and Stark-Vedres' structural folds in order to analyze the network in terms of innovation. We have said that the gender representation law had as a consequence the loss of closure in the network. According to Burt, closure dead-ends *creativity* and *brokerage* tends to creativity and innovation by importing ideas from other communities. We have seen that women tend to be part of many boards simultaneously and are key roles in connecting pairs of nodes of the network. However, the curse between closure and brokerage described in §2.3.5 makes the analysis hard to qualify, as it is easy to have suddenly too many brokers, and thus redundant information (which may be the case in Norway with this tendency of prominent directors).

On the other hand, we can see these prominent directors as nodes creating overlapping communities, or structural holes. Instead of focusing on advantageous positions for spreading innovations, Stark and Vedres focus on *how* these innovations have been first created. They also take as a case study a temporal analysis on boards of directors and place the sources of new ideas in the intersection of cohesive groups (*intercohesion*). However, if we want to perform this kind of work, we need a more complete dataset, which includes social background, companies benefits and especially the content of the information that flows over the network.

Finally, it would be interesting to go deeper in the analysis of the directors' affiliations as we could detect, not only influential directors, but also influential companies within the network.

Appendix A

Appendix A

This appendix contains some figures that have not been included in the document of this thesis but have been considered for the analysis performed. They may be considered as complementary information, but no analysis of them is given in here.

The first set of images corresponds to the cumulative degree function (CDF) for the years 2003, 2005, 2007, 2009, 2011, which have not been explained in section 5.1. They are shown in chronological order.

A. APPENDIX A

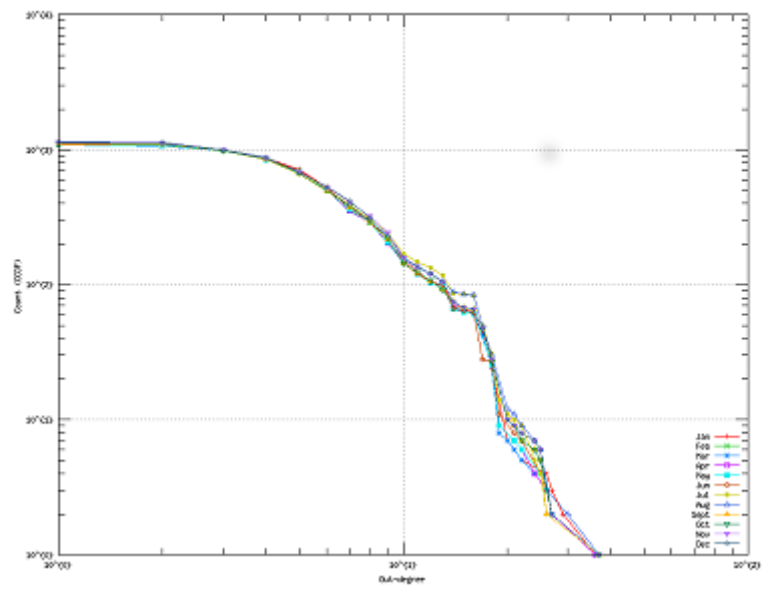


Figure A.1: CDF 2003

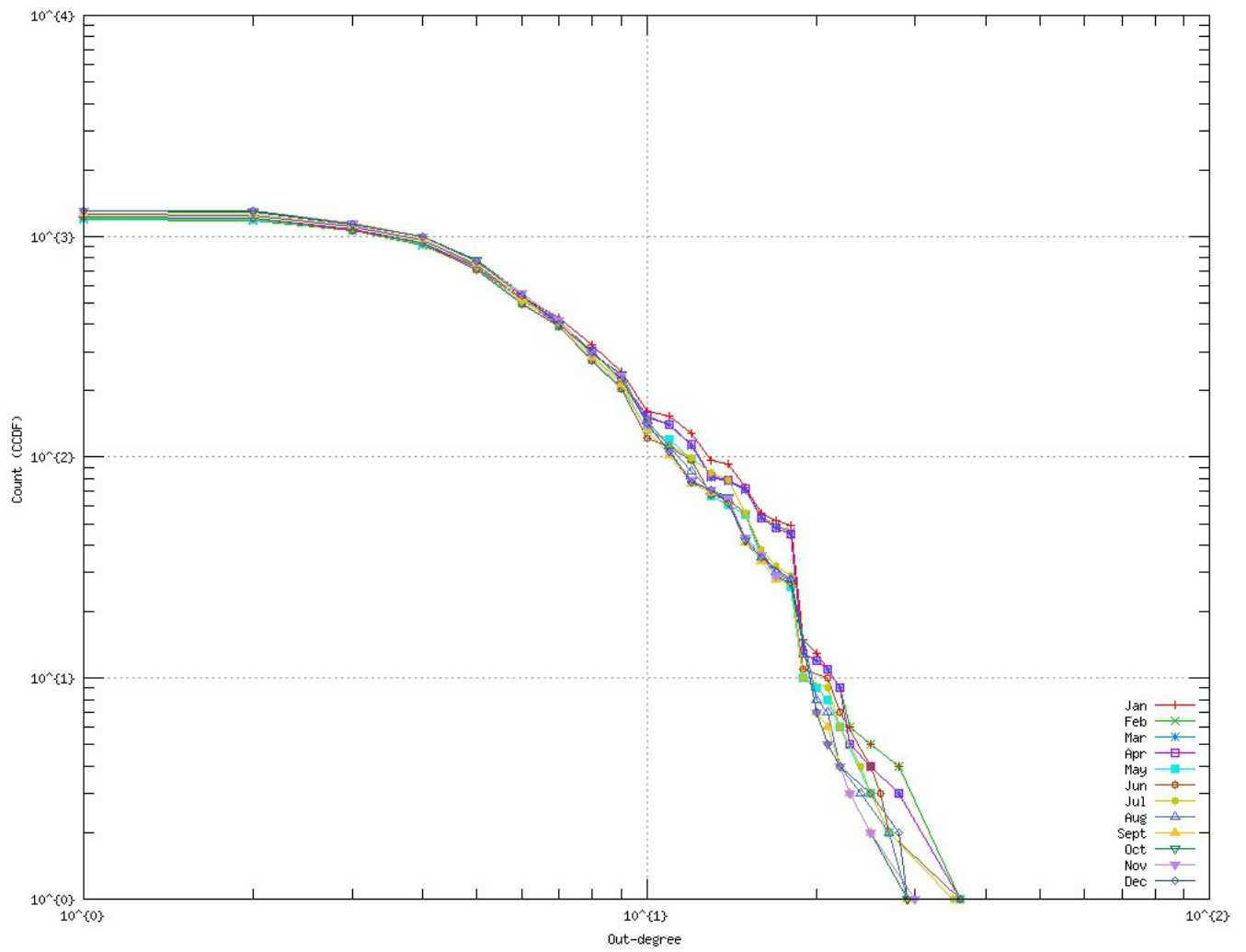


Figure A.2: CDF 2005

A. APPENDIX A

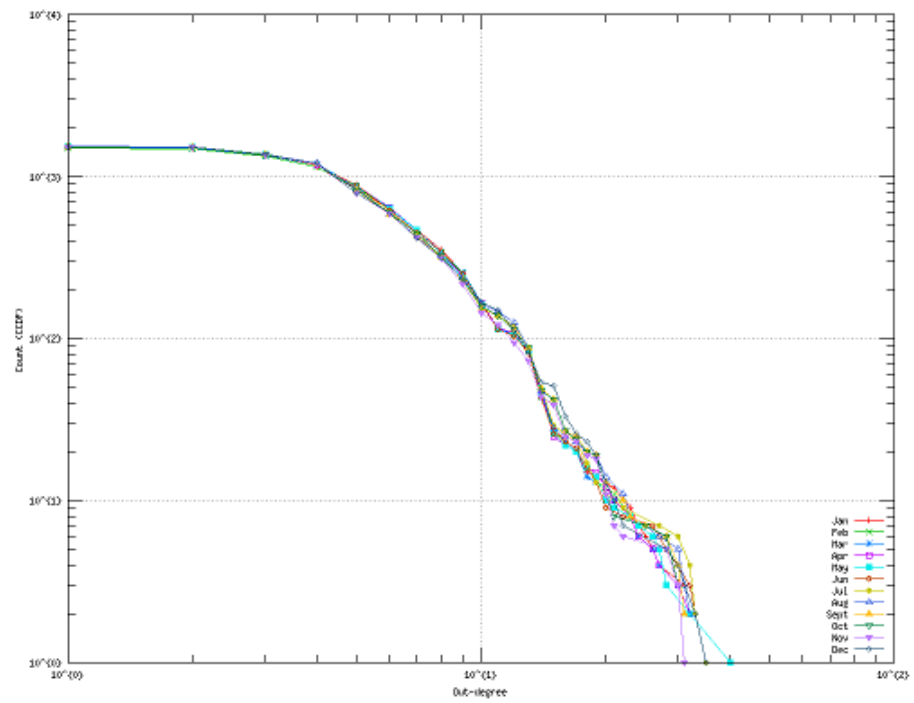


Figure A.3: CDF 2007

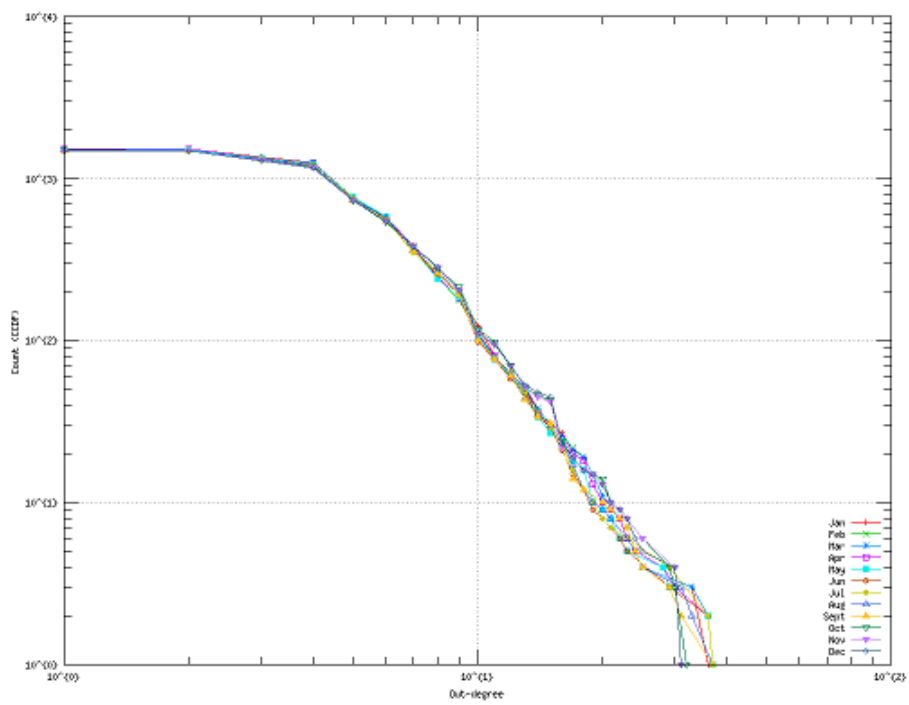


Figure A.4: CDF 2009

A. APPENDIX A

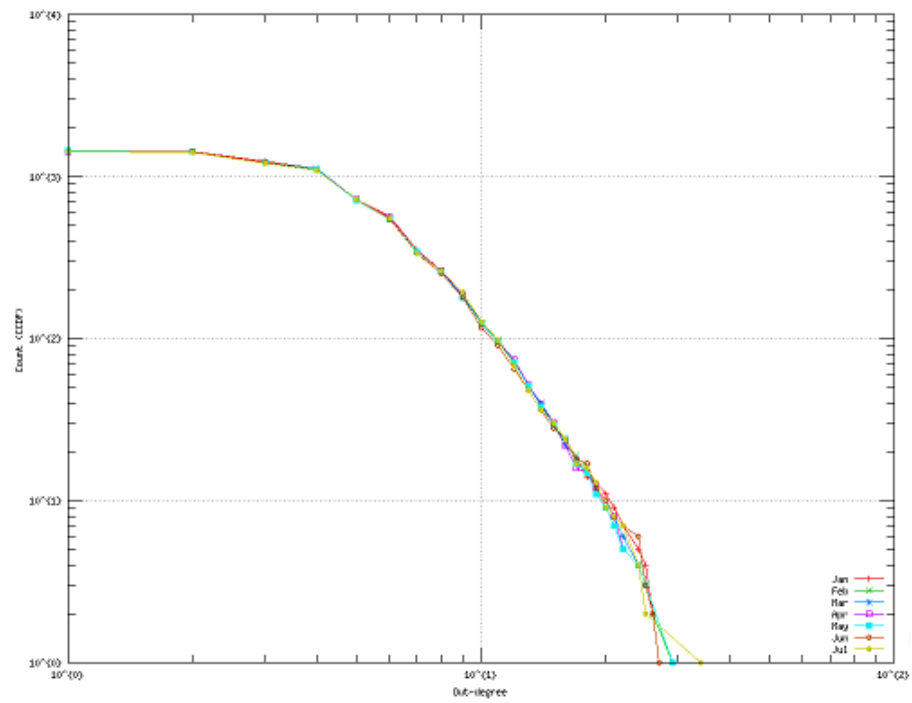


Figure A.5: ICDF 2011

Appendix B

Appendix B

The following figures correspond to monthly reports on some SNA metrics. In §5.2, annual averages are used to represent the data, as the dataset is too large for a good visualization:

B. APPENDIX B

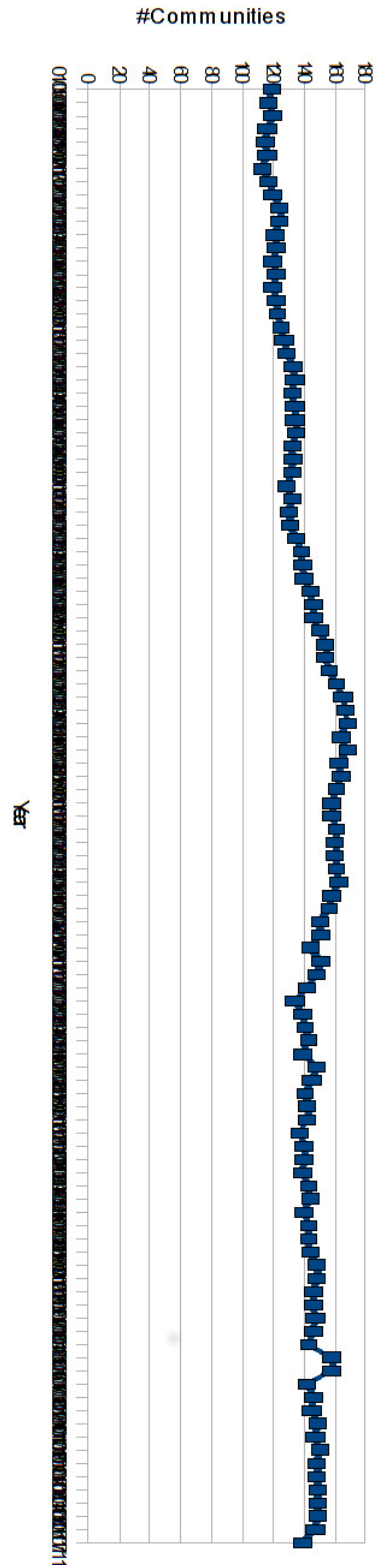


Figure B.1: Number of communities by month

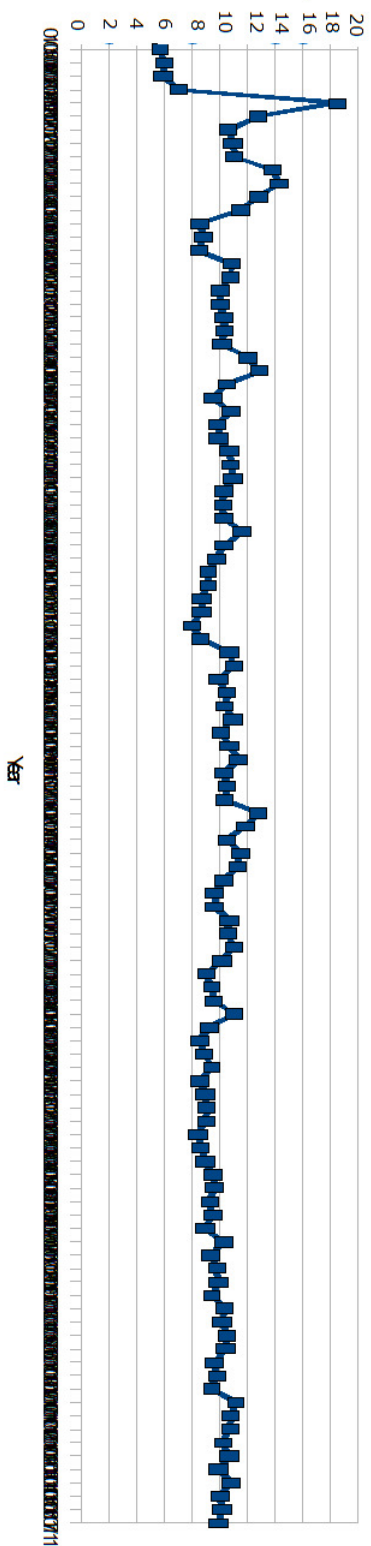


Figure B.2: Effective diameter by months

B. APPENDIX B

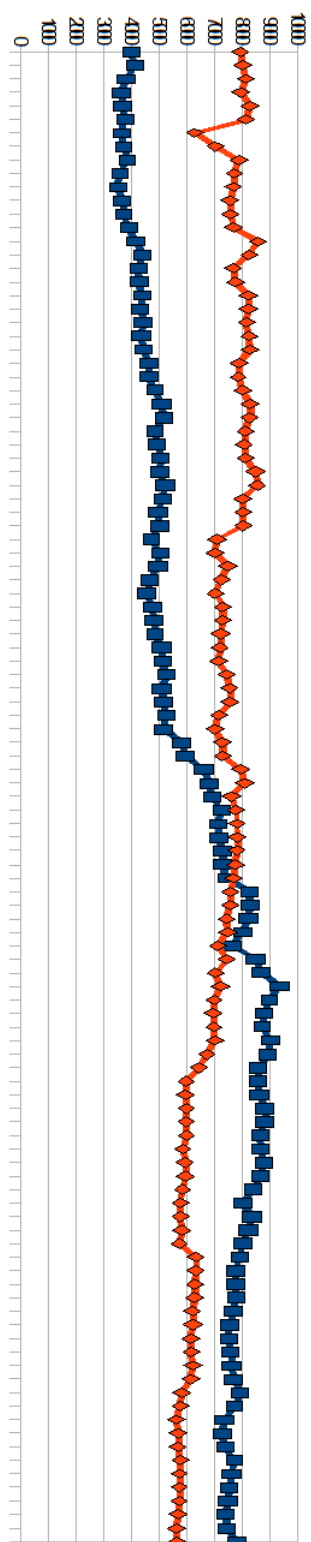


Figure B.3: Number of open and closed triads by month.

References

- [1] A. L. BARABÁSI, H. JEONG, Z. NEDA, E. RAVASZ, A. SCHUBERT, AND T. VICSEK. Evolution of the social network of scientific collaborations. *Physica A*, **311**(3–4):590–614, 2002. [23](#)
- [2] MATHIEU BASTIAN, SEBASTIEN HEYMANN, AND MATHIEU JACOMY. Gephi: An open source software for exploring and manipulating networks. In EYTAN ADAR, MATTHEW HURST, TIM FININ, NATALIE S. GLANCE, NICOLAS NICOLOV, AND BELLE L. TSENG, editors, *ICWSM*. The AAAI Press, 2009. [38](#)
- [3] P. BONACICH. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, **2**:113–120, 1972. [11](#)
- [4] STEPHEN P. BORGATTI. Identifying sets of key players in a social network. *Computational & Mathematical Organization Theory*, **12**(1):21–34, 2006. [20](#)
- [5] RONALD BURT. *Structural Holes*. Harvard Press, Cambridge, April 1992. [12](#)
- [6] RONALD BURT. *Brokerage and Closure: An Introduction to Social Capital*. Oxford University Press, Oxford, April 2005. [29](#)
- [7] J. COLEMAN. Social capital in the creation of human capital. *American Journal of Sociology*, **94**:95–120, 1988. [29](#)

REFERENCES

- [8] B. D. BAYSINGER AND H. BUTLER. Corporate governance and the Board of Directors: Performance effects of changes in board composition. *Journal of Law, Economics, and Organizations*, **1**:101–124, 1985. [30](#)
- [9] J. DUCH AND A. ARENAS. Community detection in complex networks using Extremal Optimization. *Physical Review Letters*, **72**(027104), 2005. [27](#)
- [10] VICTOR M. EGUÍLUZ, DANTE R. CHIALVO, GUILLERMO A. CECCHI, MARWAN BALIKI, AND A. VANIA APKARIAN. Scale-free brain functional networks. *Physical Review Letters*, **94**(018102):1–4, January 2005. [23](#)
- [11] SANA ELOUAER. A social network analysis of interlocking directorates in french firms. <https://www.opesc.org/>, September 2006. [33](#)
- [12] K. FAUST. Comparing social networks: Size, density, and local structure. *Metodoloski zvezki*, **2**:185–216, 2006. [16](#)
- [13] L. C. FREEMAN. Some antecedent of social network analysis. *Social Networks*, **1**:215–239, 1979. [11](#), [13](#)
- [14] L. C. FREEMAN. Centrality in social networks: Conceptual clarification. *Connections*, **19**:36–42, 1996. [25](#)
- [15] L. C. FREEMAN. Visualizing social networks. *Connections*, **1**, 2000. [27](#)
- [16] The open graph viz platform. <https://gephi.org/>, September 2012. [38](#)
- [17] RUMI GHOSH AND KRISTINA LERMAN. Community detection using a measure of global influence. *CoRR*, **abs/0805.4606**, 2008. [18](#)
- [18] K.M. GILLEY, J.E. COOMBS, AND R.H. FORD. *Board of director influence and organizational performance in entrepreneurial firms: Mediating and moderating effects*. Babson College, 2000. [30](#)
- [19] MICHELLE GIRVAN AND M. E. J. NEWMAN. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, **99**:8271–8276, 2002. [18](#), [27](#), [36](#)

REFERENCES

- [20] ROGER GUIMERA, MARTA SALES-PARDO, AND LUIS A. N. AMARA. Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E* **70**, 025101, **70**, 2004. [21](#)
- [21] ROBERT A. HANNEMAN AND MARK RIDDLE. *Introduction to social network methods*. University of California, Riverside, April 2005. [9](#)
- [22] ROSANNE JANE HAWARDEN AND STEPHEN MARSLAND. Locating women board members in gendered director networks. *Gender in Management: An International Journal*, **26**(8), 2011. [31](#)
- [23] P.W. HOLLAND AND S.LEINHARDT. Local structure in social networks. *Sociological Methodology*, **7**:1–45, 1976. [16](#)
- [24] M.G. KENDALL. *Rank Correlation Methods*. Hafner Publishing Co, New York, April 1955. [38](#)
- [25] LESKOVEC, KLEINBERG, AND FALOUTSOS. Graph evolution: Densification and shrinking diameters. In *ACM Transactions on Knowledge Discovery from Data*, **1**. 2007. [44](#)
- [26] JURE LESKOVEC, KEVIN J. LANG, ANIRBAN DASGUPTA, AND MICHAEL W. MAHONEY. Statistical properties of community structure in large social and information networks. In J. HUAI, R. CHEN, HSIAO-WUEN HON, Y. LIU, WEI-YING MA, A. TOMKINS, AND X. ZHANG, editors, *Proceedings of the 17th International Conference on World Wide Web, WWW 2008, Beijing, China, April 21-25, 2008*, pages 695–704. ACM, 2008. [42](#)
- [27] B.P. LEUNG AND J. SILBERLING. Using sociograms to identify social status in the classroom. *The California School Psychologist*, **11**:57–61, 2006. [25](#)
- [28] FREDERICK LILJEROS, CHRISTOFER R. EDLING, LUIS A. NUNES AMARAL, H. EUGENE STANLEY, AND YVONNE ABERGT. The web of human sexual contacts. *Nature*, **41**:907–908, 2001. [23](#)
- [29] C. D. MANNING, P. RAGHAVAN, AND H. SCHUTZE. *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, April 2008. [10](#), [16](#), [18](#)

REFERENCES

- [30] NINA MISHRA, ROBERT SCHREIBER, ISABELLE STANTON, AND ROBERT ENDRE TARJAN. Clustering social networks. In ANTHONY BONATO AND FAN R. K. CHUNG, editors, *WAW*, **4863** of *Lecture Notes in Computer Science*, pages 56–67. Springer, 2007. [29](#)
- [31] J.L. MORENO. *Application of the Group Method to Classification*. National Committee on Prisons and Prison Labor, New York, April 1932. [25](#)
- [32] J.L. MORENO. *Who Shall Survive?* Nervous and Mental Disease Publishing Company, Washington, DC, April 1934. [26](#)
- [33] J.L. MORENO. Autonomous groups: A new field for adult education. *Journal of Educational Sociology*, **19**(9), 1946. [26](#)
- [34] M.E.J. NEWMAN. The structure of scientific collaboration networks. In *Proceedings of the National Academy of Sciences*, **98**, pages 404–409. National Academy of Sciences, 2001. [23](#)
- [35] CHRISTINA PRELL AND JOHN SKVORETZ. Looking at social capital through triad structures. *Connections*, **28**(2), 2008. [41](#)
- [36] JOSEP MA. PUJOL, JAVIER BÉJAR, AND JORDI DELGADO. Clustering algorithm for determining community structure in large networks. *Phys. Rev. E* **74**, 016107, **74**, 2007. [27](#), [36](#)
- [37] ERIK QUALMAN. *Socialnomics: How Social Media Transforms the Way We Live and Do Business*. April 2010. [4](#)
- [38] L.S. RASHOTTE. *Social Influence*. Blackwell Publishing, 2006. [10](#)
- [39] A. REKA, H. JEONG, AND A.-L. BARABÁSI. The diameter of the world wide web. *Nature*, **401**:130–131, 1999. [22](#)
- [40] C. SEIERSTAD. *Exploring the Norwegian paradox of vertical sex segregation: Strategies and experiences in politics, academia and company boards*. PhD thesis, School of Business and Management at Queen Mary University of London, 2011. [5](#), [30](#), [35](#), [36](#), [65](#)

REFERENCES

- [41] C. SEIERSTAD AND T. OPSAHL. For the few not the many? the effects of affirmative action on presence, prominence, and social capital of female directors in norway. *Scandinavian Journal of Management*, **1**:1150–1190, 2011. [5](#), [30](#), [33](#), [34](#), [36](#), [38](#), [39](#), [53](#)
- [42] Social Network Analysis Platform. <http://snap.stanford.edu/>, September 2012. [37](#)
- [43] DEZSO SOLTAN AND ALBERT-L. BARABÁSI. Halting viruses in scale-free networks. *Phys. Rev. E* **65**, 025101, **65**, 2002. [23](#)
- [44] WILLIAM STEVENSON AND ROBERT F. RADIN. Social capital and social influence on the board of directors. *Journal of Management Studies*, **46**(1):16–44, 2009. [30](#)
- [45] M. TEIGEN. Gender quotas on corporate boards on the diffusion of a distinct national policy reform. *Comparative Social Research*, **29**, 2012. [5](#), [30](#)
- [46] S.C. VANCE. *Boards of Directors: Structure and Performance*. University of Oregon Pres. [30](#)
- [47] BALAZS VEDRES AND DAVID STARK. Structural folds: Generative disruption in overlapping groups. *American Journal of Sociology*, **115**(4):1150–1190, 2010. [19](#), [30](#)
- [48] D.J. WATTS AND S. H.STROGATZ. Collective dynamics of ‘small-world’ networks. *Nature*, **393**:440–442, 1998. [21](#), [44](#)