UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA E CONTABILIDADE
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA
MESTRADO ACADÊMICO

LUAN FALCÃO DANIEL SANTOS

NETWORK EFFECTS, CONFORMISM AND MISBEHAVIOR IN BRAZILIAN CLASSROOMS

FORTALEZA
2016

LUAN FALCÃO DANIEL SANTOS

NETWORK EFFECTS, CONFORMISM AND MISBEHAVIOR IN BRAZILIAN CLASSROOMS

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Ceará como requisito parcial para a obtenção do Título de Mestre em Economia.

Orientador: Prof. Dr. José Raimundo de Araújo Carvalho Júnior

FORTALEZA

2016

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S236n Santos, Luan Falcão Daniel.
Network Effects, Conformism and Misbehavior in Brazilian Classrooms / Luan Falcão Daniel Santos. – 2016.
79 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Faculdade de Economia, Administração, Atuária e Contabilidade, Programa de Pós-Graduação em Economia, Fortaleza, 2016.
Orientação: Prof. Dr. José Raimundo de Araújo Carvalho Júnior.

1. Peer Effects. 2. Conformism. 3. Networks. 4. Local-average Model. I. Título.

CDD 330

LUAN FALCÃO DANIEL SANTOS

NETWORK EFFECTS, CONFORMISM AND MISBEHAVIOR IN BRAZILIAN CLASSROOMS

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Ceará como requisito parcial para a obtenção do Título de Mestre em Economia.

Aprovada em

BANCA EXAMINADORA

Prof. Dr. José Raimundo de Araújo Carvalho Júnior(Orientador)
Universidade Federal do Ceará

Prof. Dr. Diego de Maria André
Universidade Federal do Rio Grande do Norte

Prof. Dr. Victor Hugo de Oliveira Silva
Instituto de Pesquisa e Estratégia Econômica do Ceará

Para minha esposa,
Flávia.

Resumo

Para entender de que forma as networks afetam o comportamento dos indivíduos, em específico o comportamento de estudantes no último ano do ensino médio dentro da sala de aula, estimamos um modelo de Network effects, o Local-average Model para duas variáveis comportamentais: fazer uma prova ou teste sem ter se preparado e colar em uma prova, a fim de entender como o comportamento dos amigos de um indivíduo afeta o comportamento do mesmo. Encontrou-se uma efeito positivo e estatisticamente significante para o network effect da probabilidade de se fazer uma prova ou teste sem ter se preparado, e um efeito positivo, mas não significante para o network effect da probabilidade de se colar em uma prova. Este resultado mostra que políticas ou ações que visam a redução da probabilidade de um estudante fazer um exame sem se preparar tem o que se chama social multiplier effect, pois além dessa política mudar o comportamento do estudante em relação a essa variável, sua mudança de comportamento afeta positivamente o comportamento das pessoas em sua network.

Palavras-chaves: Peer Effects; Conformism; Networks; Local-average Model.

Abstract

For understanding how networks affect the behavior of individuals, specifically the behavior os students in the last year of high school inside the classroom, we estimate a model of Network effects, the Local-average model for two behavioral variables: doing an exam without being prepared and cheating in an exam, in order to understand how the behavior of individual's friends affects his or her behavior. It was found a positive and statistically significant effect for the network effect of the probability of doing an exam without being prepared, and a positive, but not significant effect for the network effect of the probability of cheating in an exam. This result shows that policies or actions aiming the reduction of the probability of a student do an exam without being prepared have what is called social multiplier effect, because besides this policy change the behavior of the student regarding this variable, his or her change of behavior affects positively the behavior of people in his or her network.

Key-words: Peer Effects; Conformism; Networks; Local-average Model.

List of Figures

List of Tables

Contents

1 Introduction

It is very common to think in macroeconomic, socioeconomic and demographic variables as shaping an individual and influencing his or her microeconomic outcomes, like criminal activity, educational and labor market outcomes. However, despite these variables play an important role, we can not ignore the fact that the behavior of our social contacts influences our own behavior. In many situations and in a variety of environments, not just social, but also financial, business and scientific ones, just to mention a few, our peers are influencing our behavior and this can change some outcomes that may arise from the interactions with them. Think about criminal efforts, smoking behavior and other social diseases, the adoption of a new technology by firms, the purchase of a new product by consumers, human capital decisions, R&D efforts, public goods provision, information gathering, consumption externalities, herd behavior, location decisions by companies, bankruptcy and foreclosure decisions, etc. All these are examples of variables that may be influenced by the economic agents peers' own behavior regarding each of these variables.

A large literature has developed around the study of peer effects on individuals outcomes, as we show in the next chapter. More recently, with the development of databases containing information about the connections among the individuals in the data, econometricians could use this information for better understanding the network effects. Patacchini e Zenou (2012) developed a model in which conformism plays an important role in shaping the behavior of individuals, the local-average model. As we show in more details in chapter 5, an individual loses utility if he fails to conform with his or her peers. This way, he or she will try to get closer to the average behavior of his or her social contacts. We use their model to estimate the peer effects in the school environment on two different outcomes: the probability of doing an exam without being prepared and the probability of cheating in an exam.

The database used in this dissertation was the PAEST – Pesquisa de Aspirações e Expectativas dos Estudantes Concludentes do Ensino Médio – or Survey on Aspirations and Expectations of High School Students, the first Brazilian database to include this network structure. The PAEST collected detailed information about the context students face in their last year of high school in the northwestern state of Ceará, in Brazil. It has information about more then 2000 students from 47 public and privite schools in three geographical regions of Ceará, namely the state's capital, Fortaleza, its metropolitan area and the state's countryside.

In the chapter 2 we make a review of the literature of peer effects on a variety of outcomes, such as crime, education and labor market. As we show, the literature of peer effects can be divided in two distinct groups regarding the methodology, namely the literature of neighborhood effects and the literature of network effects. In the first, mainly because data limitation, authors suppose that all the individuals surrounding an agent in some environment, such as a school or a neighborhood, are able to influence this agent's behavior. This assumption brings an identification problem to the estimation of peer effects in econometric models of neighborhood effects, known as reflection problem, as we discuss in more details in chapter 5. The literature of network effects arises with the recent development of databases including the social and economic connections existing among the individuals in the data, making possible the use of the tools of Social and Economic Network Theory for the analysis of peer effects. When the networks present some easily identifiable features, the identification problem is overcame, and econometricians can find the real size of the peer effects.

In chapter 3 we introduce some important tools of the Social Network Theory. We begin with the characterization of a network, which is composed by a set of agents, the set of all individuals in that network, and by a graph, which is represented by a special matrix called adjacency matrix. The adjacency matrix summaries the social or economic connections that exist in the network. Basically, each row and column in the adjacency matrix indexes an individual in the network in such a way that, when two individuals are socially connected, the entrance of the adjacency matrix representing the connection between these two individuals is equal to a positive value, and zero if they are not socially connected. After the characterization of a network, we introduce some important definitions in Network Theory, such as paths, walks, n-neighborhoods and degree distribution. Such definitions are specially important for the econometric modeling of network effects and for the conditions of identification of the peer effects, as you will see in chapter 5. We then turn attention to the called centrality measures, which measure how central an individual is in a network in terms of direct and indirect connections. We give special attention to the eigenvector related measures, that have an important relation with the econometric models of network effects, including the model used in this dissertation.

In chapter 4 we talk about the database used in this research work, the PAEST, showing the demographic and geographic distribution of the data among the three regions of Ceará and between private and public schools. We also talk about networks in the database and introduce the graphical representation of the networks for each one of the classes used in this work, as well as an associated measure called degree distribution.

In chapter 5 we present Patacchini e Zenou (2012)'s local-average model. We also discuss in more detail the reflection problem and why the use of networks overcome this identification problem. The advantage of econometric models of network effects over models of neighborhood effects is shown by Bramoullé, Djebbari e Fortin (2009), which show that the reflection problem barely never arises when we use this approach for estimating peer effects. We then present the estimation of the local-average model for our dependent variables, the probability of doing an exam without being prepared and the probability of cheating in an exam.

Finally, in chapter 6, we make some final considerations about this work and propose strategies to enhance this study and improve our understanding of peer effects in education.

2 Neighborhood and Network Effects

The literature of peer effects over individuals outcomes has followed two distinct lines concerning the agent's reference group, the social group where each individual belongs, which can be thought as his or her classmates, coworkers, acquaintances, residential neighbors, firms in a production chain, etc. Just to mention a few. Although both attempt to capture the effect of the *social space* over the agent's behavior and, consequently, over the outcomes that may arrive from social or economic interactions, they are quite different in methodology. This way, the literature of peer effects can be divided into the literature of neighborhood effects and network effects, and why we make this distinction will become clear shortly.

The literature of neighborhood effects considers the reference group of an individual simply as being symmetric over all individuals within the same social space. This way, the reference group of an individual in this work would be all the people in is his or her classroom. The reference group is thought as the set of "residential neighbors". In such a way, for peer effects in crime, the reference group of a criminal may be all the people in the neighborhood he lives; for peer effects in firms production, it may be other firms in the same industrial district, regardless of being competitors or not, regardless of being connected by a production chain or not; for peer effects in education, the reference group is usually the set of classmates or schoolmates. The list goes on, but the idea is that an economic agent is influenced by the composition of his neighbors, and shocks in the composition of his or her neighborhood may change some outcome that arises from social or economic interactions.

The other strand of the literature of peer effects, the literature of network effects, takes into account the social connections that may exist among the individuals in the social space. It is reasonable that an individual in a social space, such as a classroom, do not interact with everyone in that space. Instead, he or she has a social relation with a subset of individuals in that classroom. In this setup, the reference group of each individual is different within the social space, unless they form a social structure called clique. This literature uses the tools os the social network theory, which we introduce in the chapter 3.

In this chapter, we are going to make a review on the literature of peer effects, focusing on its division between neighborhood and network effects and talk about the weakness and strengthness of each approach. In the next section, we make a review of the literature of neighborhood effects. Then we review the literature of network effects.

2.1 Neighborhood Effects

The biggest limitation for the development os models that take into account the network structure existing in any social context was certainly the lack of datasets including the social connections among the individuals in the data. Just recently some datasets with information about who is socially connected to who among the individuals inside the survey were available, such as the Add Health - The National Longitudinal Study of Adolescent to Adult Health - in United States, and the PAEST - Pesquisa de Aspirações e Expectativas dos Estudantes Concludentes do Ensino Médio - in Brazil. It was in that context that the literature of neighborhood effects has developed.

Although social interactions play an important role in how individuals shape their behavior, the neighborhood effects may arise from local shocks or through institutions in an individual's residential neighborhood, as pointed out by Topa e Zenou (2015). For instance, the bankruptcy of a local industry, the presence of churches or NGOs in a given location, etc. For education, more qualified teachers or policies aiming to improve students outcomes may influence the educational achievements of all individuals in that social space. This way, we should also distinguish the neighborhood effects as arising from social interactions and from those shocks and institutions.

Another limitation of this approach to estimate the peer effects is that econometricians face an identification problem, specifically known in the literature of peer effects as the *reflection problem*, which we discuss in more details in chapter 5. As shown by Manski (1993), when individuals interact socially, they affect each other mutually, making difficult to identify the extension of the influence of an agent's social contacts behavior over his own behavior.

A common problem for the analysis of peer effects we can face in both approaches is that individuals may sort into networks because the individuals in that network or neighborhood already have similar tastes or behave similarly. If it is not possible controlling for these similar tastes or behaviors, we must be sure that the network is formed randomly. This is possible analysing the *degree distribution*, which we discuss in more details in section 3.3. Basically, if we observe that the relative frequency of the number of connections in a network follows a Poisson distribution, then we can assume that the individuals in that network are sorted randomly. It is clear that we can not use this artifice when we do not have information about the network structure, making more difficult to justify the random sort of agents in a neighborhood.

Finally, as discussed before, the individuals in the same residential neighborhood may be exposed to local shocks or by local institutions that may be unobserved by the econometrician.

Before Manski (1993), most works focused on the effects of growing up in disadvantaged neighborhoods over educational achievement, employment, and other indicators of social welfare. For example, Corcoran et al. (1989) investigate the effects of family and neighborhood background on men's economic status. They have found the existence of substantial disadvantages in economic status for black men, men from lower-income families, and men from more welfare-dependent families or communities. In the same line, Brooks-Gunn et al. (1993) investigate the effects of neighborhood characteristics on the development of children and adolescents. They have found effects of the presence of affluent neighbors on Childhood IQ, teenage births, and school-leaving, even after the differences in the socioeconomic characteristics of families are adjusted for. These works and others before Manski (1993) usually use family and neighborhood attributes and, sometimes, mean outcomes in the social space analyzed. This bring a problem of identification to their model, as we show in more details in chapter 5, which invalidates their findings. Basically, including mean outcomes of the individuals in a neighborhood makes impossible to separate the peer effects from exogenous effects.

As exposed by Topa e Zenou (2015), subsequent works on neighborhood effects have followed two strategies to overcome the problem of model identification. The first was the use of experimental and quasi-experimental approaches. Most of these works focus on immigrant refugees sorted into locations determined by local authorities and how the characteristics of this new environment affected them. and also the relocation of families from public housing projects in poor neighborhoods to low-poverty neighborhoods.

Popkin, Rosenbaum e Meaden (1993) study the effect of relocating black low-income families who were either former or current residents in public housing and moved to subsidized housing in Chicago and its suburban areas, a governmental program called Gautreaux program. The selection of participants was not random, but the sort of agents into city neighborhoods and suburban neighborhoods was based on the availability of units, therefore quasi-random. They have found that the participants of the program who moved to suburban areas were significantly more likely to find a job after the moving than the participants who moved to city areas, even among those who had never had a job before moving.

Jacob (2004) studies the variation in housing assistance generated by public housing demolitions in Chicago during the 1990's to examine the impact of high-rise public housing on student outcomes. Families affect by the demolitions were offered vouchers to move anywhere in the metropolitan area of Chicago. He reports finding no effect of the moving on children affected by the demolitions on a wide variety of achievement measures.

He argues that the majority of households that left high-rise public housing in response to the demolitions moved to neighborhoods and schools that closely resemble those they left. Contrary to the Gautreaux experiment, Jacob (2004) finds no evidence of any impact of the demolitions and subsequent relocations on student outcomes.

Oreopoulos (2003) studies the effect on long-run labor market outcomes of adults who were assigned, when young, to substantially different public housing projects in Toronto. Same as Jacob (2004), Oreopoulos (2003) finds no evidence of neighborhood effects in a variety of outcomes, including unemployment, mean earnings, income, and welfare participation. However family differences, as measured by sibling outcome correlations, account for up to 30% of the total variance in the income and wages.

Studies based on the Moving to Opportunity program (see Ludwig, Duncan e Hirschfield (2001), Kling, Ludwig e Katz (2005) and Kling, Liebman e Katz (2007)), or simply MTO program, usually find no neighborhood effects on economic outcomes. As explained by Topa e Zenou (2015), this was a large, randomized experiment in which participants volunteered for the study, and they were randomly assigned to one of three groups: a control group, a group receiving a housing voucher without any restrictions, and a third group receiving a voucher to move to a low-poverty neighborhood. The last two groups moved to neighborhoods with significantly lower poverty rates, less crime, and in which residents reported feeling safer. Unlike the evidence of no neighborhood effects on economic outcomes, the studies found a large and significant neighborhood effect on a variety of adult mental health measures. Ludwig et al. (2012) studied the effect of the MTO program 10 to 15 years later the participants have received the housing vouchers. Again, neighborhood effects are no significant for economic outcomes, nor for physical health. They were reported to be marginally significant for mental health and highly significant for subjective well-being.

Another set of studies has focused on the resettlement of refugees into new locations selected by the host country authorities, such as Edin, Fredriksson e Aslund (2003), Aslund et al. (2011) and Beaman (2012), among others. These studies have found a significant neighborhood effect on outcomes such as employment status, earnings and educational performance.

A second approach to the estimation of neighborhood effects is based on structural models of social interactions. As explained by Topa e Zenou (2015), these models generate stationary distributions with well-defined properties over space, like excess variance across locations, or positive spatial correlations. The parameters of these models can then be estimated by matching moments from the simulated spatial distribution generated by the model with their empirical counterparts from spatial data on neighborhoods or cities.

For neighborhood effects on crime, Glaeser, Sacerdote e Scheinkman (1996) try to explain the variance of crime rates in United States cities across time, where the propensity of engagement in a crime is influenced by the individual's neighbors. They create a model where social interactions create enough covariance across individuals to explain the high cross-city variance of crime rates. The model provides an index of social interactions which suggests that the amount of social interactions is highest in petty crimes, moderate in more serious crimes, and almost negligible in murder and rape.

Topa (2001) analyzes a structural model of transitions into and out of unemployment that explicitly incorporates local interactions and allows agents to exchange information about job openings to estimate the impact of local social interaction effects on employment outcomes. He found a significantly positive amount of social interactions across neighbouring tracts. The local spillovers are stronger for areas with less educated workers and higher fractions of minorities.

In the next section, we make a review of the literature of network effects. The models discussed next use the tools of social and economic network theory, which we discuss in more details in chapter 2.

2.2 Network Effects

An agent may be influenced in two opposite ways by his or her peers. Most commonly, individuals tend to follow his or her peers actions. We can observe this in smoking behavior, criminal efforts, R&D efforts, etc. Just to mention a few. On the other hand, economic agents may want to distance themselves from the behavior of their peers. Examples include local good provision and information gathering. As the peers influence is a two-way road, we call those situations of *games of strategic complements* and *games of strategic substitutes*, respectively. Most of social and economic situations are based on complementarity, therefore, we will only discuss here the literature of network effects on games of strategic complements. We discuss in more details games of strategic complements in chapter 5, specifically the *local-average model*.

There are two different models of games of strategic complements, the *local-aggregate model* (see Calvó-Armengol e Zenou (2004) and Ballester, Calvó-Armengol e Zenou (2006, 2010).) and the *local-average model* (see Patacchini e Zenou (2012).). In the first one, individuals are assumed to be influenced by the sum of social or economic contacts he or she has. In the local-average model, we suppose that an individual is influenced by the average behavior of his or her peers, here thought as the people here or she is connected, making a distinction in relation to the residential neighborhood in the literature of neighborhood effects.

Bramoullé, Djebbari e Fortin (2009) show that the identification problem is overcame when we take into account the network structure existing when modeling peer effects. Basically, because the reference group among the individuals in a network rarely overlaps, we can easily differentiate the peer effects from the exogenous effects. In the literature of neighborhood effects, we consider that the reference group is symmetric among the individuals in a social space, mainly due to limitations in datasets, and Manski (1993) shows that this setup brings the reflection problem, making impossible to separately identify the peer effects and the exogenous effects.

Patacchini e Zenou (2012) develop the local-average model, which we discuss in more details in section 5.2. According to their model, when interacting socially or economically, individuals gain utility when they conform to the average behavior of their social contacts. They study the criminal behavior of adolescents on a variety of crimes using the Add Health dataset, which contains information about the social connections among the individuals in the data, allowing the use of social network theory. They find that conformity plays an important role for all crimes, especially for petty crimes. This suggests that, for juvenile crime, an effective policy should be measured not only by the possible crime reduction it implies but also by the group interactions it engenders.

3 A Short Introduction to Network Analysis

In this chapter, we introduce the analytical tools on how to represent and measure networks. An important feature of networks is about the positions individuals occupy in a network and how they influence and are influenced by their neighbors and indirect neighbors, or "the friends of my friends". Some measures of clustering and centrality are introduced later in this chapter, and provide a great perspective of how important they can be for the microeconomic analysis of individuals outcomes. Now, we begin with some definitions on how to represent networks and formalize the notion of players, connections and indirect connections in a network.

3.1 How to Represent a Network

The most important components of a network are the players and the relations that may exist between them. Therefore, it is really important to represent these players and their possible relations in a tractable way, that can easily represent the network structure and be used in a broad class of applications. As it follows, we present basic concepts and definitions that help on the analysis of networks and that will eventually be used in this work. For further explanations on how to represent and measure networks, see Jackson (2010).

As we said, the main components of a network are the players and their possible relations. The players, or nodes, as will become clear later, might be any economic agent, like firms, governments, politicians, organizations, web pages, etc. Not just people. In a particular network, the nodes that are involved in this net of relationships will be represented by a set, as follows:

Definition 3.1. Let $1, \cdots, n$ index the players that are involved in the network under analysis. The *set of nodes*, N, is given by $N = \{1, ..., n\}$.

The relationship between any two nodes may arise into two different ways. The first kind is an undirected relationship, in which the two nodes are either in a relationship or not. However, there are cases where one player relates to the other but this second does not relate to the first. In this case, where there is not a mutual connection between the players, we say that this kind of relation is a directed relationship. If at least one node in the set N does not have a mutual connection with some other node, we say that the network of the n players is a directed network. If it is not the case, we have an undirected network.

Undirected networks are a common form of representing networks that involve friendships and partnerships, the central issue of this dissertation. The distinction between directed and undirected networks are not just conceptual. The ways researchers can model directed and undirected networks are quite different, so, depending on how connections between the players are formed, this distinction could be extremely important. Therefore, all the following definitions and network measures are assumed an undirected network set up. For the same review on directed networks, see Jackson (2010).

The relations among the n nodes in the set of nodes, N, can be summarized by the following matrix:

Definition 3.2. Let g a real-valued $n \times n$ matrix. The relation between the node i and the node j is given by g_{ij}, where $g_{ij} = 1$ if node i relates to node j, or $g_{ij} = 0$ otherwise.

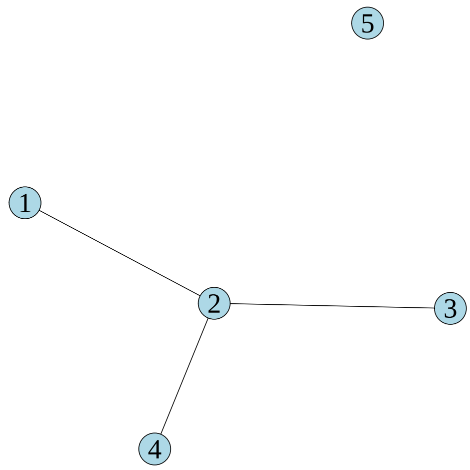
In the case of undirected networks, it is always the case in which $g_{ij} = g_{ji}$ for all i and j in N, such that g is as symmetric matrix, because the need of a mutual relationship between the players. g is often referred to as the adjacency matrix. The values of g_{ii}, the self-links or loops, may be 0 or 1, depending on the the situation under analysis.

Definition 3.3. A network, or graph, (N, g), consists of the set of nodes, $N = \{1, ..., n\}$, and the $n \times n$ real-valued adjacency matrix, g.

An alternative way of describing the links between the nodes in a network (N, g) is to define g as a set of the existing links, $g = \{ij : g_{ij} = 1\}$. If the nodes i and j are linked to each other, we say that $ij \in g$. If they are not, we say that $ij \notin g$. Saying that $ij \in g$ is equivalent to say that $g_{ij} = 1$. Seeing g as the adjacency matrix defined as Definition 3.2 or as a set of the existing links is just a matter of convenience, and they can alternate in this text, depending on the situation.

For instance, suppose a network involving 5 nodes, $N = \{1, 2, 3, 4, 5\}$, and adjacency matrix given by

Figure 3.1: Network on 5 nodes



Prepared by the author.

$$g = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

This network is depicted in figure 3.1, where the circles, or nodes, represent the players, and the straight lines, or links, represent an existing relationship between the two nodes connected by the link. The absence of a link between any two nodes indicates that these nodes do not relate to each other.

3.2 Some Important Definitions

In a network, individuals do not benefit only due to individuals they are immediately close to. Indeed, individuals in a network benefit from the whole chain of contacts that may exist between all the players. This is easy to see when you imagine the spread of information, or even the spread of a disease. I may not know a colleague of my father who is infected by a flu, but if he infect my father, it is possible that I can also be contaminated by this flu. These indirect interactions between individuals in a network are as much important as the direct contacts, so, it is important here to formalize this feature. I begin with some definitions about connectedness in a network, that are essential for network analysis.

Definition 3.4. A *path* in a network (N, g) between nodes i and j is a sequence of links $i_1i_2, ..., i_{K-1}i_K$ such that $i_ki_{k+1} \in g$ for each $k \in \{1, ..., K-1\}$, with $i_1 = i$ and $i_K = j$, and such that the sequence $i_1, ..., i_K$ is distinct.

Definition 3.5. A *walk* in a network (N, g) between nodes i and j is a sequence of links $i_1i_2, ..., i_{K-1}i_K$ such that $i_ki_{k+1} \in g$ for each $k \in \{1, ..., K-1\}$, with $i_1 = i$ and $i_K = j$.

The distinction between a path and a walk in a network is that a path involves only "handshakes" that have never occurred before, while in a walk, when trying to get in the terminal node, repeated interactions may occur.

Definition 3.6. A *geodesic* between nodes i and j is the shortest path between these nodes.

Sometimes, there are multiple ways to get in a person that is not direct connected to a node. The definition of geodesic is of a huge importance, because it is more likely that the indirect interactions between any two nodes occur via the shortest path between them.

It is easy to see that, if we set the self-links $g_{ii} = 0$, the entrance g_{ij}^k of the matrix g^k, the k^{th} power of g, tells us how many walks of size k there exist between the nodes i and j in the network (N, g). This is actually one of the possible ways of solving the famous mathematical problem of the seven bridges of Königsberg.

It is also important to consider the set of nodes in the network an individual is directly connected to, or the set of nodes one can reach through a given number of "handshakes". It arises, then, the concept of neighborhood and extended neighborhoods.

Definition 3.7. The *neighborhood* of a node i, $N_i(g)$, is the set of nodes that i is linked to. $N_i(g) = \{j : g_{ij} = 1\}$.

Given a set of nodes $S \subset N$, the neighborhood of S is the union of the neighborhoods of the nodes in S:

$$N_S(g) = \bigcup_{i \in S} N_i(g) = \{j : \exists i \in S, g_{ij} = 1\}$$

The "friends and friends of my friends" set, or the extended neighborhood of size 2, which includes the nodes i is directly connected to and all the nodes that can be reached by paths of size 2, is given by:

$$N_i^2(g) = N_i(g) \cup \left(\bigcup_{j \in N_i(g)} N_j(g) \right)$$

And, the k-neighborhood of i, which consists of all nodes that can be reached from i by paths of size less or equal than k, is given by:

$$N_i^k(g) = N_i(g) \cup \left(\bigcup_{j \in N_i(g)} N_j^{k-1}(g) \right)$$

Definition 3.8. The *degree* of a node i is the number of nodes that are directly connected to it. $d_i(g) = \#\{j : g_{ij} = 1\} = \#N_i(g)$.

Finally, we define network density.

Definition 3.9. The *density*, $d(g)$, of a network (N, g), describes the portion of the potential connections in a network that are actual connections, and may be calculated by:

$$d(g) = \frac{\sum_{i=1}^{n} d_i(g)/n}{n-1}$$

So far, we have shown several tools on how to represent a network. On the next three sections, we begin the discussion on how to measure a network, beginning with the description of the degree distribution of networks and then introducing an overall measure of clustering, that enables us to compare networks regarding its cohesiveness, the clustering measure We finish with some microeconomic measures, that explicit the role of the centrality of a node on networks, the centrality measures.

3.3 Degree Distribution

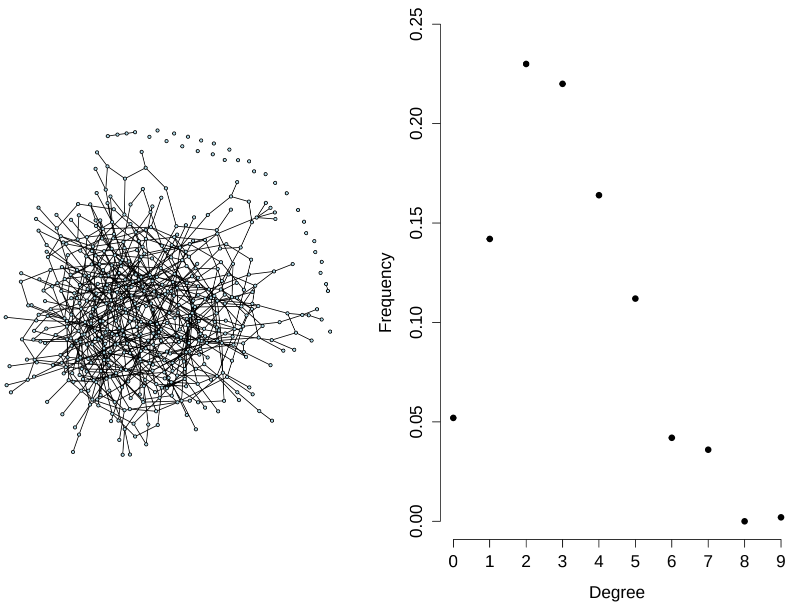
An useful way of representing a network is through its degree distribution, which summary the relative frequencies of nodes's degree. Large networks with hundreds or thousands of individuals are better represented on its degree distribution, as you can see on the network on 500 nodes depicted on figure 3.2.

Beyond this advantage regarding representation, the degree distribution also provides us with some insights about the network formation and growing process. The degree distribution on the right side of figure 3.2 is a typical example of the degree distribution of a network formed under a Poisson distribution, which it is due to Erdos e Renyi (1959) and we shortly introduce next.

3.3.1 Poisson Distribution

Consider a network (N, g) on n nodes with the probability of two nodes form a link given by p, where $0 \leq p \leq 1$. The links are formed according to a Bernoulli distribution,

Figure 3.2: Node-link representation versus degree distribution representation of a network on 500 nodes.



Prepared by the author.

such that a whole network is formed with probability given by a binomial distribution. Any network on n nodes with x links, where $0 \leq x \leq \binom{n-1}{2}$, has a probability of forming given by:

$$p^x(1-p)^{\binom{n-1}{2}-x}$$

With this completely random process of link formation given by a Bernoulli distribution, the degree distribution can easily be derived from the statistical theory. The degree distribution gives us the probability that any given node has a degree of d. Each node can be linked to $n-1$ other nodes, so, the probability of having a degree of d is given by:

$$f(d) = \binom{n-1}{d} p^d (1-p)^{(n-1)-d} \tag{3.1}$$

When $(n-1)p \longrightarrow 0$, the binomial distribution on 3.1 can be approximated by a Poisson distribution. The Poisson degree distribution is given by:

$$f(p) = \frac{e^{-\lambda}\lambda^d}{d!}$$

where $\lambda = (n-1)p$.

When links on a network are formed randomly with a fixed probability of link formation, we say that this network is a Poisson random network. The degree differences across nodes are due to a uniformly random noise on the number os links a node may has.

3.3.2 Power Distribution

The power distribution is given by the following density:

$$f(d) = cd^{-\gamma} \tag{3.2}$$

where c and γ are parameters of this distribution.

Differently from the Poisson distribution, the differences on nodes' degree arise from a cumulative process in which nodes gain connections proportionally to its current degree, such that nodes with a high degree gain connections much faster than low degree nodes. We call this process of "rich-get-richer" process.

When looking to degree distributions of networks, its usual to see fat tails when a network exhibits this process of link formation, because of the high number of low degree nodes. It is also useful to take the logarithm of both degree and its relative frequency, such that we would see a linear relation between the logarithm of degree and the logarithm of its relative frequency[1].

3.4 Clustering Measures

The spread of information or other benefits, as well as harmful things, is much faster when a network is closely-knit. Measures that are able to capture this aspect of cohesiveness in networks have its importance, because enable us to compare different networks in terms of their information transmission. Here, we introduce a way of measuring how clustered networks are.

In a network, the most clustered group possibly would be that in which all its members are connected to each other. We call this kind of cluster of *clique*. A clique is a set of nodes $S \subset N$ such that all nodes in S are connected to each other. Conceptually, we consider a clique a set with at least three nodes, otherwise, any two nodes connected could be a clique.

A simple way of calculating this aspect of cliquishness is looking to all nodes connected to a node i and verify, among all possible pairs, what fraction of them are also connected to each other. For instance, if ij and $ik \in g$, in what fraction does $jk \in g$?

[1]Taking the logarithm of equation 3.2 we have the linear relation $\log(c) - \gamma \log(d)$.

This gives us an idea of the relative number of cliques of size three there exist in the network. Hence, the *overall clustering* is given by:

$$Cl(g) = \frac{\sum_i \#\{jk \in g | k \neq j, j \in N_i(g), k \in N_i(g)\}}{\sum_i \#\{jk | k \neq j, j \in N_i(g), k \in N_i(g)\}} = \frac{\sum_{i;j\neq i;k\neq j;k\neq i} g_{ij}g_{ik}g_{jk}}{\sum_{i;j\neq i;k\neq j;k\neq i} g_{ij}g_{ik}}$$

The overall clustering ranges from 0 to 1. The closer to 1, more clustered a network is, meaning that most of the neighbors of a node are also neighbors.

Next, we introduce some measures that capture an important characteristic of the nodes, the centrality. How central are the nodes in a network and how this influence things like information flows, bargaining power, the spread of diseases, and a sort of other outcomes that may be influenced by a node's centrality.

3.5 Centrality Measures

In the last section, we showed a measure of clustering, that is able to capture an overall aspect of a network. However, we may be interested in microeconomic measures, that capture the influence of the network in the nodes' outcomes. How the neighborhood of a node influence his or her outcomes. Hence, measures that capture how central an individual is in a network can bring great insights about how his or her outcomes are influenced, other than just the economic and sociodemographic environments.

These centrality measures can be categorized into for different groups, and we introduce them next.

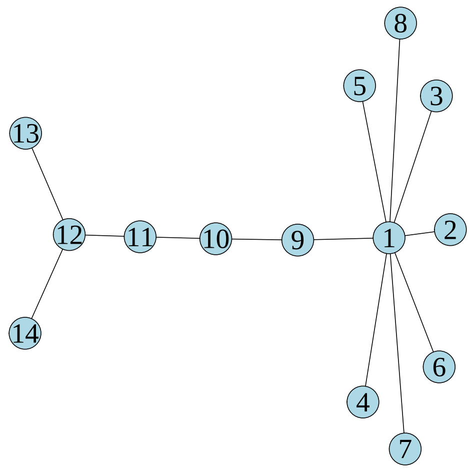
3.5.1 Degree Centrality

The degree centrality tells us how connected a node is in the network. Suppose an individual node in a network (N, g) that is directly linked to all the other $n-1$ nodes. This individual is quite central in this network, because he can reach every single node directly, without the need of any intermediary. On the other hand, a node linked to only one node on the pool o n nodes is not very central. He always need the node he is connected to for reaching any other node.

A simple way of keeping track the centrality of the nodes is look to the fraction of the nodes he is connected to among all the $n-1$ nodes he can potentially be connected to. According to Definition 3.8, the number of nodes an individual is connected to is given by its degree. So, the *degree centrality* of a node i can be calculated by:

$$Ce_i^D(g) = \frac{d_i(g)}{n-1}$$

Figure 3.3: Network on 14 nodes



Prepared by the author.

The degree centrality ranges from 0 to 1. When $Ce_i^D(g) = 0$, node i is not connected to any other node. On the other hand, when $Ce_i^D(g) = 1$, node i is connected to all the other $n - 1$ nodes in the network.

The degree centrality has some limitations, limitations that make its use very restrictive. For illustrating these limitations, suppose the network depicted in figure 3.3. Observe nodes 2 and 12. Node 2 has a degree centrality of 0.0769, while node 12 has a degree centrality of 0.2367, which means that they are connected to 7.69% and 23.67% of all nodes they could be connected to, respectively. However, node 2 is connected to node 1, which has the larger degree centrality, 0.6153, while node 12 is connected to nodes with a much lower degree centrality, 0.1538, 0.0769 and 0.0769 for nodes 11, 13 and 14, respectively. It is easy to see that node 2 is more likely to reach other nodes in the network through indirect connections, because he has a immediate connection with the node with more connections on the network. Although the node that has only one connection is more likely to reach other nodes in the network through indirect connections, the degree centrality for him does not capture this feature. Other centrality measures are needed in order to capture this aspect.

3.5.2 Closeness Centrality

The closeness centrality measures, differently from the degree centrality, which considers only direct connections, capture how easily a node can reach each other. By its definition, it is easy to see its advantage over the degree centrality measure. Instead of looking only to the number of nodes an individual is connected to, here we consider the whole network and how central a node is considering their direct and indirect connections.

Let $l(i, j)$ the number of links on the geodesic[2] between i and j. Averaging $l(i, j)$ over j we have a measure that captures the average path distance between node i and all the other nodes. The bigger it is, the bigger the overall distance from node i to all the other nodes, therefore, less central. It is usual to take the inverse of this measure, such that bigger values mean more centrality:

$$Ce_i^C(g) = \frac{n-1}{\sum_{j \neq i} l(i,j)}$$

This measure ranges from 0 to 1. Even in networks with large diameters[3], and, because of this, a particular node may have a large average distance, when we take the inverse, this distance get closer to 0. On the other side, the least average distance possible is 1, which means that a node is directly connected to all the other nodes in the network. When we take the inverse of the average distance, this node remains with 1 as closeness centrality.

For the network depicted in figure 3.3, the closeness centrality for nodes 2 and 12 is 0.3333 and 0.2766, respectively, which confirms our suspicious that node 2 was more likely than node 12 to reach other nodes through indirect connections, and shows that the closeness centrality overcome the limitations that arise from the use of degree centrality.

Although the average path distance attain to the task of capturing all of the neighborhood effect of a network on an individual, it weights equally nodes that are far away from the individual and nodes directly connected to it. A particular measure that overcome this problem is the *decay centrality.* Consider a decay parameter δ, such that $0 < \delta < 1$. Setting $l(i, j) = \infty$ for nodes that are not path connected with i, the decay centrality,

$$Ce_i^{de}(g) = \sum_{j \neq i} \delta^{l(i,j)}$$

gives more weight for nodes that are closer to the individual, and weight zero for nodes not path connected.

3.5.3 Betweenness Centrality

As the name suggests, the betweenness centrality measures are based on how important a node is in order to connect two distinct nodes. Indirect connections are just

[2]See definition 3.6.

[3]The largest distance between any two nodes.

attainable if it is possible to get in the terminal node by the intermediation of all the nodes that are between the two individuals. If one of these intermediary nodes are not willing to make this intermediation, the indirect connection will not occur.

The first to see the importance of how well situated a node is in terms of intermediate indirect connections was Freeman (1977), and I introduce his betweenness centrality measure next.

Let $P_i(kj)$ the number of geodesics between k and j that goes through i, and $P(kj)$ the whole number of geodesics between k and j. $P_i(kj)/P(kj)$ gives us an idea on how important node i is in connecting k and j. If it is close to 0, i is not very important in connecting the two nodes. It lies in a few paths that connect k and j, so it is unlikely he will be contacted for intermediating the connection. On the other hand, if it is equal to 1, he is always needed for making the connection happen, and, therefore, more central. Averaging $P_i(kj)/P(kj)$ over all combinations of k and j provides an idea on the overall importance of i in connecting the network. Freeman's measure is given by:

$$Ce_i^B(g) = \sum_{k \neq j : i \notin \{k,j\}} \frac{P_i(kj)/P(kj)}{\binom{n-1}{2}}$$

For the network depicted on figure 3.3, the node with larger betweenness centrality is node one, 0.8076, which means that it lies on 80, 76% of all shortest paths on the network he or she belongs. Nodes 2, 3, 4, 5, 6, 7, 8, 13 and 14 have a betweenness centrality of 0, which means that they are never asked to intermediate any indirect connection.

3.5.4 Eigenvector Related Centrality Measures

Eigenvector centrality measures relies on the premise that a node's importance is determined by its neighbors' importance. A first approach to this issue was made by Katz (1953).

The *Katz prestige* of a node i, $P_i^k(g)$, is the sum of the prestige, or centrality, of its neighbors, weighted by theirs respective degrees:

$$P_i^k(g) = \sum_{j \neq i} g_{ij} \frac{P_j^k(g)}{d_j(g)} \tag{3.3}$$

According to its definition, the Katz prestige of a certain node increases with the prestige of its neighbors, compensated by their degree. Even connected to a highly prestigious node, if this node has a lot of other connections, his prestige is diluted among all nodes he is connected to. This can be thought as the time available for a node be connected to other node. More the number of neighbors a prestigious node has, less the time available for each of its neighbors access him.

Considering all the n nodes in a network (N, g), we have the following system for the Katz prestige measure defined in 3.3:

$$\begin{aligned}
P_1^k(g) &= \frac{g_{12}}{d_2(g)} P_2^k(g) + \cdots + \frac{g_{1n}}{d_n(g)} P_n^k(g) \\
P_2^k(g) &= \frac{g_{21}}{d_1(g)} P_1^k(g) + \cdots + \frac{g_{2n}}{d_n(g)} P_n^k(g) \\
&\vdots \\
P_n^k(g) &= \frac{g_{1n}}{d_1(g)} P_1^k(g) + \cdots + \frac{g_{n-1,n}}{d_{n-1}(g)} P_{n-1}^k(g)
\end{aligned} \tag{3.4}$$

If we set the self-links $g_{ii} = 0$, we can write the system of equations in 3.4 in the matrix form as follows:

$$P^k(g) = \hat{g} P^k(g) \tag{3.5}$$

Where $P^k(g)$ is a $n \times 1$ column vector and $\hat{g}$ a $n \times n$ matrix.

Observe that, in order to calculate the Katz prestige measure in 3.5, we need just to find the eigenvector $P^k(g)$ belonging to the unit eigenvalue.

Following Katz (1953), Bonacich (1972) proposes a centrality measure that does not normalize the adjacency matrix by the nodes' degree. His measure, known as *eigenvector centrality*, uses the idea that the centrality of a node is proportional to the sum of the centrality of its neighbors:

$$C_i^e(g) = \frac{1}{\lambda} \sum_{j \neq i} g_{ij} C_j^e(g)$$

Again, letting the self-links $g_{ii} = 0$, we have the following system in matrix form:

$$\lambda C^e(g) = g C^e(g) \tag{3.6}$$

Like in the Katz prestige measure, finding the eigenvector centrality measure in 3.6 is a simple task of calculating the eigenvector $C^e(g)$ belonging to the eigenvalue λ. It is a convention to look for the the eigenvector associated with the largest eigenvalue.

Katz (1953) also introduce a measure of centrality that covers the idea of the power of indirect connections. Remember from section 3.1 that a way of keeping track of how many walks of size k there exist between two distinct nodes is to look to the k^{th} power of the adjacency matrix. Defining a, such that $0 < a < 1$, we can look to the centrality of a node as the weighted sum of the walks that it has emanating from it. More distant nodes receive less weight. The second Katz prestige measure is given by:

$$
\begin{aligned}
P^{K2}(g,a) &= ag\mathbb{I} + a^2g^2\mathbb{I} + a^3g^3\mathbb{I} + \cdots \\
&= (1 + ag + a^2g^2 + \cdots)ag\mathbb{I} \\
&= (\mathbb{I} - ag)^{-1}ag\mathbb{I}
\end{aligned}
$$

Bonacich (1987) extends the second Katz prestige measure as follows:

$$
Ce^B(g,a,b) = (\mathbb{I} - bg)^{-1}ag\mathbb{I} \tag{3.7}
$$

$Ce^B(g,a,b)$, the *Bonacichi centrality*, evaluates walks of length k to other nodes by a factor b^k times the base value of the terminal node, allowing b to differ from a. The parameter b captures how the value of being connected to someone decays with distance, and parameter a captures the base value on each node (JACKSON, 2010).

4 Database

The database used in this research was the *PAEST – Pesquisa de Aspirações e Expectativas dos Estudantes Concludentes do Ensino Médio* – or Survey on Aspirations and Expectations of High School Students. The PAEST collected detailed information about the context students face in their last year of high school. It has information about family structure, socio-demographics, educational achievement and performance, networks through family and friends, sources and quality of information about undergraduate studies and expectations about labor market and educational choices of students in the third grade of high school in the northwestern state of Ceará, in Brazil.

Before talking about the demographic and geographic distribution of the data, it worths to comment about the sample unit in PAEST. The smallest sample unit in the PAEST is, indeed, the student. However, like in the American database Add Health, used in Patacchini e Zenou (2012) for the same purpose as ours, the PAEST has also information about the network structure existing in each classroom the student sampled belongs, allowing us to use models of network effects to study peer effects over educational outcomes, which makes it possible to distinguish the peer effects from the exogenous effects, as shown by Bramoullé, Djebbari e Fortin (2009). This way, the classroom is also the sample unit in the PAEST.

Schools in the cities of Fortaleza, Caucaia, Maracanaú, Eusébio, Sobral, Juazeiro do Norte and Crato, composing three distinct geographical groups, to know, the state's capital (Fortaleza), the Metropoliotan Area of Fortaleza (Caucaia, Maracanaú and Eusébio) and the state's countryside (Sobral, Juazeiro do Norte and Crato), were sampled according to their demographic distribution in the state of Ceará.

In Brazil, there is a clear difference regarding school infrastructure and teaching quality between public schools, which are free of charge, and private schools, in which tuition fees have a large variation from school to school. In general, parents enroll their children in public schools when they can not afford for private education, since public schooling is far from providing good quality education. There are only a few exceptions to this rule, and the large majority include military schools and the Federal Institutes of Education, Science and Technology, the IF'*s*. Said that, the PAEST also stratifies the sample between private and public schools, in order to reflect as closely as possible the

Table 4.1: Geographic distribution of the data in the PAEST.

Region	Private school	Public school	Total
Fortaleza	244	807	1,051
Metropolitan Area of Fortaleza	79	525	604
Ceará countryside	189	498	687
TOTAL	512	1,830	2,342

Prepared by the author.

Table 4.2: Class summary.

	School name	School type	Shift	Class name
1	Santo Inácio	Private	Afternoon	3A
2	Presidente Humberto Castelo Branco	Public	Morning	3D
3	Presidente Humberto Castelo Branco	Public	Afternoon	3E
4	Liceu do Ceará	Public	Afternoon	3A
5	Liceu do Ceará	Public	Morning	3A
6	Figueiredo Correia	Public	Morning	3A
7	Figueiredo Correia	Public	Afternoon	3E
8	Salome Bastos	Private	Morning	3A
9	José de Alencar	Public	Morning	3B
10	José de Alencar	Public	Afternoon	3F
11	Colégio Luciano Feijão	Private	Afternoon	3
12	Colégio Luciano Feijão	Private	Morning	3
13	EEFM Dona Maria Amalia Bezerra	Public	Afternoon	3C
14	EEFM Dona Maria Amalia Bezerra	Public	Morning	3A
15	General Eudoro Correia	Public	Morning	3B
16	General Eudoro Correia	Public	Morning	3D
17	General Eudoro Correia	Public	Afternoon	3F
18	Liceu de Caucaia	Public	Morning	3A
19	Liceu de Caucaia	Public	Morning	3B
20	Liceu de Caucaia	Public	Afternoon	3D

Prepared by the author.

composition of the population. The data is also stratified into morning and afternoon shifts, since most schools in Brazil have these two distinct shifts, with no differences regarding syllabus. In such a way, the data can account for possible differences between students in the two shifts.

Altogether, students and their respective classes from 36 public schools and 11 private schools were sampled, with at least one class in each shift being interviewed, whenever possible. Just in small number of cases either one class or more than two classes in a same school were sampled. The final number of students sampled by the PAEST, divided according geographical region and school type is given in table 4.1.

For this research, a subset of 20 classes from the PAEST were made available. The

Table 4.3: Geographic distribution of the data in the sample.

Region	Private school	Public school	Total
Fortaleza	68	213	281
Metropolitan Area of Fortaleza	0	94	94
Ceará countryside	64	80	144
TOTAL	132	387	519

Prepared by the author.

20 classes came from 10 schools, being 7 public schools and 3 private schools. Each class, with school name, school type, shift and class name is given in table 4.2. Our sample is composed by 519 students from all the three geographical regions of Ceará. The exact stratification of the sample used in this research is given in table 4.3.

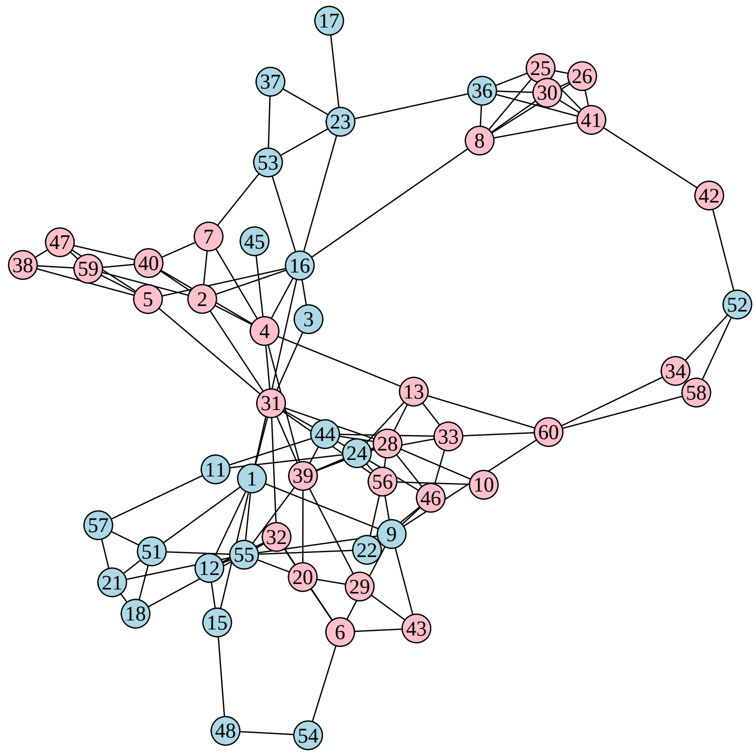
4.1 Networks in PAEST

The PAEST asks each student to list up to five other students in his classroom as his friends. Not always, when a student lists other as his or her friend, this second one lists the first. As some students were absent or refused to participate in the PAEST, a student can also list someone who was not interviewed, but who is part of the network.

In this work, we will consider that, when a student lists other as his friend, there is a mutual relationship between them, even if the second one has not listed the first. In other words, the networks are undirected. Other assumptions can be made about the the nature of the relationships in the networks. We can consider the networks as being directed or give more weight for links in which students list each other. The assumption o undirected network, however, seems to make more sense than the assumption of directed network, since it is very unlike that the information flow has only one direction in networks of friendship.

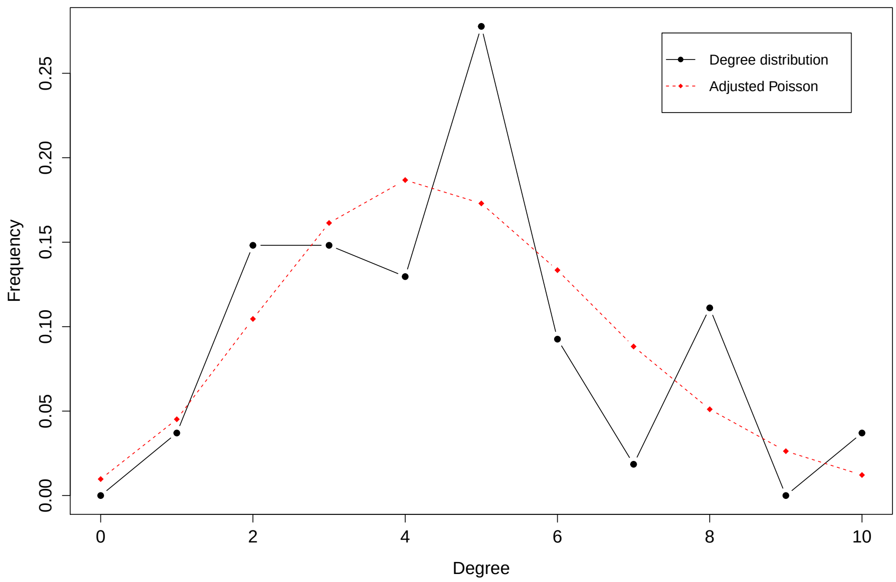
We show the degree distribution of each one of the 20 classes in appendix A, as well as an adjusted Poisson distribution and the graphical representation of the networks. As an example, we show here the graphical representation of the social network for class 3A of Santo Inácio School, afternoon shift, in figure 4.1, and its degree distribution in figure 4.2. The blue nodes represent male individuals and the pink nodes female individuals. As you can see in figure 4.2, the degree distribution clearly follows a Poisson distribution, which means that, at least in this case, the individuals are randomly assigned to that classroom. We can observe the same for all the other 19 classes, as we you can observe in appendix A.

Figure 4.1: Social network for class 3A of Santo Inácio School, afternoon shift.



Prepared by the author.

Figure 4.2: Degree distribution for the network depicted on figure 4.1.



Prepared by the author.

5 Econometric Model

When economists or sociologists try to study the effects of social contacts over individuals' outcomes, or, in other words, peer effects, it is usual to approach this issue including in the econometric model a broad measure summarizing the neighborhood characteristics[1], usually an average of the characteristics of all individuals living or interacting in a same environment.

However, as shown by Manski (1993), this approach has a serious problem of identification, known in the literature as the *reflection problem*. Think about an individual's outcome as a personal choice of effort in exerting some activity. His choice of effort depends on some of his observable characteristics, and also may depends on the choices of effort of his social contacts. As an individual's choice of effort depends on his observable characteristics, the peers' choices of effort also depend on their observable characteristics. Hence, we have to differentiate between the effect of peers' choice of effort, or the endogenous peer effects, and peers' characteristics that do impact on their choice of effort, or the exogenous contextual effects. When the reference group, the group in which each individual belongs, is the same for all individuals, it is impossible to separately identify the endogenous peer effect and the exogenous contextual effect. For illustrating this, suppose the standard peer-effect model, the linear-in-means model, described as following:

$$y_{i,r} = \phi E(y_r) + \gamma E(x_r) + \beta x_{i,r} + \varepsilon_{i,r} \tag{5.1}$$

where $y_{i,r}$ is the outcome of individual i belonging to the reference group r, that arises from some social interaction, (like a GPA, number of criminal activities, etc.), $E(y_r)$ the average of outcome y among all individuals in the reference group r, $x_{i,r}$ an observable characteristic, or a vector of observable characteristics, of individual i in the reference group r (like gender, income, age, educational level, etc.), $E(x_r)$ denotes the average of the observable characteristics of all individuals belonging to reference group r, and $\varepsilon_{i,r}$ an stochastic error term. The parameters ϕ and γ represent the endogenous peer effect and the exogenous contextual effect, respectively. As we said the reference group r is the

[1]A neighborhood here has a different meaning of the neighborhood defined in definition 3.7. Here, we are referring to all the members in a network.

same for all individuals, so, the reference group of an individual in our data set would be his whole class, disregarding the network structure that, indeed, exist.

We show next that it is impossible to identify the endogenous peer effect, ϕ, and separate it from the contextual exogenous effect, γ. Let $E(\varepsilon_{i,r}|y_r, x_r) = 0$. Taking the conditional expectation on both sides of equation 5.1 and solving the resulting expression for $E(y_r)$, we have:

$$E(y_r) = \left(\frac{\gamma + \beta}{1 - \phi}\right) E(x_r) \tag{5.2}$$

Replacing 5.2 on 5.1 yields:

$$y_{i,r} = \left(\frac{\phi(\gamma + \beta) + \gamma(1 - \phi)}{1 - \phi}\right) E(x_r) + \beta x + i, r + \varepsilon_{i,r} \tag{5.3}$$

As you can see on equations 5.1 and 5.3, there are more structural parameters than parameters in the reduced form in 5.3, and it is not possible to separately identify the endogenous peer effect and the exogenous contextual effect.

Identifying the endogenous peer effect is an important task not just in a empirical and technical point of view, but also because of its social multiplier effect, the multiplying power that some policy aiming an individual has to spread over this individual's peers and his indirect contacts. For illustrating this, consider an educational program targeting criminal individuals. An individual affected by this policy may stop his criminal activities, and then he can influence his peers to reduce, or even stop, their criminal activities, and then these individuals may influence their peers in the same way, and so forth. On the other side, policies affecting exogenous variables, like the gender composition of a classroom, will probably not have a social multiplier effect. Fortunately, with the use of the network structure is possible to work around the identification problem.

When the reference group is set as the social contacts of the individuals, or, in other words, when we take into account the network structure of a group of individuals, the reflection problem barely never arises, because the reference group of each agent is different, as shown by Bramoullé, Djebbari e Fortin (2009).

Next, we introduce semi-anonymous graphical games, from where some of the econometric models of network effects, in special the local-average model, are derived, and then we present our empirical strategy for estimating the peer effects in our data set.

5.1 Semi-anonymous Graphical Games

In a social context, the decision for a player to take a given action may depend on the popularity of this action among his or her social or economic contacts. As an example, think about smoking or criminal behavior, R&D efforts, the adoption of a new technology by firms, the allocation of time between study and leisure, etc. However, as the player influenced by the actions of his/her social contacts is also part of the social contacts of other players, his/her action also influence them. Hence, all these examples are also examples of strategic interactions, and, therefore, can be modeled by graphical games, in this case, by the called anonymous and semi-anonymous graphical games.

In a anonymous graphical game, the decision of a player to take a given action depends on the number of other players taking this same action in a symmetric way. His or her decision does not depend on who is taking the action. Differently, in a semi-anonymous graphical game, his or her decision in taking an action depends on the number of his direct contacts, or friends, taking the action.

There are two different classes of semi-anonymous graphical games, relatively to how the utility of taking or not an action changes with the number of friends taking this action. When the utility increases, we say that the strategic interaction is a game of *strategic complements*, and includes most of the social and economic situations, like those cited above. On the other side, when the utility of taking an action decreases with the number of contacts taking the same action, the interaction is said to be a game of *strategic substitutes*, and encompass situations like the provision of a public good and information gathering (JACKSON, 2010). Next, we formally define games of strategic complements, since they encompass the relation we are modeling on this dissertation.

Definition 5.1. Let $u_i(x_i, m) : \mathbb{R}_+^2 \longrightarrow \mathbb{R}$ a real-valued utility function, where x_i denotes the action of player i, with $x_i = 1$ if the player is taking a given action and $x_i = 0$ otherwise, and m the number of players in $N_i(g)$, the neighborhood of i, with $x_j = 1$. Let $m > m'$. A semi-anonymous graphical game exhibits *strategic complements* if it satisfies:

$$u_i(1, m) - u_{d_i}(0, m) > u_i(1, m') - u_i(0, m')$$

for all $i \in N$.

The model we introduce next, the local-average model, is an example of a semi-anonymous graphical game of strategic complements, since the utility of taking an action increases with the average sum of active links an individual has. In this model, the social multiplier is due to the average effort of a player's friends in exerting some activity, like studying or committing a crime.

5.2 The Local-average Model

Let $N_r = \{1, \ldots, n_r\}$ a finite set of nodes in the adjacency matrix g_r, for $r = 1, \ldots, \overline{r}$, where r index the network wherein player i belongs and $\overline{r}$ is the number of networks in our sample. On this game, each player decides the level of effort to exert in some activity. The level of effort of individual i in network r is denoted by $y_{i,r}$. If we set the self-links $g_{ii} = 0$, the utility function of player i in network r as function of his effort and the effort of his peers is given by:

$$u_{i,r}(\mathbf{y}_r, g_r) = (a^*_{i,r} + \eta^*_r + \varepsilon^*_{i,r})y_{i,r} - \frac{1}{2}y^2_{i,r} - \frac{\lambda}{2}(y_{i,r} - \overline{y}_{i,r})^2 \tag{5.4}$$

where $\mathbf{y}_r$ is an n_r-dimensional vector of efforts. This utility function has two parts, an individual part and a part representing the local-average effect, given by the last term on the righthand side of equation 5.4. The term $a^*_{i,r}$ is a measure that captures the observable characteristics of player i (like gender, age, years of study, etc.) and the observable average characteristics of player i's direct contacts (like the fraction of male or female friends, their average age, their average years of study, etc.), and can be written as:

$$a^*_{i,r} = \sum_{m=1}^{M} \beta_m x^m_{i,r} + \frac{1}{d_i(g_r)} \sum_{m=1}^{M} \sum_{j=1}^{n_r} g_{ij,r} x^m_{j,r} \gamma_m$$

where $\mathbf{x}_{i,r} = \{x^1_{i,r}, \ldots, x^M_{i,r}\}$ is a set of M variables for the observable characteristics of player i, and β_m and γ_m parameters associated to the variables $x^m_{i,r}$ and $x^m_{j,r}$, respectively. Back to the utility function in 5.4, η^*_r denotes the unobservable network characteristics, an unobserved heterogeneity among networks, and $\varepsilon^*_{i,r}$ a stochastic error term.

On the local-average effect, given by

$$\frac{\lambda}{2}(y_{i,r} - \overline{y}_{i,r})^2$$

the term $\overline{y}$ denotes the average level of effort of player i's social contacts, and is given by:

$$\overline{y}_{i,r} = \frac{1}{d_i(g_r)} \sum_{j=1}^{n_r} g_{ij,r} y_{j,r} = \sum_{j=1}^{n_r} g^*_{ij,r} y_{j,r} \tag{5.5}$$

where $d_i(g_r)$ is the degree of player i in network r, as defined in definition 3.8, and $g^*_{ij,r} = g_{ij,r}/d_i(g_r)$. The matrix with each entrance given by $g^*_{ij,r}$, the matrix g^*_r, is a row-normalization of the adjacency matrix g_r.

The local-average effect captures how the choices of effort by player i's social contacts influence player i's own choices of effort. On equation 5.4, observe that the utility of player i decreases with his distance from the average level of effort of his social contacts. This way, each individual has as an objective to minimize the social distance between himself and his reference group. For instance, it is likely that an individual with low performance in mathematics in a group of individuals with high performance in the subject will try to improve his performance to conform with his peers. Because of this, the parameter λ is sometimes called the *taste for conformity*. An individual loses utility if he does not conform to his social contacts.

Each player will choose a level of effort that maximizes his utility function. Therefore the Nash equilibrium in pure strategies for this graphical game, $\mathbf{y}_r^* \in \mathbb{R}^{n_r}$, satisfies, for all r and all $i \in N_r$:

$$y_{i,r}^* = \phi \sum_{j=1}^{n_r} g_{ij,r}^* y_{j,r}^* + a_{i,r} + \eta_r + \varepsilon_{i,r} \tag{5.6}$$

where $\phi = \lambda/(1+\lambda)$, $a_{i,r} = a_{i,r}^*/(1+\lambda)$, $\eta_r = \eta_r^*/(1+\lambda)$ and $\varepsilon_{i,r} = \varepsilon_{i,r}^*/(1+\lambda)$.

It is possible to show that, if $\phi < 1$, the system of best-reply functions in 5.6 has a unique solution, given by:

$$\mathbf{y}_r^* = (\mathbf{I}_{n_r} - \phi g_r^*)^{-1} \alpha_r g_r^* \mathbf{I}_{n_r} \tag{5.7}$$

for $r = 1, \ldots, \overline{r}$. Where $\alpha_r = a_{i,r} + \eta_r + \varepsilon_{i,r}$.

Observe that the Nash equilibrium on equation 5.7 is the same as the Bonacichi Centrality represented on equation 3.7. This way, the parameter ϕ captures how the benefit of being path-connected to someone decays with the distance between them.

5.2.1 Conditions for Identification of the Local-average Model

Differently from the linear-in-means model in equation 5.1, on the local-average model, each individual is influenced by his or her peers in a different way, since no one has the same reference group. Hence, each player is differently influenced by his/her peer's choices for effort. In contrast, when the reference group is the same for all individuals, the peer effects are an externality that affects identically every individual in a neighborhood. Besides not being a plausible hypothesis, it presents empirical problems that make its use too restrictive. In this section, we explain why the use of network effects works around the identification problem, making its use ideal for finding endogenous peer effects.

Let r_i the reference group of individual i on network r. Rewriting the linear-in-means model for the hypothesis of different reference groups and including a network unobservable heterogeneity, η_r, we have the following econometric model:

$$y_{i,r_i} = \phi E(y_{r_i}) + \gamma E(x_{r_i}) + \beta x_{i,r_i} + \eta_r + \varepsilon_{i,r_i} \tag{5.8}$$

where $E(y_{r_i}) = \overline{y}_{i,r}$, as defined on equation 5.5, $E(x_{r_i})$ denotes the average characteristics od player i's social contacts, and $\gamma E(x_{r_i}) + \beta x_{i,r_i} \equiv a_{i,r}$, as in equation 5.6. Model 5.8 is exactly the same as the Nash equilibrium of the local-average model on equation 5.6. If we set the self-links $g_{ii,r} = 0$, we can write the model in 5.8 in matrix form, as follows:

$$\mathbf{Y}_r = \phi g_r^* \mathbf{Y}_r + \mathbf{X}_r \boldsymbol{\beta} + g_r^* \mathbf{X}_r \boldsymbol{\gamma} + \eta_r \mathbf{1}_{n_r} + \boldsymbol{\varepsilon}_r \tag{5.9}$$

where $\mathbf{Y}_r$ is an $n_r \times 1$ vector of observations on the dependent variable, $\mathbf{X}_r$ is an $n_r \times M$ matrix of observations on the M exogenous variables, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ an $M \times 1$ vector of parameters, $\boldsymbol{\varepsilon}_r$ a stochastic error term, $\mathbf{1}_{n_r}$ an $n_r \times 1$ vector of 1, and ϕ a scalar parameter for the endogenous peer effects. Assume $E(\boldsymbol{\varepsilon}_r | g_r, \mathbf{X}_r)$=0. Then, the model on equation 5.9 is similar to a spatial autoregressive model.

We can eliminate the unobservable heterogeneity term $\eta_r \mathbf{1}_{n_r}$ deviating the model from its mean, using the deviation from group mean projector $\mathbf{J}_r = \mathbf{I}_{n_r} - \frac{1}{n_r} \mathbf{1}_{n_r} \mathbf{1}_{n_r}^T$. Pre-multiplying the model on equation 5.9 by $\mathbf{J}_r$, we have:

$$\mathbf{J}_r \mathbf{Y}_r = \phi \mathbf{J}_r g_r^* \mathbf{Y}_r + \mathbf{J}_r \mathbf{X}_r \boldsymbol{\beta} + \mathbf{J}_r g_r^* \mathbf{X}_r \boldsymbol{\gamma} + \mathbf{J}_r \boldsymbol{\varepsilon}_r \tag{5.10}$$

since $\mathbf{J}_r \mathbf{1}_{n_r} = 0$.

Bramoullé, Djebbari e Fortin (2009) demonstrate that, if $\phi\beta + \gamma \neq 0$, the identification of the local-average model on 5.10 is possible, since $\mathbf{J}_r g_r^{*2} \mathbf{X}_r$, $\mathbf{J}_r g_r^{*3} \mathbf{X}_r$, ... can be used as instrumental variables. In a network, social connections are not necessarily transitive, which means that if a player i is connected to two other players, k and j, for instance, k and j are not necessarily connected to each other. Hence, it is unlikely that the average characteristics of "friends' friends", $g_r^{*2} \mathbf{X}_r$, may be perfectly correlated with individuals' own characteristics, $\mathbf{X}_r$, and with the average characteristics of his friends, $g_r^* \mathbf{X}_r$. Therefore, $\mathbf{J}_r g_r^{*2} \mathbf{X}_r$, $\mathbf{J}_r g_r^{*3} \mathbf{X}_r$, ... are excellent instrumental variables to identify the endogenous peer effects and the exogenous contextual effects. Bramoullé, Djebbari e Fortin (2009) show that, if $\mathbf{I}_{n_r}$, g_r and g_r^2 are linearly independent, the model and, thus, the network effects, is identified. If $g_r^2 \neq 0$ for each r, which means that there is at least one path of size two between any two nodes, then $\mathbf{I}_{n_r}$, g_r and g_r^2 are linearly independent.

However, if individuals are not randomly assigned into networks, the estimated network effects may not represent the true population parameter. As pointed out by Topa e Zenou (2015), if the variables that drive this process of selection are not fully observable, potential correlations between unobserved group-specific factors and the target regressors are the major sources of bias. We must then be sure that individuals in a network were randomly sorted into that network, to avoid a bias in the estimated network effect. As showed in chapter 3, when links on a network are formed randomly with a fixed probability of link formation, the degree distribution of this network follows a Poisson distribution, and the network with this feature is called a Poisson random network. However, the use of network fixed effects works around this problem. Even if individuals self-select into

Table 5.1: Maximum likelihood estimation of network effects.

Autoregressive variable	Estimated ϕ	p-value
Doing an exam without being prepared	0.26257	0.0268
Cheating in an exam	0.12626	0.6011

Prepared by the author.

networks and then link formation takes place, if linking decisions are uncorrelated with the observable variables, this two-step model of link formation generates network fixed effects, as pointed out by Bramoullé, Djebbari e Fortin (2009). This way, by multiplying the local-average model by the deviation from group mean projector, $\mathbf{J}_r$, we can identify the peer effects correctly.

5.3 Results

As said in the last section, in order to identify the endogenous peer effects and separate it from the exogenous contextual effects, the matrices $\mathbf{I}_{n_r}$, g_r and g_r^2 must be linearly independent for each network in our database, and this condition is satisfied when $g_r^2 \neq 0$, which means that there is at least one path of size two between any two nodes in the network r. For the 20 networks used in this work, $g_r^2 \neq 0$. This way, we can use the local-average model in equation 5.10 to find the peer effects and its social multiplier effect.

According to the table 5.1, the estimated network effect, ϕ, for the variable *cheating in an exam*, is statistically insignificant, since the null hypothesis of ϕ=0 is rejected only for significance levels greater than 60.11%. However, the variable *doing an exam without being prepared* is statistically significant for any significance level greater than 2.68%.

Remember that $\phi = \lambda/(1 + \lambda)$, where the parameter λ is the called *taste for conformity*. For the variable *doing an exam without being prepared*, $\lambda = 0.207964707$, and $\lambda/2 = 0.1039824$, which means that an one-unit increase in the individual i's taste for conformity, or, equivalently, in the average behavior of his reference group regarding doing an exam without being prepared, increases individual i's probability of doing an exam without being prepared in 10.39%.

Also, notice that individual i's behavior increases with the average behavior of his reference group, showing that the social interaction among the students is, indeed, a game of strategic complements, as we stated in section 5.1.

6 Final Considerations

The actions for which we have estimated the local-average model in the last chapter, which are doing an exam without being prepared and cheating in an exam, can be considered behaviors driven by a game of strategic complements. It is not hard to accept that, when an individual not much willing to commit any of those two misbehaviors becomes part of a group in which the individuals there do it in a greater extent, he or she will probably increase the probability of doing an exam without being prepared and/or cheating in an exam. Our results support this statement, at least for the variable *doing an exam without being prepared*, since the positive sign of the network effect, or, specifically, the taste for conformity, evidences that a player tends to follow the actions of his or her peers. A positive sign for the network effect in the local-average model indicates that a player will increase or decrease his or her choices of effort with the respective increasing or decreasing choices of effort of his or her peers.

The results found here have also implications for educational policies of reduction of poor academic performance of high school students in the process of applying for the entrance exam in Brazilian universities, the ENEM. Such policies have what we previously denominated social multiplier effect. An educational policy that encourages a student to previously study the contents taught in the classroom, so that he or she can apply his or her knowledge in an exam, will not affect only this student affected by this policy, but will also affect, in a positive way, the behavior of those students in the social circle of this student. This result may help to reduce the cost of implementation of such policies, since policy makers may not be considering the social multiplier effect that policies aiming to change a social behavior have.

Patacchini e Zenou (2012) show that the value of $\hat{\phi}$ decreases with the increasing of the set of controls, indicating that the estimated network effect may be capturing important confounding factors. However, qualitative results remain unchanged. Following this finding, an important improvement to this work would be collect information about factors that make us able to better differentiate between the endogenous network effects and the exogenous contextual effects.

For many other variables in a school environment we can argue that we observe the effects of conformism as shaping behavior of students. Patacchini e Zenou (2012) studies the effects of conformism in the criminal behavior of high school students. Conformism may also plays an important role in academic performance measured by grades or GPA; In the choice of time dedicated to study; in sexual behavior; in career choices, etc. A natural extension to this work would be, therefore, estimate the network effects for these variables.

We can make other assumptions regarding the social structure of the networks. We can make a robustness check by testing undirected networks versus directed networks, or a weighted network against a non weighted network. As explained in the chapter 4, a student in the survey lists up to five other students in the classroom as his or her friend. Not necessarily a listed student will list the first as his friend, but it is somewhat clear that there is a social relation between them, although only one of them lists the other as his or her friend, which is what usually happens in networks of friendship. Then, we can test the assumptions of directed networks and weighted networks, in which we give a weight greater than one to links connecting students who cite each other, against the hypothesis of undirected network.

For last, there are other models of games of strategic complements, like the local-aggregate model. Liu, Patacchini e Zenou (2014) propose a test to evaluate whether the local-average model is more relevant in some activities than the local-aggregate model, and vice versa. It is possible that, instead of conformism, the sum of active links plays the role of shaping individual behavior.

Bibliography

ASLUND, O. et al. Peers, neighborhoods, and immigrant student achievement: Evidence from a placement policy. *American Economic Journal: Applied Economics*, v. 3, n. 2, p. 67–95, 2011.

BALLESTER, C.; CALVó-ARMENGOL, A.; ZENOU, Y. Who's who in networks. wanted: The key player. *Econometrica*, v. 74, n. 5, p. 1403–1417, 2006.

BALLESTER, C.; CALVó-ARMENGOL, A.; ZENOU, Y. Delinquent networks. *Journal of the European Economic Association*, v. 8, n. 1, p. 34–61, 2010.

BEAMAN, L. A. Social networks and the dynamics of labor market outcomes: Evidence from refugees resettled in the u.s. *Review of Economic Studies*, v. 79, n. 1, p. 128–161, 2012.

BONACICH, P. Factoring and weighting approaches to status scores and clique identifcation. *Journal of Mathematical Sociology*, v. 2, n. 1, p. 113–120, 1972.

BONACICH, P. Power and centrality: a family of measures. *American Journal of Sociology*, v. 92, n. 5, p. 1170–1182, 1987.

BRAMOULLé, Y.; DJEBBARI, H.; FORTIN, B. Identification of peer effects through social networks. *Journal of Econometrics*, v. 150, n. 1, p. 41–55, 2009.

BROOKS-GUNN, J. et al. Do neighborhoods influence child and adolescent development? *American Journal of Sociology*, v. 99, n. 2, p. 353–395, 1993.

CALVó-ARMENGOL, A.; ZENOU, Y. Social networks and crime decisions: The role of social structure in facilitating delinquent behavior. *International Economic Review*, v. 45, n. 3, p. 939–958, 2004.

CORCORAN, M. et al. Effects of family and community background on men's economic status. NBER Working Paper No. 2896. 1989.

EDIN, P.-A.; FREDRIKSSON, P.; ASLUND, O. Ethnic enclaves and the economic success of immigrants - evidence from a natural experiment. *The Quarterly Journal of Economics*, v. 118, n. 1, p. 329–357, 2003.

ERDOS, P.; RENYI, A. On random graphs. *Publicationes Mathematicae*, v. 6, n. 1, p. 290–297, 1959.

FREEMAN, L. C. A set of measures of centrality based on betweenness. *Sociometry*, v. 40, n. 1, p. 35–41, 1977.

GLAESER, E. L.; SACERDOTE, B.; SCHEINKMAN, J. A. Crime and social interactions. *The Quarterly Journal of Economics*, v. 111, n. 2, p. 507–548, 1996.

JACKSON, M. O. *Social and Economic Networks*. Princeton: Princeton University Press, 2010.

JACOB, B. A. Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in chicago. *The American Economic Review*, v. 94, n. 1, p. 233–258, 2004.

KATZ, L. A new status index derived from sociometric analysis. *Psychometrika*, v. 18, p. 39–43, 1953.

KLING, J.; LIEBMAN, J.; KATZ, L. Experimental analysis of neighborhood effects. *Econometrica*, v. 75, n. 1, p. 83–119, 2007.

KLING, J. R.; LUDWIG, J.; KATZ, L. F. Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics*, v. 120, n. 1, p. 87–130, 2005.

LIU, X.; PATACCHINI, E.; ZENOU, Y. Endogenous peer effects: local aggregate or local average? *Journal of Economic Behavior & Organization*, v. 103, n. 1, p. 39–59, 2014.

LUDWIG, J. et al. Neighborhood effects on the long-term well-being of low-income adults. *Science*, v. 337, n. 6101, p. 1505–1510, 2012.

LUDWIG, J.; DUNCAN, G. J.; HIRSCHFIELD, P. Urban poverty and juvenile crime: Evidence from a randomized housing-mobility experiment. *The Quarterly Journal of Economics*, v. 116, n. 2, p. 655–679, 2001.

MANSKI, C. F. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, v. 60, n. 3, p. 531–542, 1993.

OREOPOULOS, P. The long-run consequences of living in a poor neighborhood. *The Quarterly Journal of Economics*, v. 118, n. 4, p. 1533–1575, 2003.

PATACCHINI, E.; ZENOU, Y. Juvenile delinquency and conformism. *The Journal of Law, Economics, & Organization*, v. 28, n. 1, p. 1–31, 2012.

POPKIN, S. J.; ROSENBAUM, J. E.; MEADEN, P. M. Labor market experiences of low-income black women in middle-class suburbs: Evidence from a survey of gautreaux program participants. *Journal of Policy Analysis and Management*, v. 12, n. 3, p. 556–573, 1993.

TOPA, G. Social interactions, local spillovers and unemployment. *Review of Economic Studies*, v. 68, n. 2, p. 261–295, 2001.

TOPA, G.; ZENOU, Y. Handbook of regional and urban economics. In: ______. Amsterdam: Nikki Levy, 2015. v. 5A, cap. Neighborhood and Network Effects, p. 561–624.

A Networks Representation and Degree Distributions

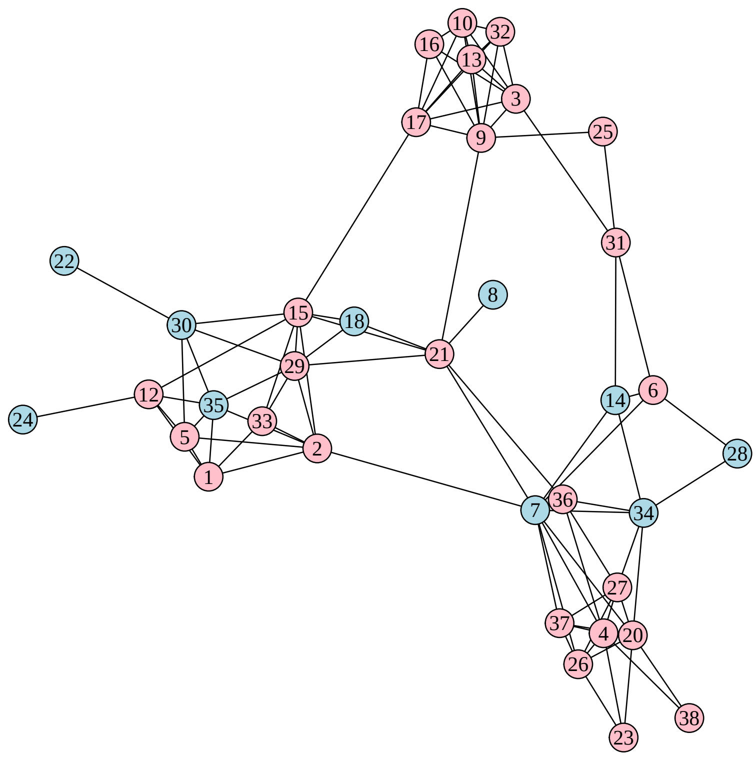


Figure A.1: Social network for class 3A of Salomé Bastos School, morning shift.

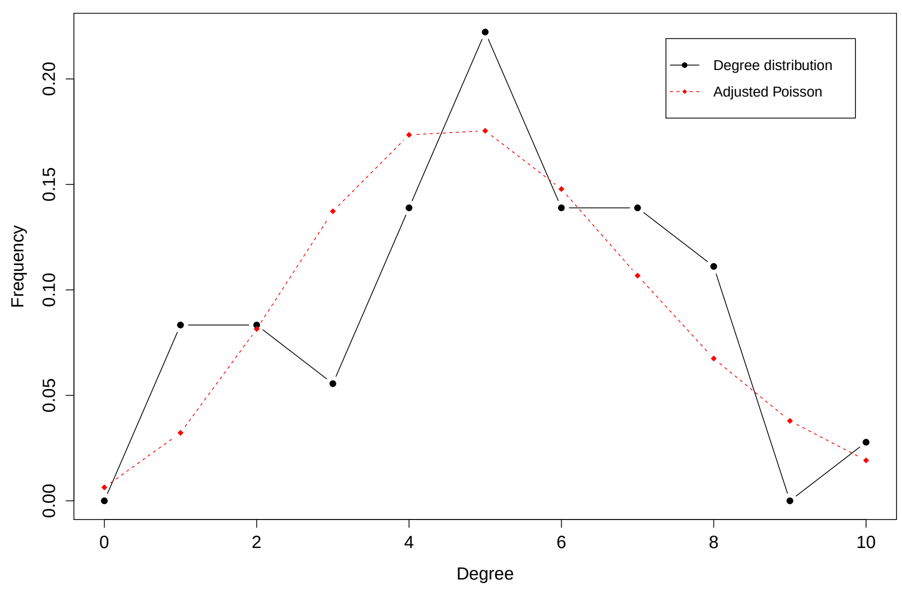


Figure A.2: Degree distribution for the network depicted on figure A.1.

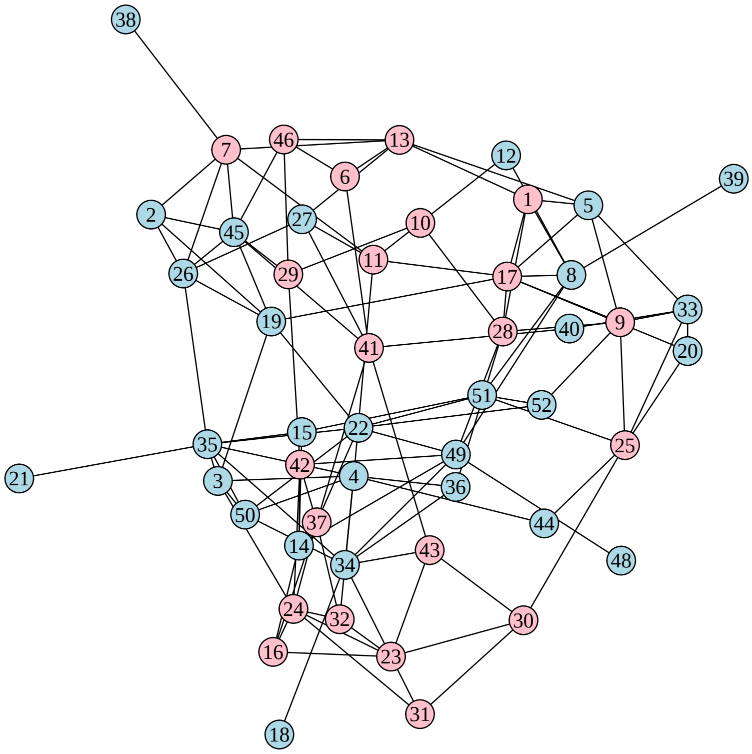


Figure A.3: Social network for class 3 of Luciano Feijão School, morning shift.

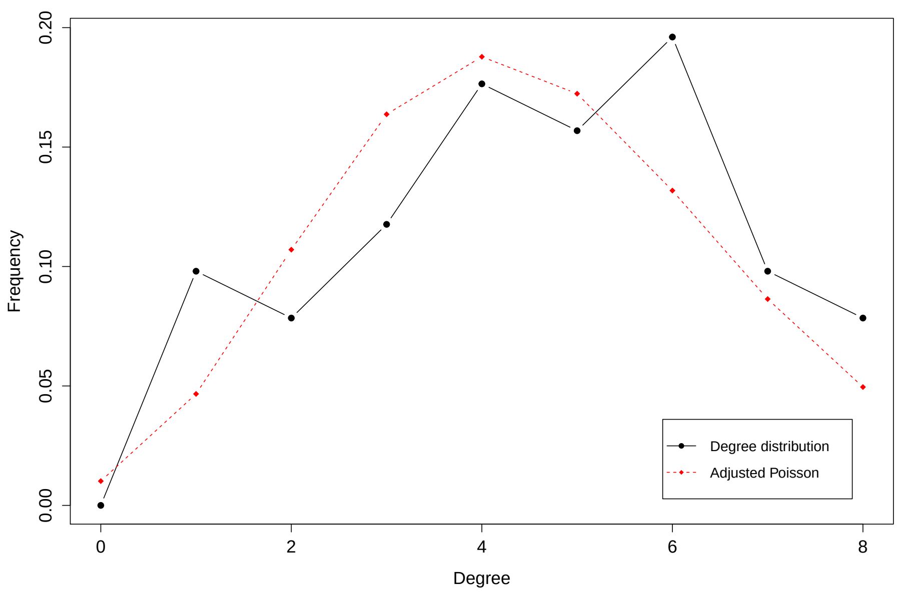


Figure A.4: Degree distribution for the network depicted on figure A.3.

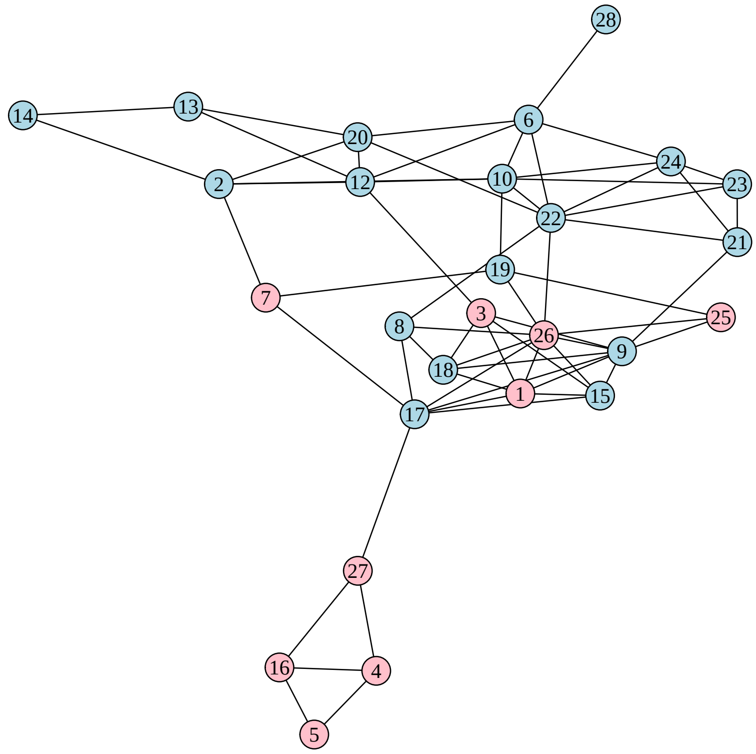


Figure A.5: Social network for class 3 of Luciano Feijão School, afternoon shift.

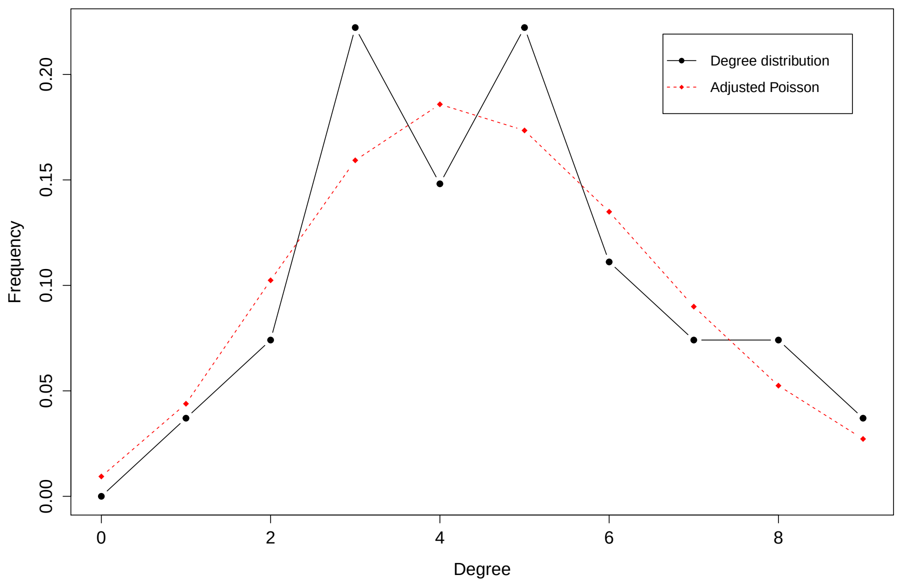


Figure A.6: Degree distribution for the network depicted on figure A.5.

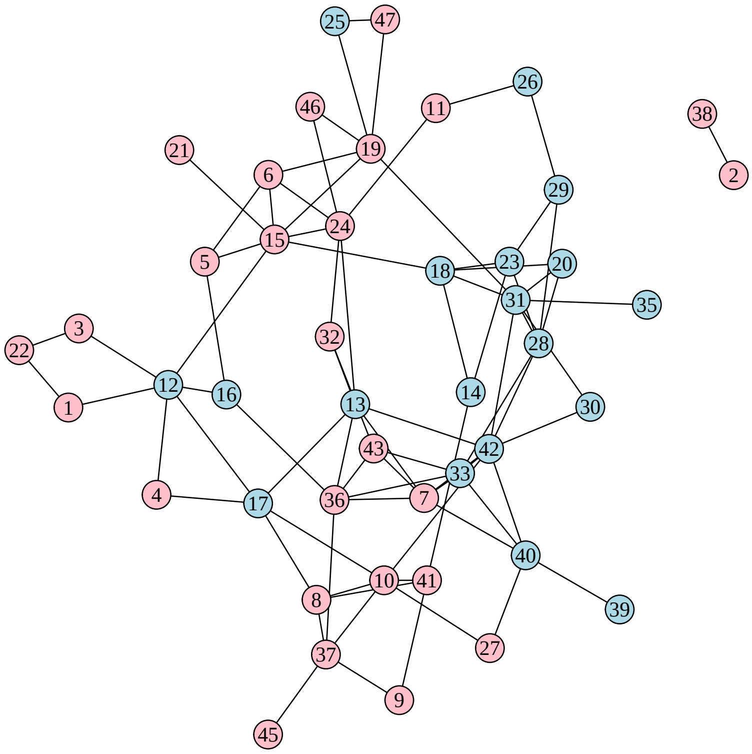


Figure A.7: Social network for class 3B of José de Alencar School, morning shift.

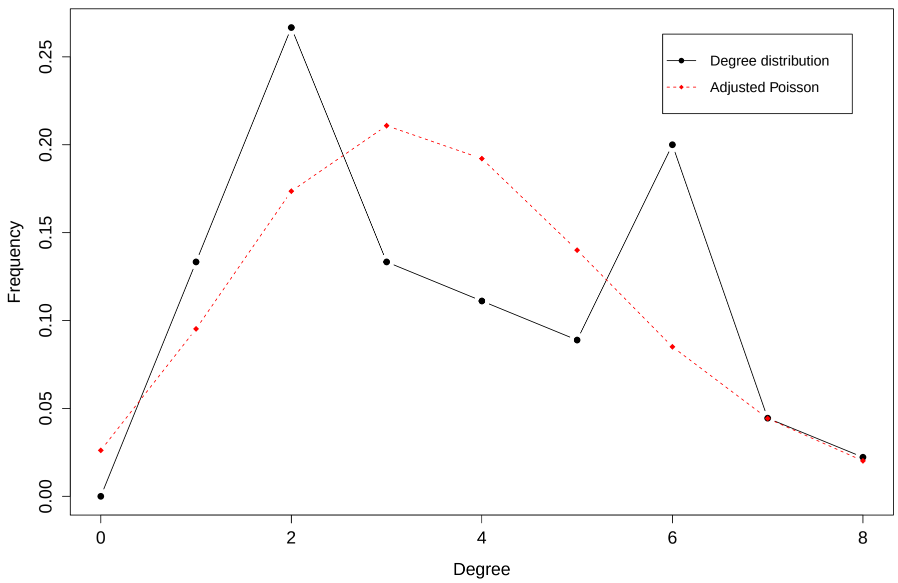


Figure A.8: Degree distribution for the network depicted on figure A.7.

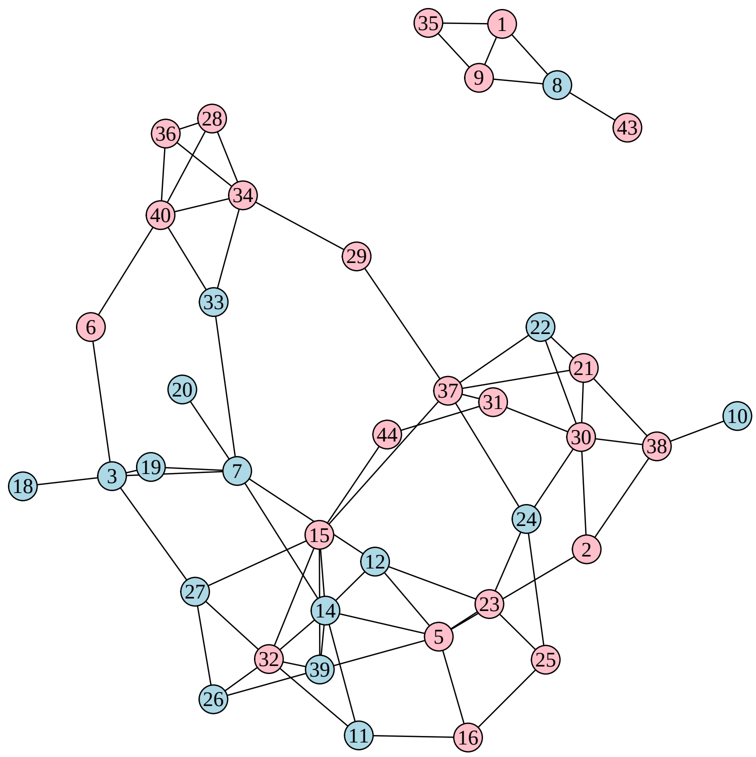


Figure A.9: Social network for class 3F of José de Alencar School, afternoon shift.

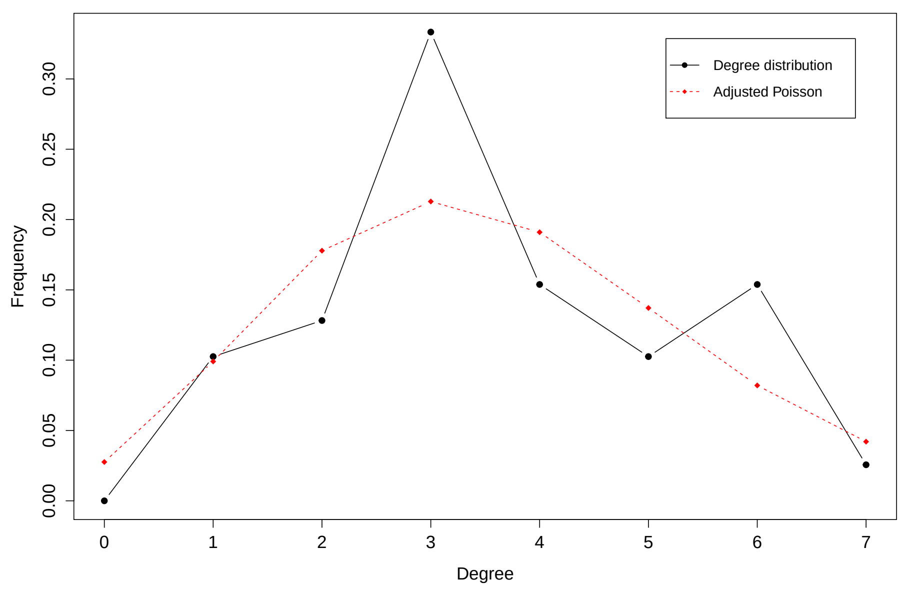


Figure A.10: Degree distribution for the network depicted on figure A.9.

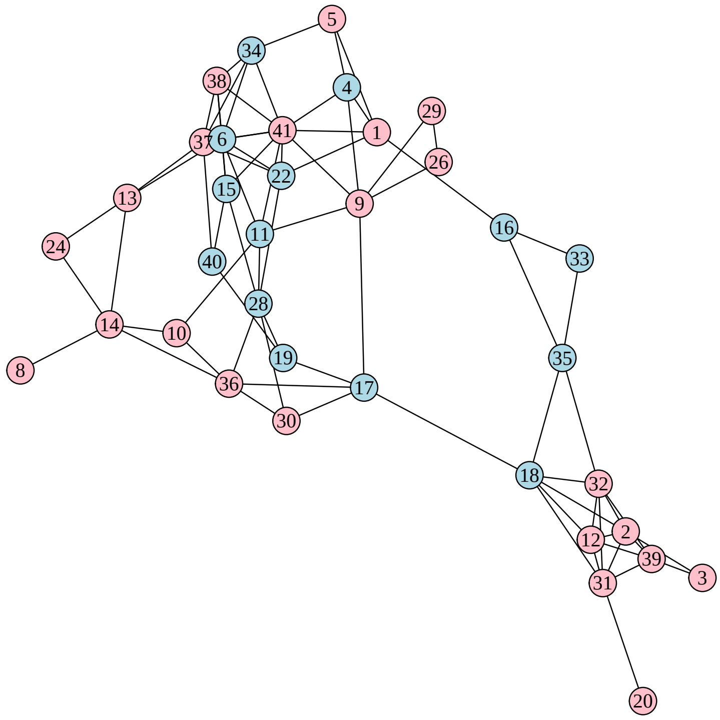


Figure A.11: Social network for class 3D of Presidente Humberto Castelo Branco School, morning shift.

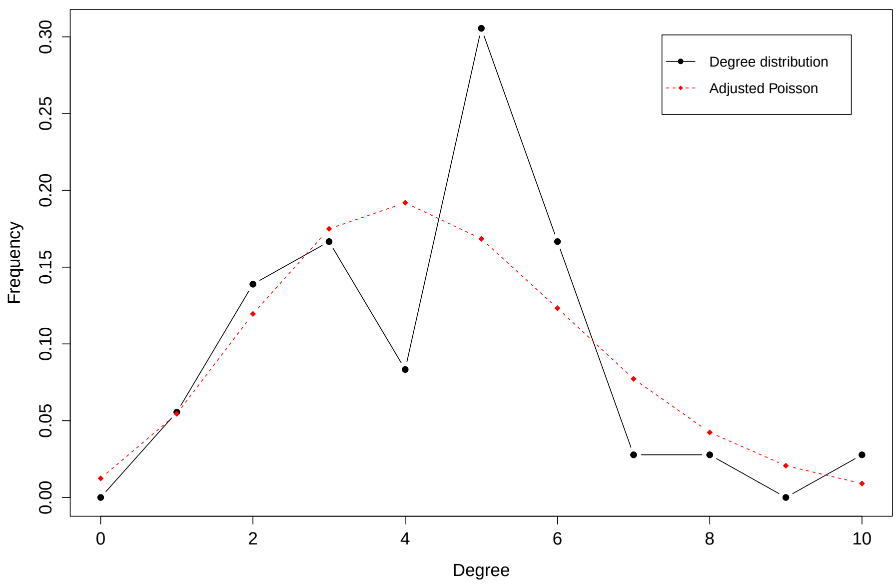


Figure A.12: Degree distribution for the network depicted on figure A.11.

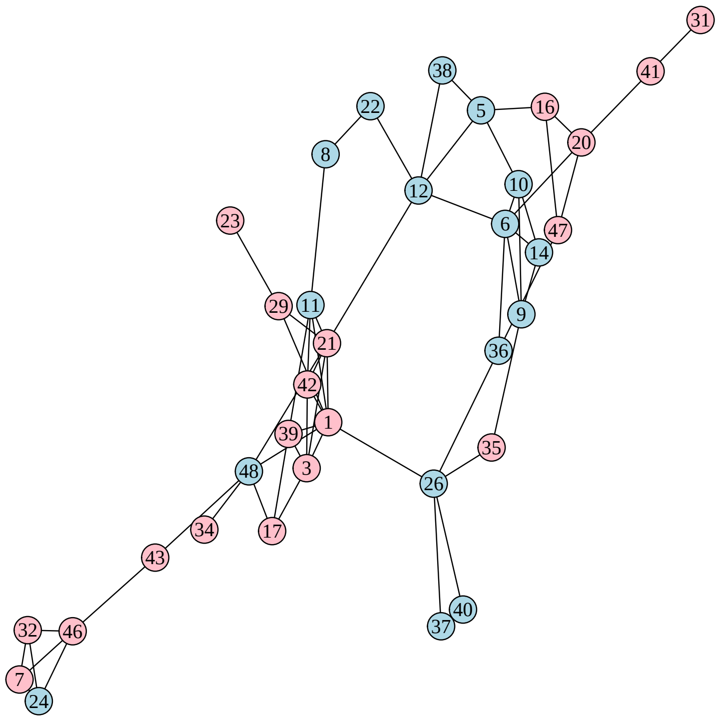


Figure A.13: Social network for class 3E of Presidente Humberto Castelo Branco School, afternoon shift.

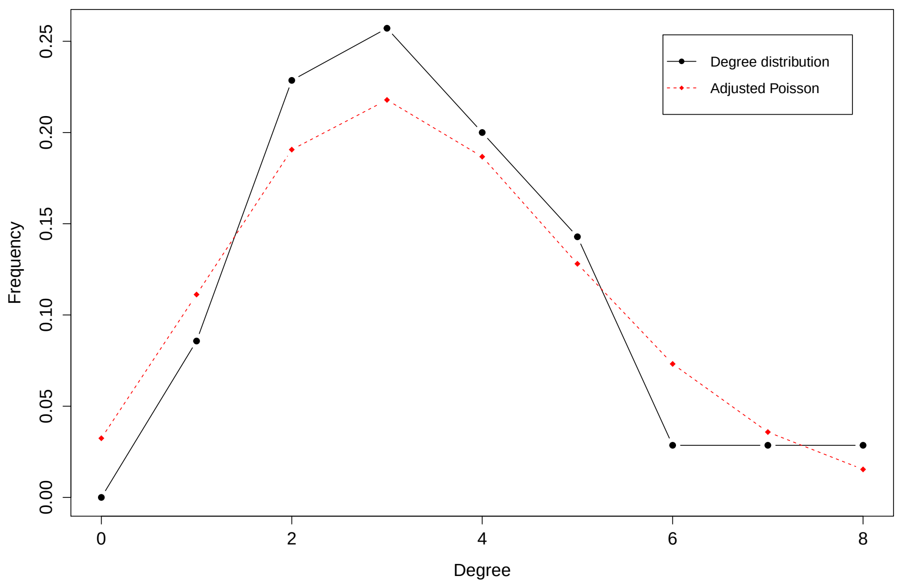


Figure A.14: Degree distribution for the network depicted on figure A.13.

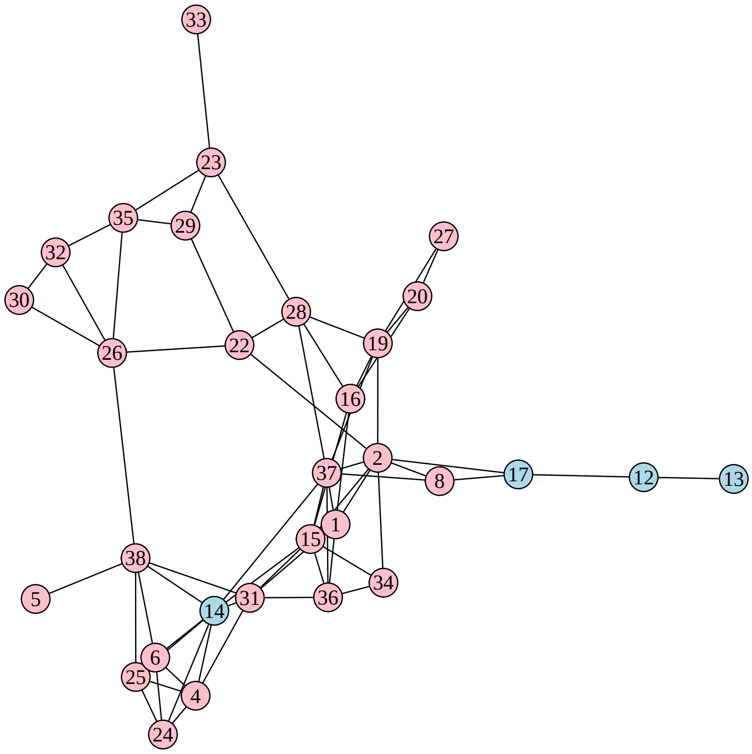


Figure A.15: Social network for class 3E of Figueiredo Correia School, afternoon shift.

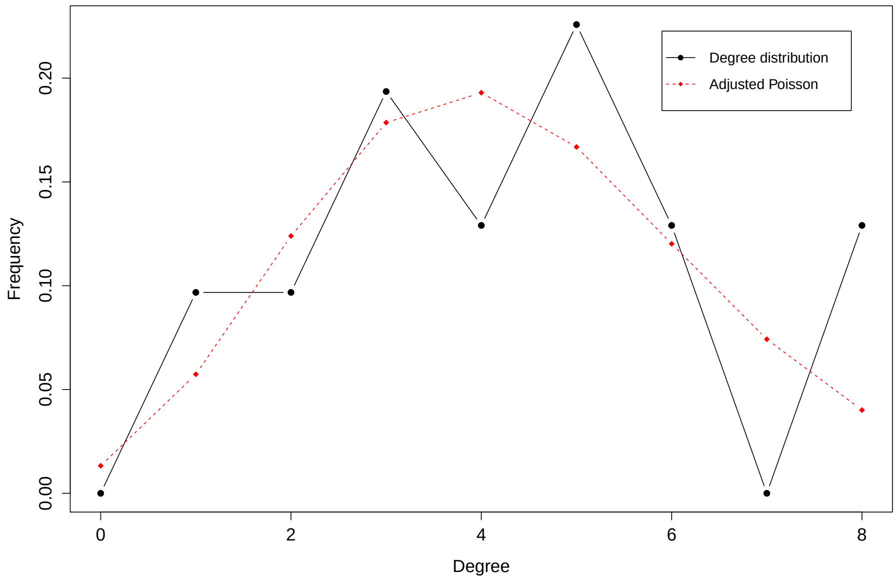


Figure A.16: Degree distribution for the network depicted on figure A.15.

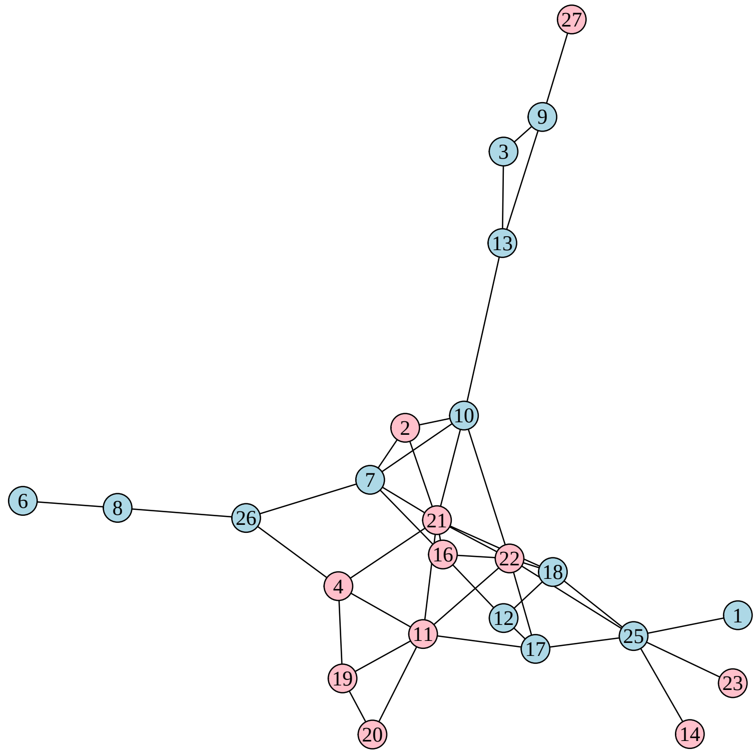


Figure A.17: Social network for class 3A of Figueiredo Correia School, morning shift.

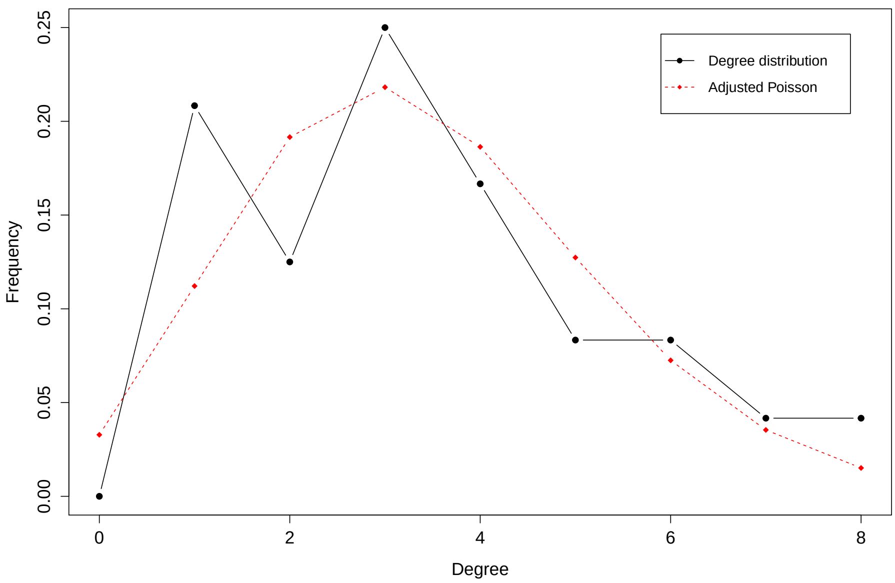


Figure A.18: Degree distribution for the network depicted on figure A.17.

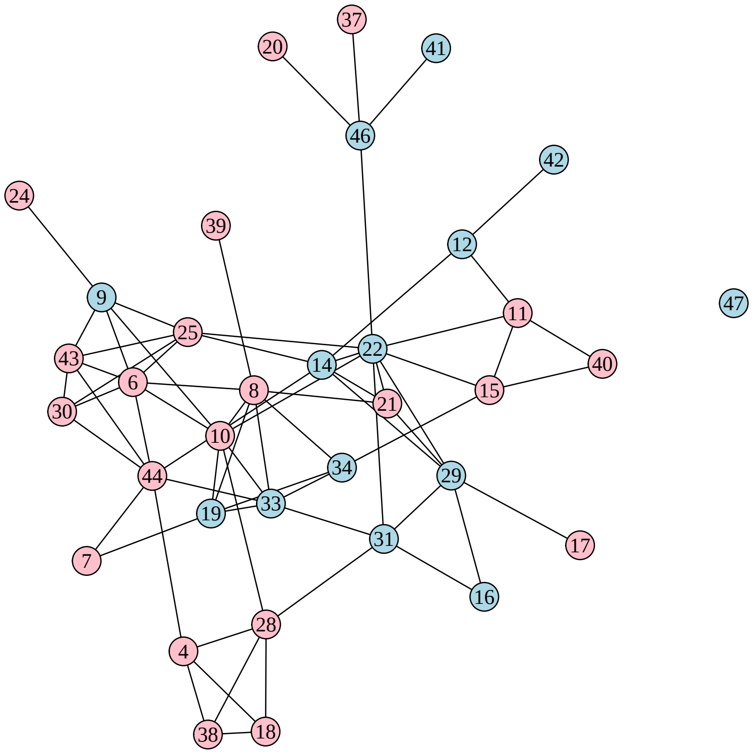


Figure A.19: Social network for class 3A of Liceu do Ceará School, morning shift.

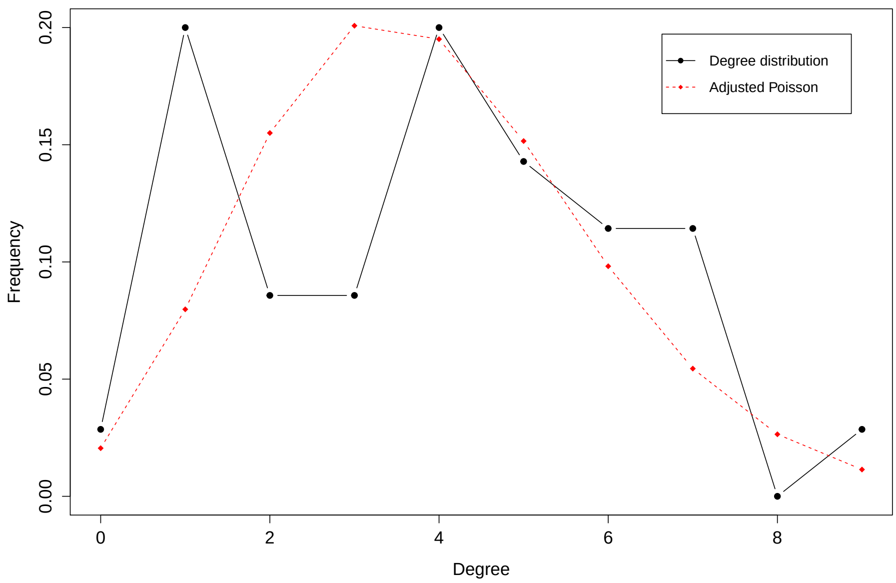


Figure A.20: Degree distribution for the network depicted on figure A.19.

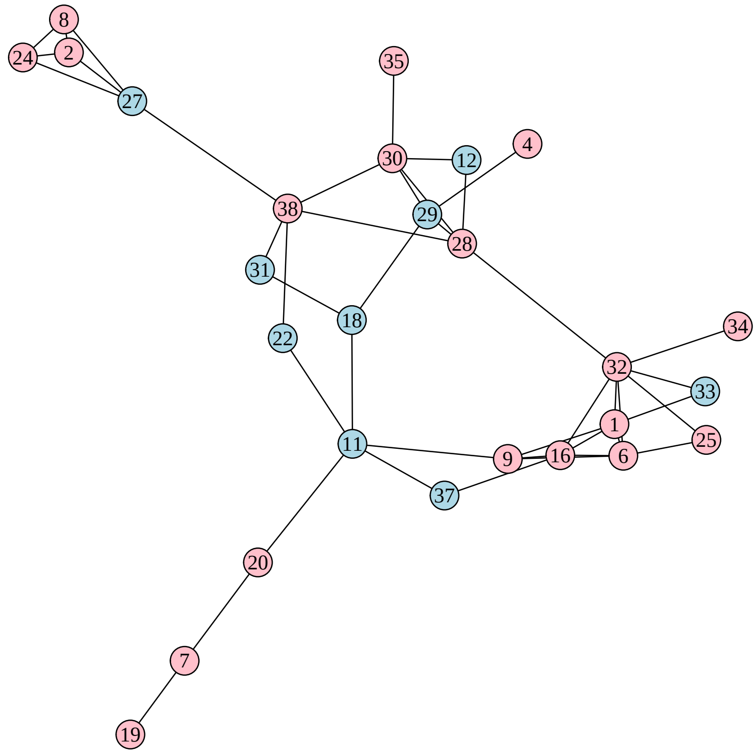


Figure A.21: Social network for class 3A of Liceu do Ceará School, afternoon shift.

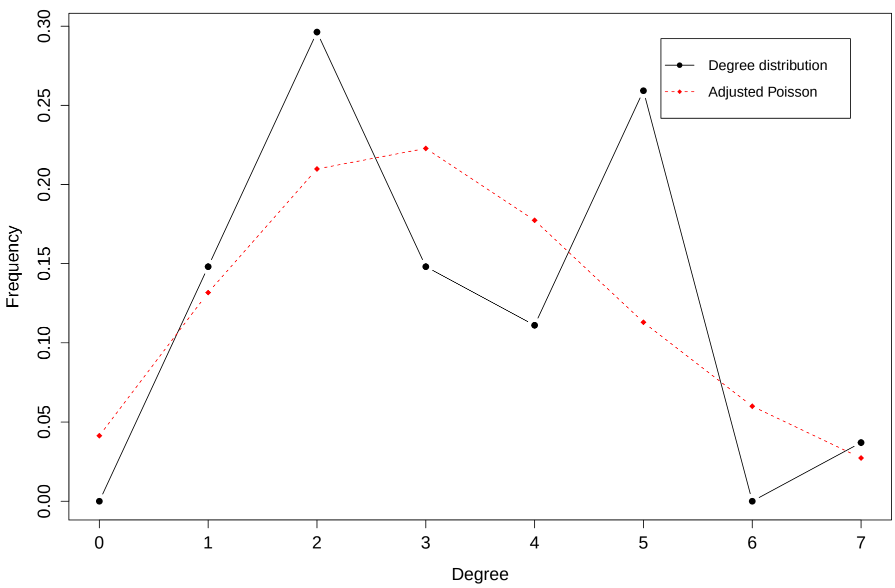


Figure A.22: Degree distribution for the network depicted on figure A.21.

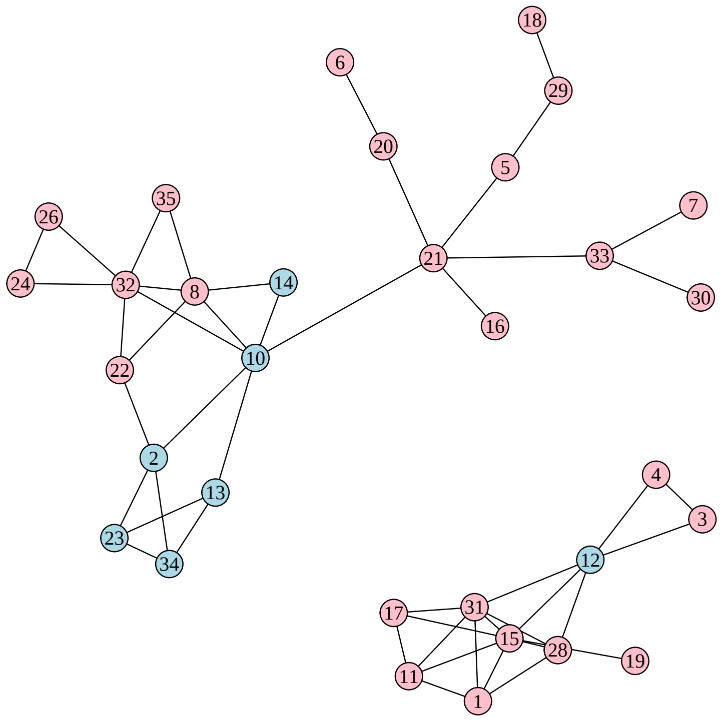


Figure A.23: Social network for class 3C of Dona Maria Amália Bezerra School, afternoon shift.

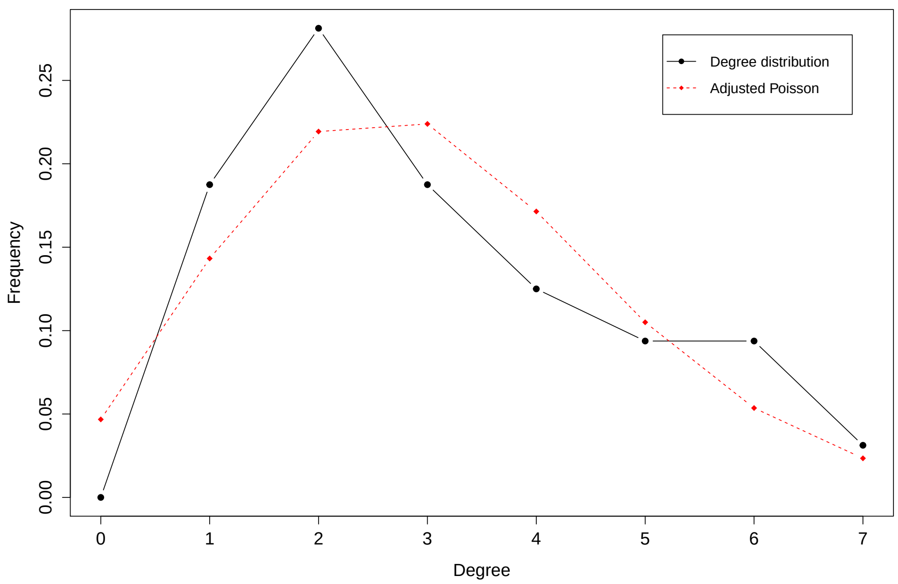


Figure A.24: Degree distribution for the network depicted on figure A.23.

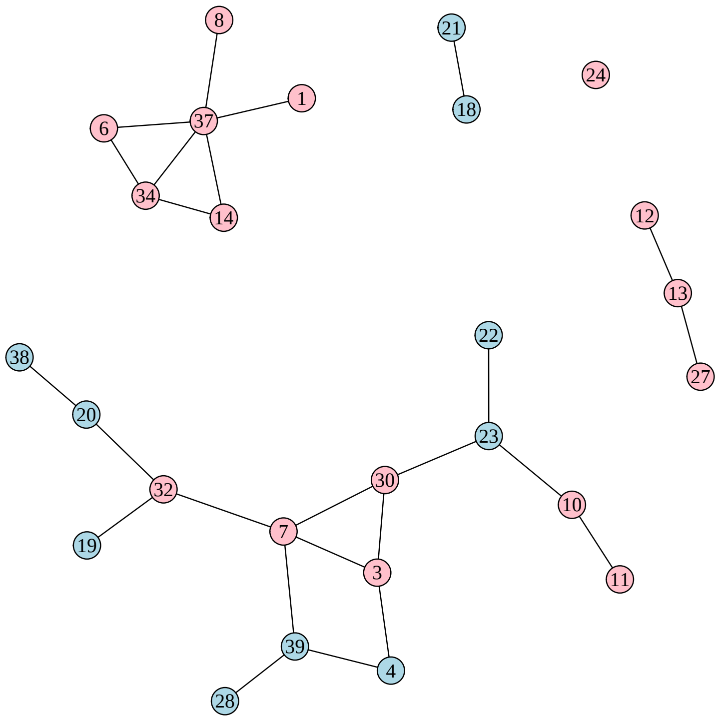


Figure A.25: Social network for class 3A of Dona Maria Amália Bezerra School, morning shift.

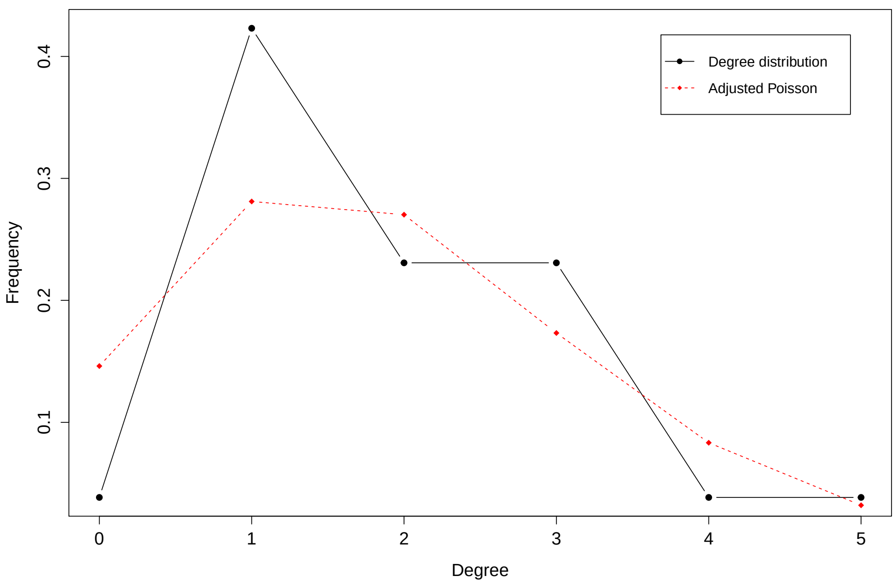


Figure A.26: Degree distribution for the network depicted on figure A.25.

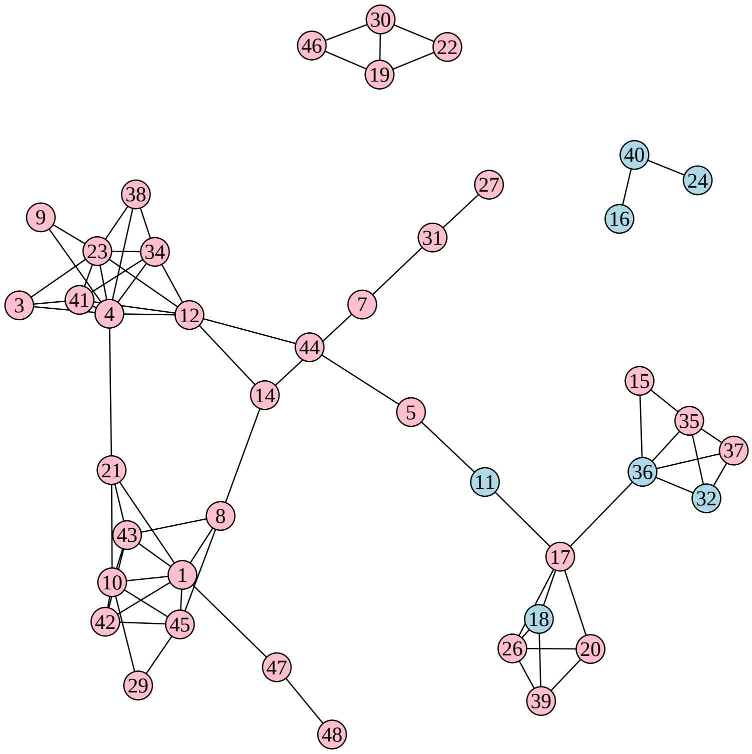


Figure A.27: Social network for class 3B of Liceu de Caucaia School, morning shift.

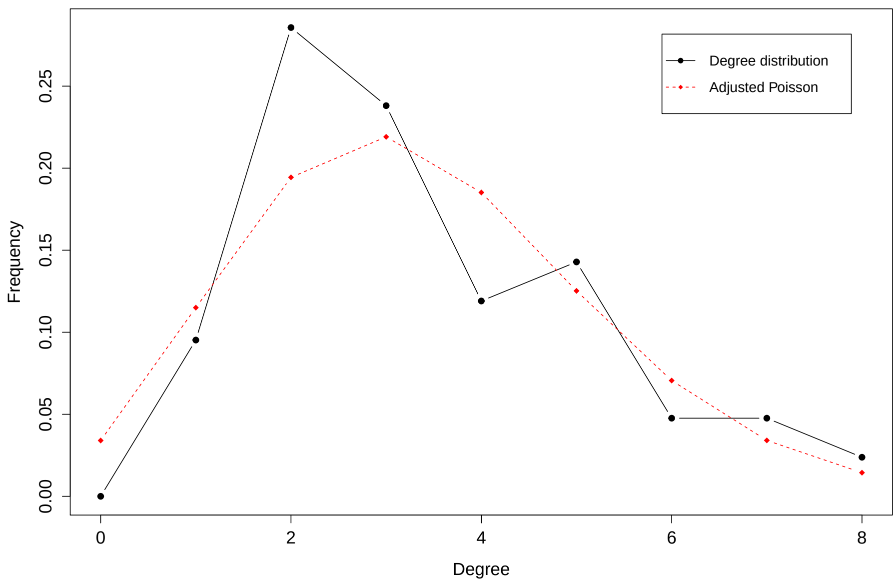


Figure A.28: Degree distribution for the network depicted on figure A.27.

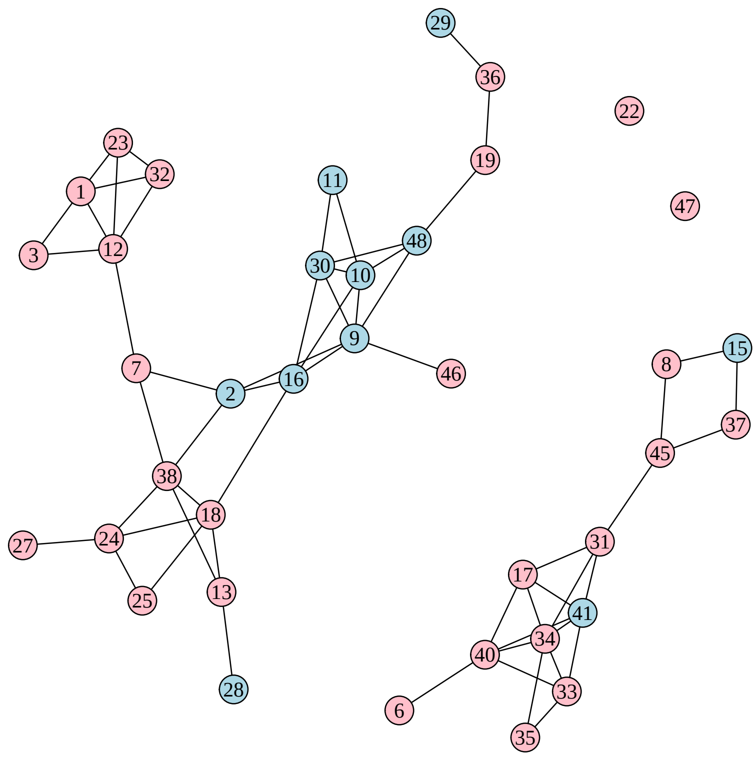


Figure A.29: Social network for class 3D of Liceu de Caucaia School, afternoon shift.

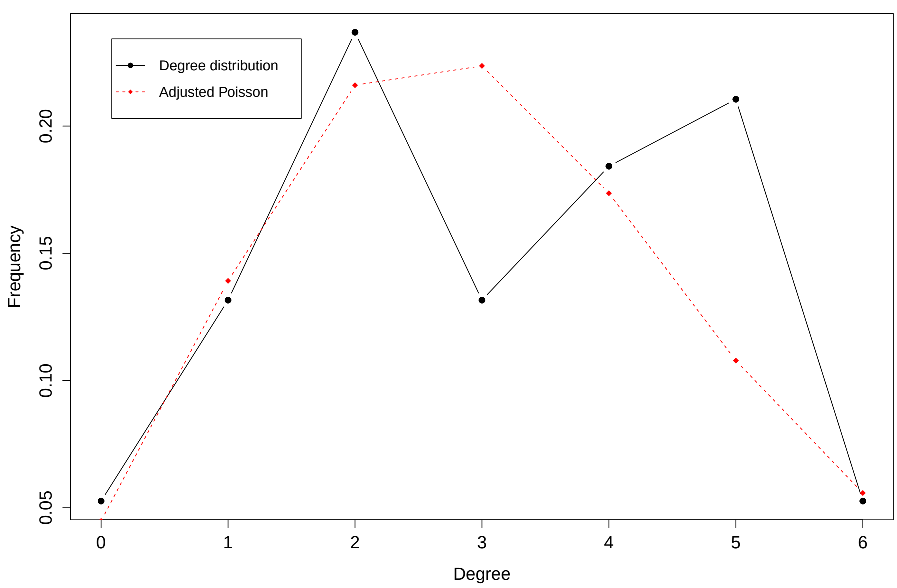


Figure A.30: Degree distribution for the network depicted on figure A.29.

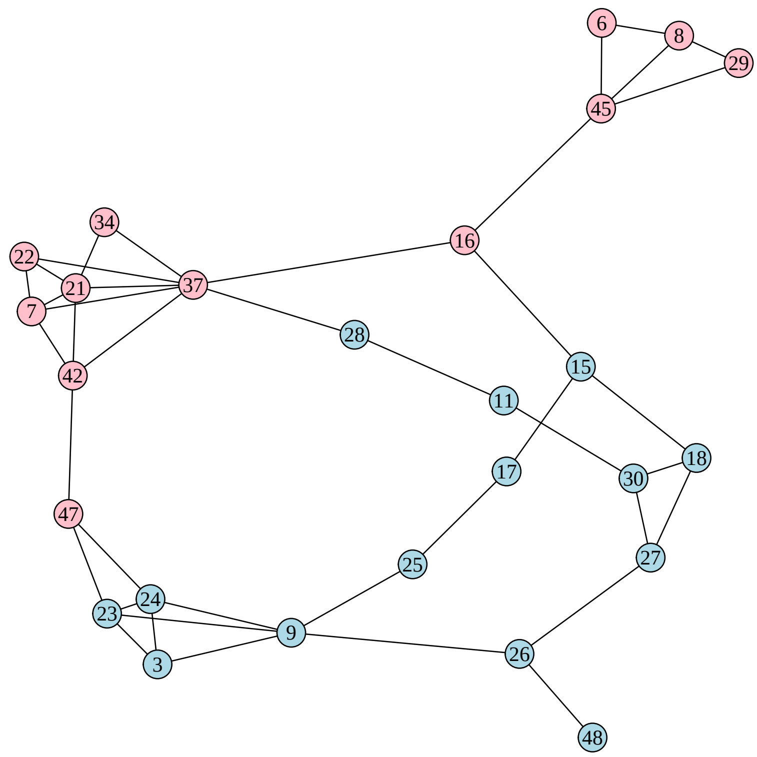


Figure A.31: Social network for class 3A of Liceu de Caucaia School, morning shift.

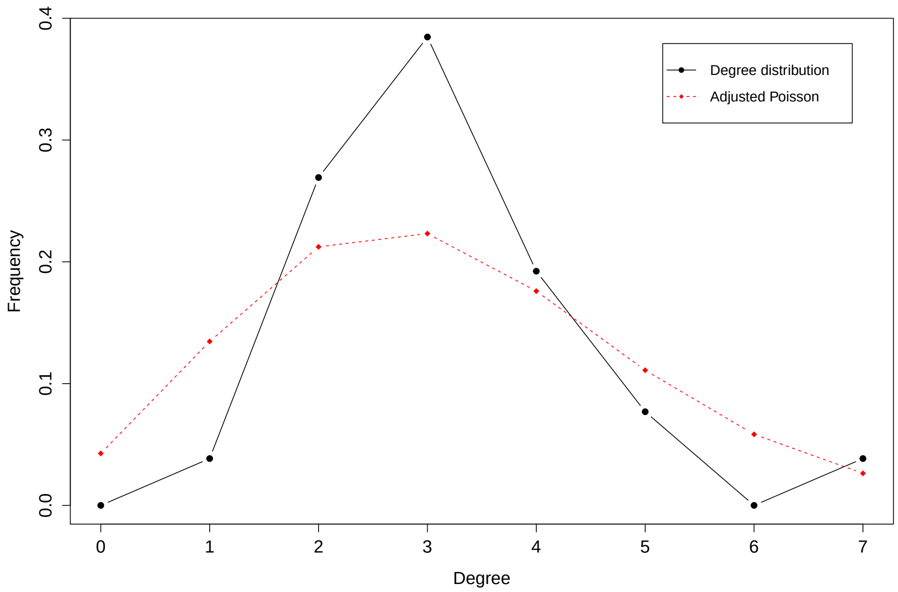


Figure A.32: Degree distribution for the network depicted on figure A.31.

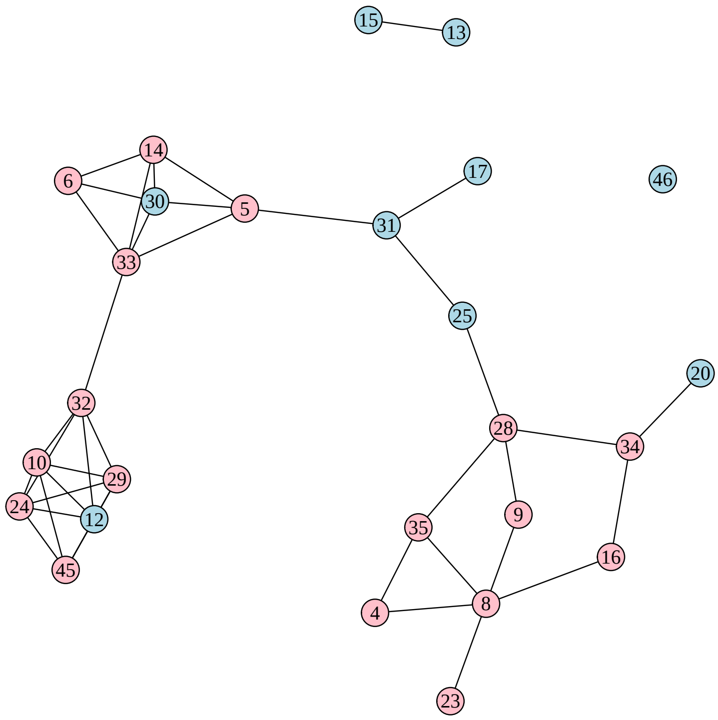


Figure A.33: Social network for class 3F of General Eudoro Correia School, afternoon shift.

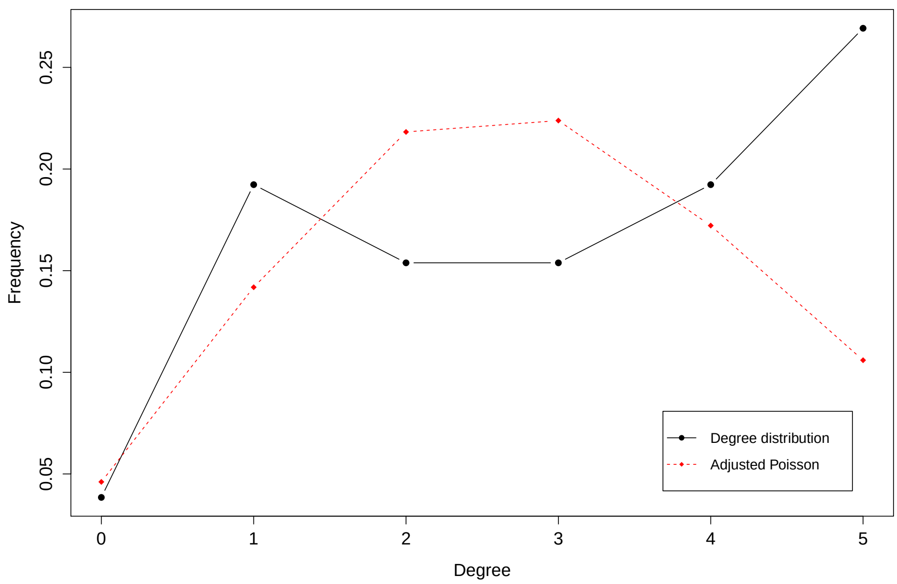


Figure A.34: Degree distribution for the network depicted on figure A.33.

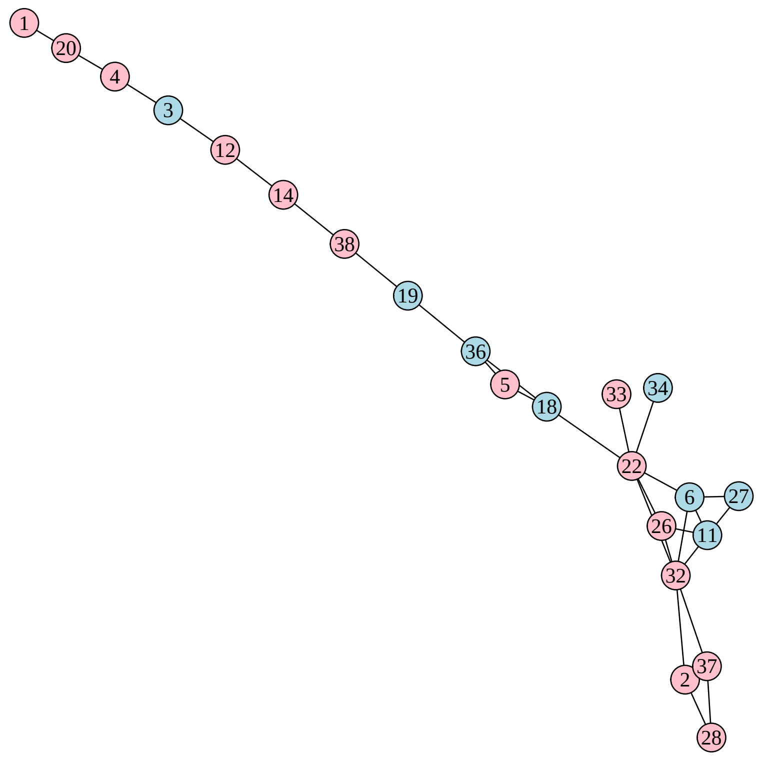


Figure A.35: Social network for class 3B of General Eudoro Correia School, morning shift.

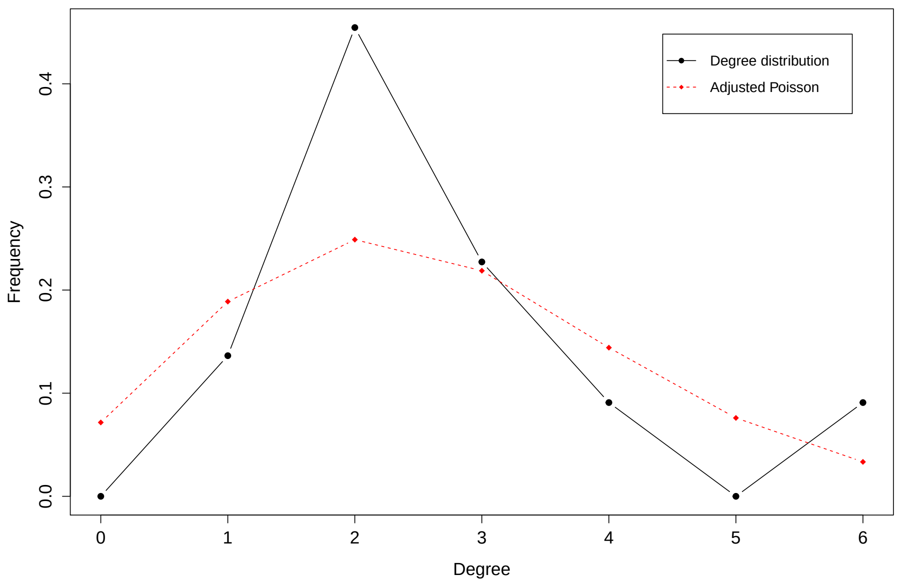


Figure A.36: Degree distribution for the network depicted on figure A.35.

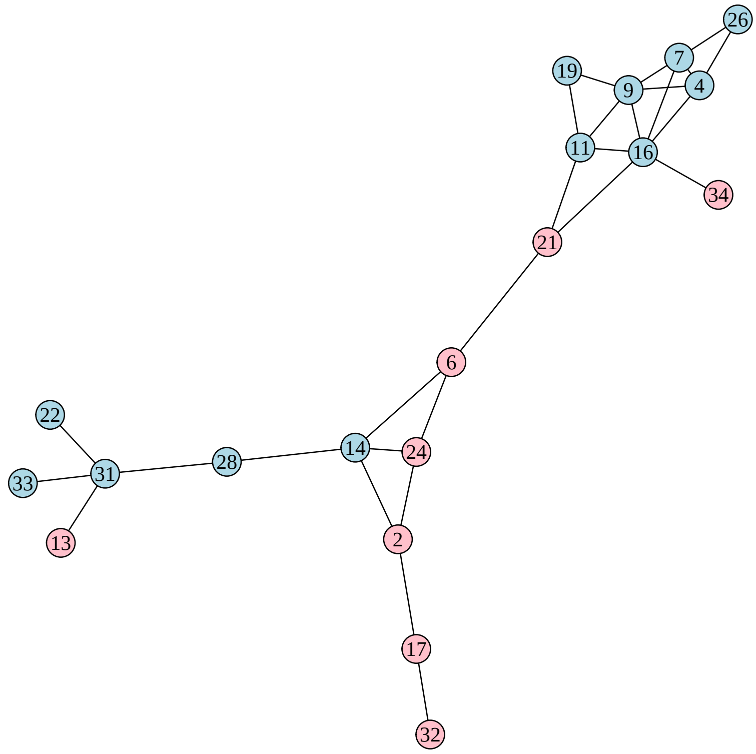


Figure A.37: Social network for class 3D of General Eudoro Correia School, morning shift.

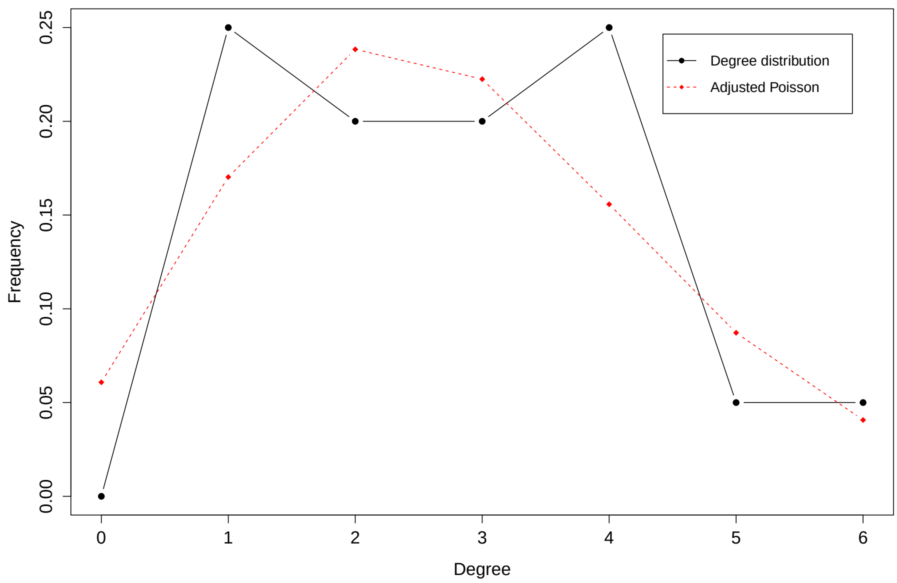


Figure A.38: Degree distribution for the network depicted on figure A.37.

DECLARAÇÃO

Atesto, para os devidos fins, que a dissertação de mestrado intitulada "Network Effects, Conformism and Misbehavior in Brazilian Classrooms", de autoria de Luan Falcão Daniel Santos, foi devidamente revisada. O material está em consonância com a gramática normativa da língua inglesa.

Fortaleza, 18 de Julho de 2016

Adriana Fontenele Pinheiro Padilha

Adriana Fontenele Pinheiro Padilha
Professora aposentada do Município de Fortaleza
Graduada em Letras Português/Inglês (licenciatura) com habilitação no ensino da língua portuguesa, inglesa e suas respectivas literaturas pela Universidade Estadual do Ceará - UECE