Creditworthiness using Social Capital: A Theoretical Framework

Roberto Almeida Shimizu

Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo

Master's in Computer Science

November 30th, 2023

Resumo

Confiança de Crédito Usando Capital Social: Uma Estrutura Teórica
Dissertação de Mestrado em Ciência da Computação
Autor: Roberto Almeida Shimizu
Professor orientador: Flavio Soares da Silva
Instituto de Matemática e Estatística, Universidade de São Paulo
30 de novembro de 2023

O capital social é crucial em comunidades pequenas, facilitando transações econômicas e arranjos de empréstimos solidários, como Associações Rotativas de Poupança e Crédito (ROSCAs) e empréstimos peer-to-peer. Estudos recentes, focalizando empréstimos peer-to-peer em fintechs como Prosper e Lending Club, indicam que empréstimos dentro de redes pessoais têm maiores taxas de adimplência do que os convencionais.

Os Sistemas de Confiança e Reputação, essenciais na ciência da computação, são utilizados para avaliar a confiabilidade em diversas redes, tais como em redes de comércio eletrônico e mais recentemente incluindo redes sociais. Esses sistemas operam com dados explícitos e implícitos para estabelecer medidas de confiança. A presente pesquisa analisou algoritmos voltados para a computação de confiança e reputação, representando o capital social em redes. Foram escolhidos algoritmos como eigentrust e tidaltrust para aplicação em estudos de caso reais de crédito solidário, incluindo um ROSCA e dois empréstimos ponto a ponto.

O estudo buscou entender a eficácia desses algoritmos e os desafios de coletar dados sobre confiança e reputação. Adicionalmente, investigou-se como esses algoritmos poderiam prever solidez de crédito em redes sociais de aldeias indianas, utilizando dados do estudo "A Difusão da Microfinança".

Este trabalho visa contribuir para o desenvolvimento de métodos de pontuação (escore) de crédito mais inclusivos, focados em populações menos privilegiadas. Através de uma abordagem interdisciplinar, o estudo destaca o potencial de utilizar capital social e mecanismos de confiança para fomentar a inclusão financeira e empoderamento econômico em comunidades pequenas.

Palavras-chave: Capital social, ROSCAs, Empréstimos peer-to-peer, Sistemas de Confiança e Reputação, Inclusão financeira, algoritmos em grafos.

Abstract

It is well known that in small communities, social capital plays a very important role in allowing economic transactions, especially in solidary lending arrangements, such as Rotating Savings and Credit Associations (ROSCAs), peer-to-peer loans among family and friends, and so on. More recently there were lots of academic papers analyzing the performance of peer-to-peer credit in fintechs such as Prosper, Lending Club, Kiva. These studies reveal an intriguing trend: loans funded within personal networks often exhibit higher repayment rates compared to conventional loans.

Nevertheless, there remains a significant gap in research, particularly in understanding the quantification of social capital within social networks and its subsequent application in assessing an individual's creditworthiness.

Trust and Reputation systems is an area of computer science which have attracted attention in various networking environments, including online social networks, wireless communication networks, multi-agent systems, and peer-to-peer networks. Trust and Reputation systems can use explicit or implicit information for decision making.

The research, embarked on a bibliographic analysis of algorithms specifically designed for computing trust and reputation, which act as proxies for social capital in social networks.

The algorithms selected for this study were eigentrust [37], tidaltrust [25] and graph algorithms [40]. They were applied in real cases of solidary credit arrangements observed in the community, more especifically a case of a ROSCA (Consorcio entre amigos) and two cases of peer-to-peer loans (Emprestimos entre amigos e familiares) that were derived from the former ROSCA. The objective was to evaluate the effectiveness and limitations of these algorithms and to address the challenges inherent in gathering trust and reputation data, considering its sensitive nature..

Furthermore, the research extends to examining how these algorithms can predict the likelihood of creditworthiness using real dataset of indian villages, database provided in the paper "The Diffusion of Microfinance" By Abhijit Banerjee, Arun G. Chandrasekhar, Esther Duflo and Matthew Jackson.

The goal of this research is to contribute to the development of more inclusive credit scoring methods, particularly targeting less-privileged segments of the population. By integrating interdisciplinary approaches, the study provides valuable insights into the potential of leveraging social capital and trust mechanisms to enhance financial inclusion and support economic empowerment in small communities.

Acknowledgements

This dissertation could not be accomplished without:

my wife Roberta for supporting me during this period.

my daughter's Tatiana and Catarina for inspiring me.

my professor Flavio Soares da Silva for the confidence.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Can social capital generated from social networks be applied as an alternative or complement to the traditional bank score methods?" From a computational point of view, this question can be rephrased as "What types of data structures and mathematical models in the literature take social capital as input and predict the creditworthiness of an individual?" For example, what if reputable members of someone's social network endorsed him or her about his or her creditworthiness based on feedback of past interactions, and after a series of computations we arrive with a probability of how much this person could be trusted in another financial transaction? Can we use "wisdom of the crowds" [62] to yield in a reasonable level of confidence a probability of creditworthiness of a person? In the absence of a minimum number of endorsements for a given person, what are other information about the topology of one's social network that could help to infer about his or her creditworthiness?

The theoretical motivation comes from a growing body of research in economics and computer science on whether Social Capital affects economics outcomes. This is particularly beneficial in micro-finance, where the targeted customers are mostly economically under-privileged people and are not well screened in traditional bank scoring methods, which basically assess payment history and size of asset collateral. In Brazil, it is estimated that 48% of population are unbanked, however, that does not mean that this segment does not use any finance solution in their daily lives. On the contrary, because of their vulnerable situation, they find numerous informal arrangements [50], most of them through their networks of family and friends. These networks have unique characteristics such as solidarity values and proximity of members that reduce information asymmetry and monitoring costs. In addition, the importance of trust and social reputation, works as social mechanism to reduce incentives to moral hazard between borrowers and lenders.

In some sociotechnical systems [2] such as e-commerce platform or product review websites, users can explicitly declare trust, normally as the result of positive interactions with the trustee. There are already a vast academic literature about reputation management methods and models to measure trust. However, when it comes to personal credit

and finance, obtaining social information and transforming them into a usable format is difficult. Another issue with such information is its reliability, which can be mitigated if there is credible verification and if the mechanisms designs in gathering the information assure rational behavior.

1.1 Problem Description: financial exclusion of poor population

The banking system have several mechanisms to reduce information asymmetry embedded in all and every financing operation.

Every credit transaction is an inter-temporal contract with an uncertain outcome. The exchange between the credit lender and the credit borrower does not rely solely on material collateral, but rather on the borrower's capacity to honour commitments. The economic basis of every credit organization is its ability to evaluate the situation and predict the borrower's behavior. At the heart of this exchange relationship, there is an assymetry of information between the credit borrower and the credit lender, which in turn causes two types of problems: adverse selection and moral hazard.

Adverse selection [18] refers to the tendency of those who seek to ask for credit to be a non-random selection from the population - more particularly, to be those who expect to have the highest expected credit delinquency. It also refers to the kind of *precontractual opportunism* that arises when one party to a bargain has private information about the something that affects the other's net benefit from the contract and when only those private information implies that contract will be especially disadvantageous for the other party agree to a contract.

Moral hazard [18] is a form of *postcontractual opportunism* that arises when actions required or desired under the contract are not frely observable and so the person taking them may chose to pursue his or her private interests at other's expense.

Therefore it is paramount in any credit operation a throughly and detailed analysis of the motives for the credit that will be used for, the individual history of the credit taker and if possible information about any asset collateral. The banking manager then relies on statistical techniques to determine a score (at which scale of credibility the individual credit proponent is placed), to decide whether he concedes the credit [50].

In addition to the cost that this operation involves, it is evident that a traditional credit organization (banks) tends to exclude population that lives in poverty, because they are unable to offer significant physical or financial collateral and a reliable record of banking transactions, which are a big problem to a population segment that just need liquidity the most.

However, this problem has been in part overcome by several organization capable to

lend small amounts of money to the poor people, in a way that would be incompatible with the high transaction costs of the traditional banks. These types of organization can offer to these borrowers' valid substitutes for individual collateral, and to the lenders low-cost alternatives to imperfect creditworthiness information. In fact, they provide credit to the poor based on "social collateral", through which borrowers' reputation, or the social networks to which they belong, take the place of the traditional physical or financial collateral. Since these arrangements build, to various degrees, on the extent and strength of personal relationships, they provide a fertile ground for the analysis of the role of social capital in the provision of credit.

The peer monitoring, group lending, or "solidarity group" are largely used approaches to credit delivery based on the assumption that the poor represent a much lower credit risk than the formal financial sector generally assumes and that, under specific circumstances, they can be trusted to repay small uncollateralized loans, using a lending methodology that relies on traditional and personal interactions among borrowers[50]. These principles of solidarity constitute a branch of microfinance.

The most visible and studied example of group-based lending is the Grameen Bank in Bangladesh[6]. The bank was started in 1976 by Mohammad Yunus, a professor at the University of Chittagong, as a research project. By 1994, the Bank had served half of all villages in Bangladesh, with a total membership of more than 2 million, of which 94% are women. Using the group lending methodology and transparent lending decisions, the Bank has consistently reported repayment rates in excess of 95%. The main innovation of this microfinance program stems from the observation that potential customers of these programs have a comparative information over the lender, which could be mobilized to develop mutually advantageous financial services. In fact, the constitution of groups is a necessary condition to the concession of credit, as these groups are responsible for selecting the members (thereby smoothing the hidden information problem), creating mechanisms that subordinates the concession of one credit to the payment of another - "Joint-Liability", and reduce delinquency rates through an effective invisible monitoring. These adapted forms of social pressure cause the solidarity groups to necessarily assume transaction costs and responsibilities, i. e., those which were now in great part performed by the group's members. Therefore, the cost of lending on average is significantly lower, allowing it to disburse larger amounts and to reach poorer people (through riskier loans) than those of traditional banks.

Self-selection of group members is a major element of this process. The Bolivian BancoSol program[6], which relies on groups of five to seven borrowers, includes groups which previously existed as ROSCAs, contributing to their high success rates as microfinance institutions.

The Banco do Nordeste (BNB) created the CrediAmigo[58] program in 1997 to provide microcredit services to informal entrepreneurs in Northeast Brazil. It is the best consoli-

dated program of its kind in Latin America and it uses extensively the above group-based lending principles (self-selection and "joint-liability") introduced by Grameen Bank. By the end of 2016, CrediAmigo had 2 million active clients who had received around R$ 2,8 billion (US$ 1.7 billion) in loans, with an average value per loan of R$ 1.900,00 (US$ 1,144.58) at interest rates of between 1.7% and 2% per month, depending on the type of product. The program has run for 20 years and now operates in around 2000 of Brazil's 5570 municipalities with some 500 service outlets ("units"). From the risk management point of view, it is important to note that the CrediAmigo default rate is relatively low (around 1.9% in 2016). In 2014 the default risk rate was better: a mere 0.6%, following several years when even lower average rates were recorded.

Karlan[38] experimented with FINCA-Peru, a group lending organization in Peru that provides loans with joint-liability). It has found evidence that individuals with a stronger social connection to their fellow group members (i.e., either living closer or being of a similar culture) have higher repayment rates and higher savings.

Examples of such organization are the rotating credit and savings associations (ROSCAs)[32], the local moneylenders (usurers), trade credit and the group-based microfinance programs (Grameen Bank) [6].

Rotating Savings and Credit Associations (ROSCAs) [6] are a "response by a socially connected group to credit market exclusion, and a widespread way to crystallize social relations in an informal – yet often formally run – system of internal credit delivery [7]. The concept of ROSCA is known as "chit funds" in India, "kye" in Korea, "partners" in Jamaica, "susu" in Ghana, "njangeh" in Cameroon, "tontine" in Senegal, "pasanaku" in Bolivia, "panderos" in Peru, "Tanda" in Mexico, "caixinhas e consórcios" in Brazil, and in many other countries. Izumida [34] describes the Tanomoshi-Kou system, introduced in Japan in the 12th or the 13th century as the oldest form of ROSCA.

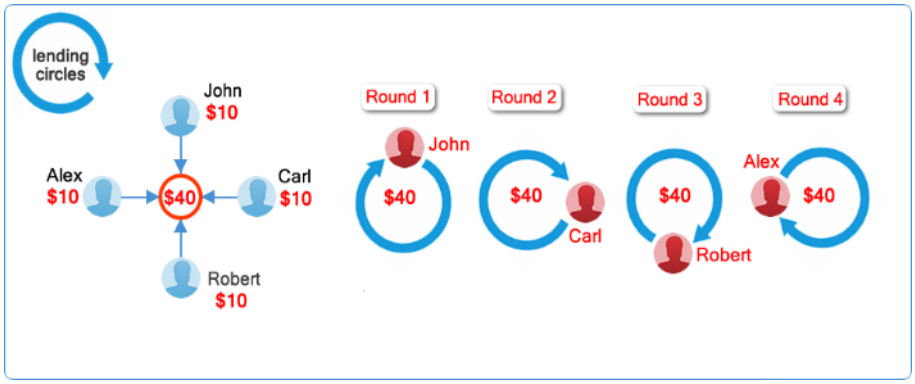


Figure 1.1: Dynamics of a Rotating Savings and Credit Association (ROSCA)

A ROSCA [32], depicted in Figure 1.1, is a group of men and/or women who contribute to a collective fund which is at regular intervals distributed – either randomly, or by auction, or by collective decision – to one of the group's members. In effect, all the

members of the group (except the last person in the rotation) receive an advance that they repay through their contribution to the fund for the duration of the cycle: the earlier and individual's position in the rotation, the larger the credit he/she receives, and the lower the risk he/she faces (if a person fails to contribute after he/she receives the fund, only those members after him/her will be adversely affected). At the end of a cycle, i.e., when each member has received the fund once, the ROSCA is dismantled or, more often, reconstituted with the same or similar membership. This creates the possibility that well-performing member can move up in the rotation, providing an incentive for good payment record that spans several rotations, and – if new entrants are chosen carefully – potentially reinforces the efficiency of the association.

ROSCAs work well as long as individuals value the benefits of membership in the association more than the benefits of defaulting, as a result, all members contribute to the fund even after they have received the total group collection. Therefore, the main defining characteristic of a performing association lies in the reduction of the risk of opportunistic behavior (moral hazard). The costs of default include social mechanisms that extend beyond the domain of the ROSCA into community wide sanctions such as peer pressure and social ostracism, which affect every aspect of that individual's social and economic life.

Although the cases above show strong evidences of the impact of social relations in success of this credit arrangements, it is still very poor the research of how the trust sentiment embedded in the social relations of the members of such above example of credit arrangements can be captured in a computational data structure. Then, how it can be used to compute scores of trust, which could be used for other applications, for example recommendations or complementary information for credit in banks.

Chapter 2

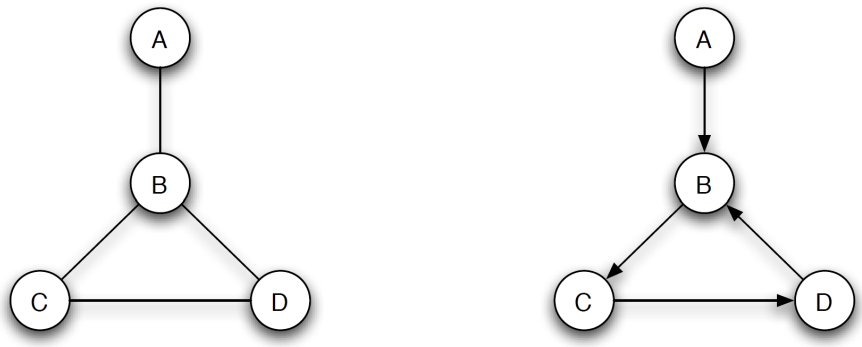
Literature Review

2.1 Social Capital and Economic Outcomes

In order to discuss social capital in a social network, it is necessary to congregate ideas coming from computer science, economics and sociology. These ideas help to reason about the network structure, strategic behavior of the participants and the social influence they produce in others.

2.1.1 Social Network Structure and Social Capital

For analyzing network structure, we rely heavily on *graph* theory. A *graph* is a way of specifying relationships among a collection of items. A *graph* consist of a set of objects, called *nodes*, with certain pairs of these objects connected by links, called *edges*. They are useful because they serve as mathematical models of network structures. In the context of social networks, *nodes* are people or groups of people, and *edges* represent some kind of social interaction. In many social contexts however, we want to express asymetric relationships, and in this case we define a *directed edge* as an edge that links one node to the other, with the direction being important. Graphically we represent these edges as arrows. The category of graphs that have direct edges are called *directed graphs or digraphs*, as opposed to *undirected graphs*, in where the relationships are reciprocal. Figure 2.1 exhibits examples of *undirected and directed graphs*.



(a) *A graph on 4 nodes.* (b) *A directed graph on 4 nodes.*

Figure 2.1: Two graphs: (a) an undirected graph, and (b) a directed graph

Figure 2.2 shows a social network in the form of a graph. Observing its visual appearance we can already see some of the complexity that social network structures contain. It is generally difficult to summarize the whole network succinctly; there are parts that are more or less densely interconnected, sometimes with central cores containing most of the links, and sometimes with natural splits into multiple tightly-linked regions. Participants in the network can be more central or more peripheral; they can straddle the boundaries of different tightly-linked regions or sit squarely in the middle of one [17].

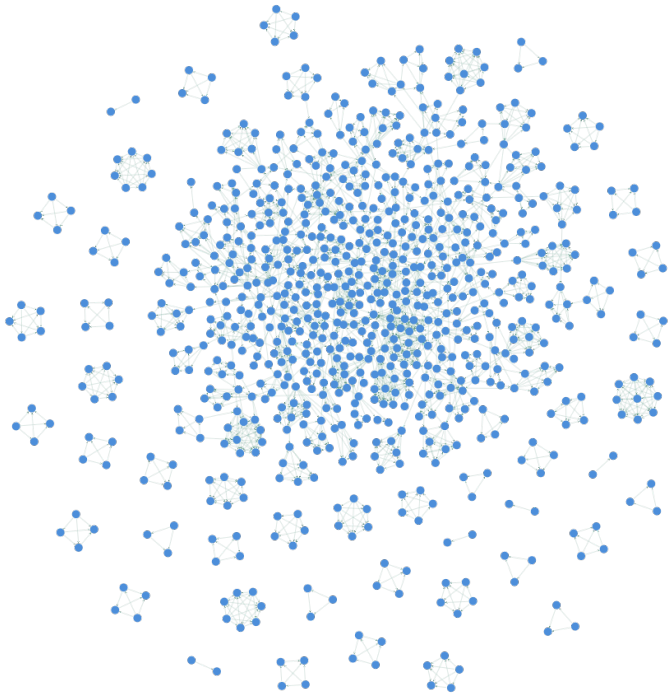


Figure 2.2: Graph of social network of village in India

The definition of a *path* is a sequence of nodes with the property that each consecutive pair in sequence is connected by an *edge*. We say that a *graph* is *connected* if for every

pair of nodes, there is a *path* between them. Moreover, we normally call a *connected component* of a *graph* a subset of nodes such that every node in the subset has a path to each other and the subset is not part of a larger set. In large and complex social networks, its often observed the existence of a *giant component*, a deliberately informal term for a connected component that contains a significant fraction of all the nodes

In addition to simply asking whether two nodes are connected by a *path*, it is also interesting in most settings to ask how long such a *path* is. We define the length of a *path* to be the number of steps it contains from beginning to end, in other words, the number of edges in the sequence that comprises it. There could be more than one path between two nodes, with different lengths. However, we define the *distance* between two nodes to be length of the *shortest path* between them. The most efficient way to calculate the *shortest path* between two nodes is by using the *breadth-first search* algorithm, which uses recursion to keep discovering nodes layer-by-layer from the source node, building each new layer from the nodes that are connected to one node away in the previous layer, until reaching the destination node. [17].

The Small-World Phenomena

In addition of the existence of a giant component in very large social networks, in which we have paths connecting a given person to all other people, these path lengths were found to be surprisingly small. The most famous experiment was conducted by Stanley Milgram in the 1960's [41]. In the experiment, Milgram decided to investigate the so-called *small-world problem*, the hypothesis that everyone on the planet is connected by just a few intermediaries. In his experiment, a few hundred people from Omaha tried to get a letter to a target—a complete stranger in Boston. But they could only send the letter to a personal friend whom they thought was somehow closer to the target than they were. When Milgram looked at the distribution of how much intermediaries took the letters to reach the target, depicted in Figure 2.3, he found that on average the letter changed hands only about six times. This finding has revealed the notion that everyone can be connected by a chain of acquaintances roughly six links long. Since then, several other experiments were repeated, including more recently using very large online social networks, and all of them have six-degrees as a ballpark [43]. The existence of all these short paths has substantial consequences for the potential speed with which information can spread through society, as well as for the potential access that the social network provides to people with very different characteristics from one's own [17].

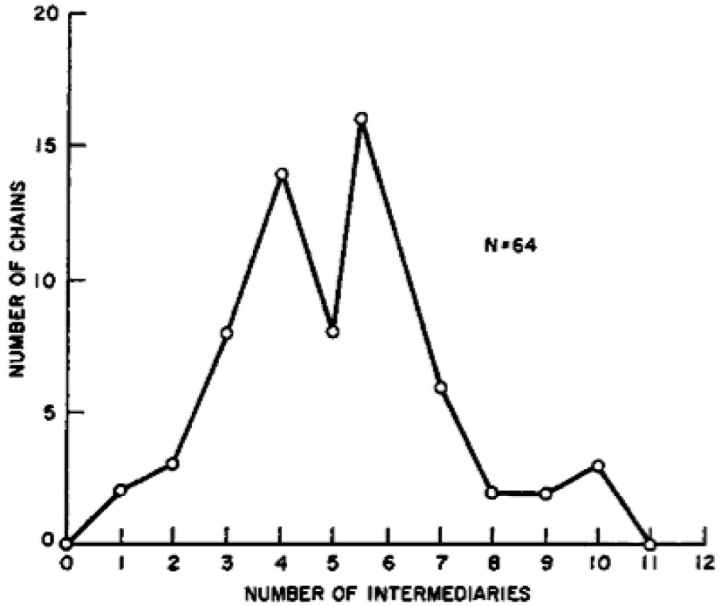


Figure 2.3: A histogram from Milgram's paper on their small-world experiment

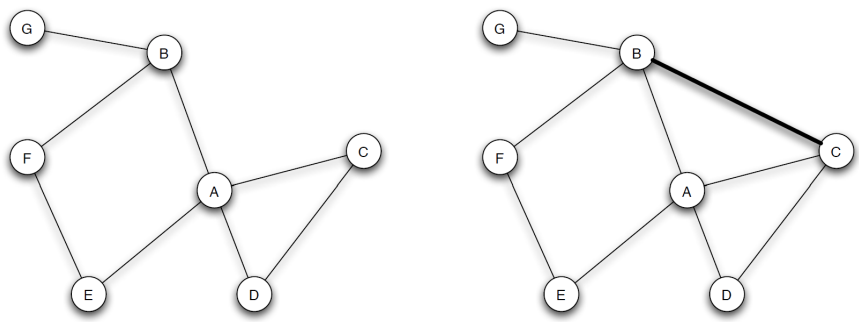
In a social network, different nodes can play structurally distinct roles in dissemination of information. One of the most famous work is the research Mark Granovetter, also in the 1960s. In his research [29], he interviewed people who had recently changed employers to learn how they discovered their new jobs. He found that many people learned information leading to their current jobs through personal contacts. But perhaps more strikingly, these personal contacts were often described by interview subjects as *acquaintances* rather than close friends.

In order to better understand why, it is important to grasp some concepts such as *triadic closure*, *clustering coefficient*, *bridges and local bridges* and *the strength of weak ties.*

Triadic closure principle

If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future.

Consider the network in Figure 2.4, if nodes B and C have a friend A in common, then the formation of edge BC closes a triangle in which all three nodes A, B and C will have edges connecting each other.



(a) *Before B-C edge forms.* (b) *After B-C edge forms.*

Figure 2.4: The formation of the edge between B and C illustrates the effects of triadic closure, since they have a common neighbor A

Possible reasons for triadic closure: One reason why B and C are likely to become friends, when they have a common friend A, is simply based on the opportunity for B to meet C. Second, is the fact each of B and C is friends with A, gives them basis for trusting each other that an arbitrary pair of unconnected people may lack. And a third one is based on incentive A may have to bring B and C together.

Clustering coefficient

The basic role of triadic closure in social networks has motivated the formulation of simple social network measures to capture its prevalence. One of these is the clustering coefficient [45, 60], calculated as:

$$C_i = \frac{2e_i}{k_i(k_i - 1)} \tag{2.1}$$

where e_i is number of connections between friends of i and k_i is the number of friends of i.

In Figure 2.5, in the first case, all the 4 friends of i are connected within themselves, totalizing 6 connections, then applying the formula it yields 1. In the second case, there are 3 connections among the 4 friends of i, then applying the formula it yields 0.5. And finally, in the third case, there are no connections among the 4 friends of i, then applying the formula it yields 0. Another way to think about clustering coefficient is that it yields the fraction of pairs of node $i's$ friends that are connected with each other.

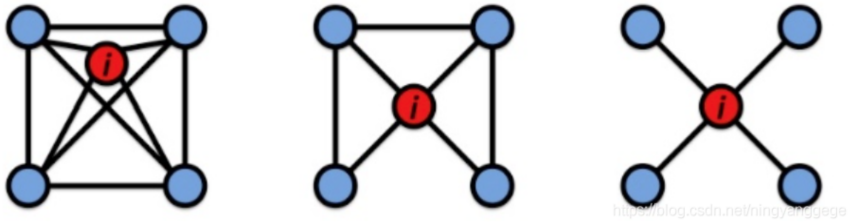


Figure 2.5: Three cases of network structure that yield different clustering coefficient for node i

The clustering coefficient of a node i ranges from 0 (when none of the node i's friends are friends with each other) to 1 (when all of the node i's friends are friends with each other), and the more strongly triadic closure is operating in the neighborhood of the node i, the higher the clustering coeffcient will tend to be.

Local Brigdes

In Figure 2.6, the person A has three friends (C, D, E) forming a closed knit-group, in where they all know each other. In addition, person A has another friend B, who belongs to a different group of people. We say that an edge joining two nodes A and B in a graph is a *local bridge* if deleting the edge would cause A and B to lie in two different components. In other words, this edge is literally the only route between its endpoints: the nodes A and B.

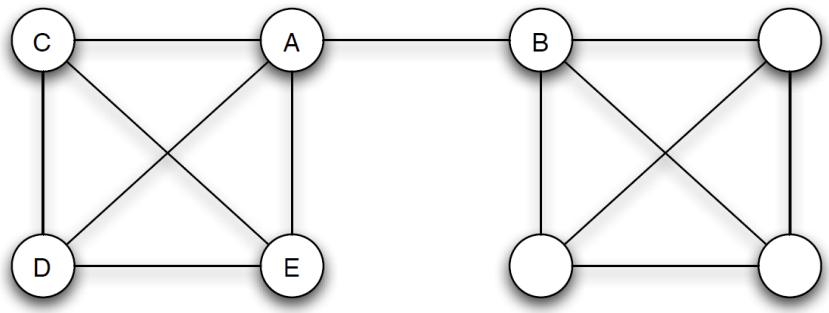


Figure 2.6: The A-B edge is a local bridge, meaning that its removal would place A and B in distinct connected components.

Local bridges provide their endpoints with access to other parts of the network, and therefore to new sources of information, that they would otherwise be far away from. This is the context to interpret Granovetter's observation about job-seeking. We might expect that if a node like A is going to get new information, the kind that leads to a job opportunity, it might come from a friend like B, connected by a *bridge*. The closed

knit-group that A belongs, although having people eager to help, often have the same information that A has.

Strength of Weak Ties

In order to understand fully Granovetter's observation, we need to be able to distinguish the relative strength of the relationship: a notion of *strong ties*, links that represent closer friendship and greater frequency of interaction, and *weak ties* as links that correspond to less closer relationship, for example *an acquaintance*. Granovetter then suggested an relation between strong and weak ties and triadic closure, that may be called *strong triadic closure principle*. This principle assumes:

If a node A has edges to nodes B and C, then B-C edge is especially likely to form if A's edges to B and C are ***strong*** *ties.*

The consequence of it entails the following claim: *If a node A in a network satisfies the strong triadic closure principle and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie.*

In other words, assuming the *strong triadic closure principle* and a sufficient number of strong ties, the local bridges in a network are necessarily weak ties.

Figure 2.7 shows a network in where all nodes satisfy the *strong triadic closure principle.*

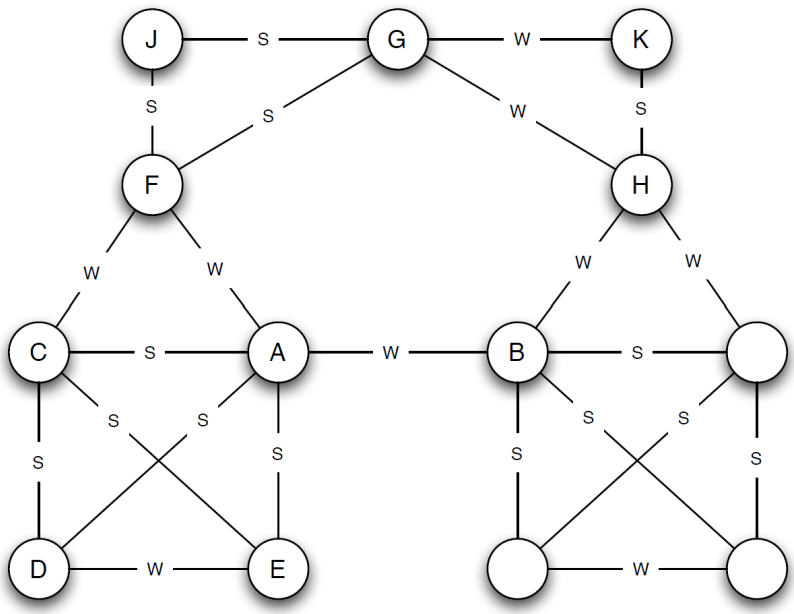


Figure 2.7: Each edge of the social network is labeled here as either a strong tie (S) or a weak tie (W), to indicate the strength of the relationship. The labeling in the figure satisfies the strong triadic closure principle

This concludes a series of arguments that support Granovetter's findings that corre-

lates tie strength with local bridges. It give us a framework to think about the *strength of weak ties* (here considered as acquaintances), in enabling connections to people from other parts of the network, and therefore enabling transmission of new information in between.

Effect of node position

In social networks, access to edges that span different groups is not equally distributed across all nodes: some nodes are positioned at the interface between multiple groups, with access to boundary-spanning edges, while others are positioned in the middle of a single group. What is the effect of this heterogenity? We can formulate an answer to this question as a story about the different experiences that nodes have in a network like the one in Figure 2.8 particularly in the contrast between the experience of a node such as A, who sits at the center of a single tightly-knit group, and node B, who sits at the interface between several groups.

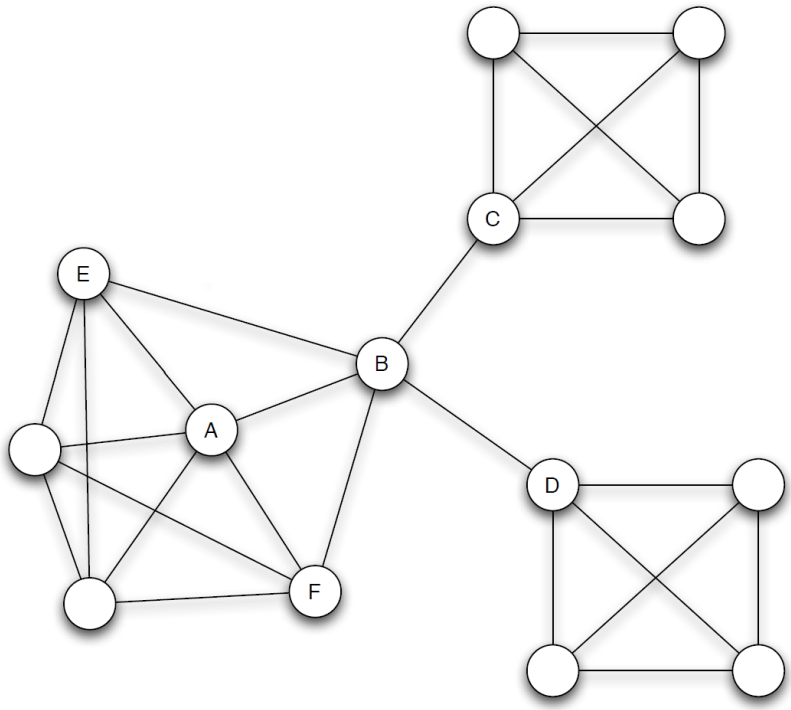


Figure 2.8: The contrast between densely-knit groups and boundary-spanning links is reflected in the different positions of nodes A and B in the underyling social network.

Embeddeness

In Figure 2.8, node A's set of network neighbors has been subject to considerable triadic closure; A has a high clustering coefficient. We define the *embeddedness* of an edge in a network to be the number of common neighbors the two endpoints have. Thus, for

example, the A-B edge has an *embeddedness* of two, since A and B have the two common neighbors E and F. So what stands out about A is the way in which all of his edges have significant *embeddedness*. We also observe that local bridges, such as B-C, are precisely the edges that have an *embeddedness* of zero since they were defined as those edges whose endpoints have no neighbors in common.

A long line of research in sociology has argued that if two individuals are connected by an *embedded edge*, then this makes it easier for them to trust one another, and to have confidence in the integrity of the transactions (social, economic, or otherwise) that take place between them [14, 13, 28]. Indeed, the presence of mutual friends puts the interactions between two people "on display" in a social sense, even when they are carried out in private; in the event of misbehavior by one of the two parties to the interaction, there is the potential for social sanctions and reputational consequences from their mutual friends [17].

No similar kind of deterring threat exists for edges with zero embeddedness, since there is no one who knows both people involved in the interaction. In this respect, the interactions that B has with C and D are much riskier than the embedded interactions that A experiences. Moreover, the constraints on B's behavior are made complicated by the fact that he or she is subject to potentially contradictory norms and expectations from the different groups she associates with [15].

In contrast, a line of research in sociology, catalyzed by influential work of Burt [10], has argued that network positions such as node B's, at the ends of multiple local bridges, confer a distinct set of equally fundamental advantages. First, it is an informational one: B has early access to information originating in multiple, non-interacting parts of the network. A second, related kind of advantage is based on the way in which standing at one end of a local bridge can be an amplifier for creativity [9]. And third, B's position in the network provides an opportunity for a kind of social "gatekeeping", he or she regulates the access of both C and D to the tightly-knit group she belongs to, and she controls the ways in which her own group learns about information coming from C's and D's groups. This provides B with a source of power in the organization, and one could imagine that certain people in this situation might try to prevent triangles from forming around the local bridges they're part of for example, another edge from C or D into B's group would reduce B's gatekeeping role.

All of these arguments are framed in terms of individuals and groups deriving benefits from an underlying social structure or social network; as such, they are naturally related to the notion of *social capital*.

Social Capital

In Alejandro Portes's review of the topic, he writes [46], "Consensus is growing in the literature that *social capital* stands for the ability of actors to secure benefits by virtue of

membership in social networks or other social structures".

Another concept of *social capital* is associated with Putnam [57, 48] who views it as a set of "horizontal associations" between people: *social capital* consists of social networks ('networks of civic engagement") and associated norms that have an effect on the productivity of the community. The key feature of *social capital* in this definition is that it facilitates coordination and cooperation for the mutual benefit of the members of the association.

Social capital can be constructed either as an individual attribute that captures the embedded resources within the individual's social relations to generate a personal economic benefit[30], or as a group attribute that enhances group member's transactional efficacy for economic gain[12].

The *social capital* definition from Granovetter[27] says economic actions of individuals are determined by how the social relations between involved agents are. These social and economic relations are very much immersed in social networks, grounded by reciprocal trust either for the purpose of enabling economic transactions or more broad social relations.

Ferrary[20] built on this definition, **highlighting the roles of social networks and trust**. Social networks, due to obligations that bond some of its members and the nature of information that they exchange, alter the economic regulation. In this context, one can define social network as a group of individuals in which the frequency of economic interactions and the density of social relations enable to reduce the uncertainty related to moral hazard, enabling discriminate precisely honest members from dishonest ones. According to Ferrary[20], there are 3 conditions of existence and strengthening of trust in an economic relation within a social network: proximity, temporality and personal nature of the relations:

- proximity primarily geographic, moreover recently communication advances are enabling virtual proximity. The key is that proximity favors a frequency and quality of the interaction between agents.

- temporality means interaction in a logic of repeated games with same players, which reduce the cost of information and enable a mutual learning between the agents. Therefore, any member who has a longstanding relation in a social network tends to be seen more trustworthy than new ones.

- personal nature of relation, this is a subtle concept, members of a social network are not necessarily only motivated to do economic relations with another member due to the economic reasons, but also due to social and psychological reasons. And sometimes these decisions have a more symbolic meaning that justifies it and strengthens the relationship rather than anything else.

Glaeser et al.[19] made a couple of experiments using "Trust Games" with Harvard undergraduates to try to measure *social capital* as an individual attribute. In these experiments they examined trust and trustworthiness relating the outcomes of the trust games with attitudinal questions (General Social Survey – GSS) and background characteristics. They found that i) measures of past trusting behavior are better than the abstract attitudinal questions in predicting subjects' experimental choices in the trust games; ii) although questions about trusting attitudes do not predict trusting behavior, such questions do appear to predict trustworthiness; iii) the degree of social connection between the sender and recipient (the number of friends they have in common and the duration of their acquaintanceship) generally predicts the levels of trust and trustworthiness in the two-person trust game; iv) subjects who are paired with a partner of difference race or nationality send back less money to their partner; v) background characteristics capturing the level of status and organization membership (variables meant to serve as proxy for an individual's own social capital) strongly predict the amount of money that sender's receive back from recipients.

Social capital, given its informational and solidarity benefits, can ameliorate potential inefficiencies caused by information imperfection and to enhance market efficiency. A core proposition of *social capital* theory is that valuable resources are embedded within networks of relationship, providing collectively owned capital for their members to use as a credit[62].

Borgatti, Jones, and Everett [8], summarizing discussions within the sociology community, observe two important sources of variation in the use of the term *social capital*. First, *social capital* is sometimes viewed as a property of a group, with some groups functioning more effectively than others because of favorable properties of their social structures or networks. Alternately, it has also been considered as a property of an individual; used in this sense, a person can have more or less *social capital* depending on his or her position in the underlying social structure or network. A second, related, source of terminological variation is based on whether *social capital* is a property that is purely intrinsic to a group - based only on the social interactions among the group's members - or whether it is also based on the interactions of the group with the outside world.

Closure and Bridging as Forms of Social Capital

A view at this level of generality does not yet specify what kinds of network structures are the most effective for creating *social capital*.

The writings of Coleman [13] and others on *social capital* emphasize the benefits of *triadic closure* and *embedded edges* for the reasons discussed above: they enable the enforcement of norms and reputational effects, and hence can help protect the integrity of social and economic transactions.

Burt, on the other hand, discusses *social capital* as a tension between *triadic closure*

and *brokerage*, with the former referring to Coleman's conception and the latter referring to benefits arising from the ability to "broker" interactions at the interface between different groups, across structural holes.

The contrasts are also related to Robert Putnam's dichotomy between bonding capital and bridging capital [344]; these terms, while intended informally, correspond roughly to the kinds of *social capital* arising espectively from connections within a tightly-knit group and from connections between such groups.

The notion of *social capital* thus provides a framework for thinking about social structures as facilitators of effective action by individuals and groups, and a way of focusing discussions of the different kinds of benefits conferred by different structures. Networks are at the heart of such discussions — both in the way they produce closed groups where transactions can be trusted, and in the way they link different groups and thereby enable the fusion of different sources of information residing in these groups.

2.1.2 Organizations leveraging Social Capital

More recently platforms such as Facebook and LinkedIn have changed the landscape by greatly facilitating the creation and maintenance of many social relations and making them highly visible, more and more online platforms are seeking to leverage these social relations for economic activities such as lending, car sharing and rentals[39]. As individuals connected by powerful social networking tools transact with each other, it is inevitable that economic decisions are embedded in social relations. Such is the case with online peer-to-peer (P2P) lending where individual lenders collectively bid on loan request by individual borrowers in an online platform supported by social networking tools.

There are several peer-to-peer lending platforms worldwide, such as Prosper, Lending Club, Zopa, Funding Circle, Kiva, Lenddo and PPiDai. As with crowdfunding platforms, P2P leverages the "wisdom of the crowd"[21, 62] by allowing multiple lenders to collectively fund a loan. P2P lending also provides online social networking functions so that lenders and borrowers can declare friendship with one another[62]. These friendships include both existing offline social relations and newly formed online groups. P2P lending platforms provide tools for members to formally recognize these social relations, together with benefits such as the ability to broadcast loan request to friends, and to receive notifications of friends' borrowing and lending activities. The ability to leverage friendship networks in borrowing/lending activities is a key difference between online P2P lending and traditional lenders such as banks.

Table 2.1 shows some academic papers that analyze funding and lending outcomes of P2P lending platforms using data-driven classifiers such as logistic regression. All the results are consistent with the fact that social ties in the context of P2P lending improve funding and repayment outcomes.

Fredman and Jim[21] found that borrowers' friendship networks were consistently significant predictors of lending outcomes. Lin et al.[42] found that the number of friends that a borrower has and the number of friends that actually bid on a loan increase the probability of successful funding of a loan and reduce interest rates and ex post default rates.

On the other hand, Chen et al.[11] showed that social connections from group networks (groups formed online, with no previous connection offline) improves funding performance but has no impact on repayment performance.

Sukamaningsih[54] made a throughout assessment of the academic papers in the topic of trust in peer-to-peer lending and observed that P2P lending still lack risk assessment methodologies. His paper proposed a model of trust between lender and borrower based on Elaboration likelihood Model (ELM) and literature review. This model, depicted in figure 2, categorizes trust factor for borrowers into three categories: hard information as part of central route and soft information and social capital as parts of peripheral route. Moreno et al.[3] also performed a literature review and identified 10 success factors in peer-to-peer lending, dividing them into three groups: borrower factors, platform information and lender factors. Among them, trust is one of the key factors, which is not considered in tradition banking.

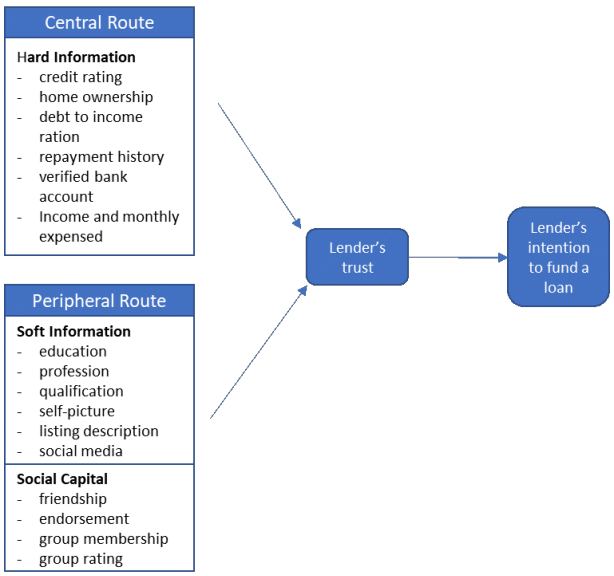


Figure 2.9: Proposed Trust Model for peer-to-peer lending

Bao, Ding and Mohlmann [5] pointed weaknesses of interdependencies between organizational components and platform design in modern digital P2P platforms such as LendingClub and Upstart (US), Renrendai (China), and Zopa (UK). These plataforms rely too much in the characteristics of the loan itself, not much different from than traditional banks.

Although some progress has been made in proposing a more comprehensive P2P lend-

Table 2.1: Seminal papers that analyze P2P lending outcomes with social networks

Year	Authors	Paper	P2P platform	Findings	Methods
2012	Haewon Yum, Byungtae Lee, Myungsin Chae	From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms	Popfunding.com (Korea)	- Lenders seek the wisdom of the crowds when information on creditworthiness of new borrowers are very limited.	Logistic Regression
2008	Seth Freedman, Ginger Zhe Jim	Do Social Networks Solve Information Problems for Peer-to-Peer Lending? Evidence from Prosper.com	Prosper.com (USA)	- Loans with friend endorsements and friend bids have fewer missed payments and yield significantly higher rates of return than other loans. - Group leader rewards motivated leaders to fund lower quality loans in order to earn their rewards.	Logistic Regression
2009	Mingfeng Lin, N.R. Prabhala, Siva Viswanathan	Social Networks as Signaling Mechanisms: Evidence from Online Peer-to-Peer Lending	Prosper.com (USA)	- The stronger the social tie, and the more verifiable and visible to lenders, the greater the probability to attract funds. - Social networking variables reduce the interest rates on funded loans; - Borrowers with strong and verifiable relationships are less likely to default.	Logistic Regression
2015	De Liu, Daniel Brass, Yong Lu, Dongyu Chen	Friendships in Online Peer-to-PeerLending: Pipes, Prisms, and Relational Herding	PPDai.com (China)	- distrust of "friend bid" by third parties, supported by view that friends of borrowers feel a social obligation to endorse or support their friends; - people are more likely to follow the "wisdom of the crowds" when those crowds include friends rather than strangers.	Logistic Regression
2015	Xiangru Chen, Lina Zhou, Difang Wan	Group social capital and lending outcomes in the financial credit market: An empirical study of online peer-to-peer lending	Prosper.com (USA)	- Group social capital generated from group networks improves funding performance but has no impact on repayment performance. - Group networks are formed in a virtual online environment where group members are anonymous and have little opportunity for a face-to-face interaction. - As opposed to friendship networks, where relationships are formed outside the online context with a real life connection.	Logistic Regression
2015	Seth Freedman, Ginger Zhe Jim	The information value of online social networks: Lessons from peer-to-peer lending	Prosper.com (USA)	- Borrowers with social ties are consistently more likely to have their loans funded and receive lower interest rates - However, most borrowers with social ties are more likely to pay late or default. - Only endorsements from friends who also contribute money to the loan themselves produce consistently better *ex post* repayment.	Logistic Regression

ing model, it still very incipient the number of papers assessing specifically the trust or social capital parts and incorporating them into a credit scoring model.

Wei et al.[61], to the best of knowledge is the first paper that develop a series of models to study the impact of social tie formation and compare the accuracy of credit scores with and without network data. They did not however evaluate the performance on real data.

Tan et al.[55] claimed to be among the first paper to empirically predict repayment in P2P loans based on social networks. In their research, they used microfinance loans data together with Facebook profiles and full history of interactions with friends, and applied both attribute-based models and graph-based models. Overall they showed that social network based prediction alone improve the repayment by more than 18% as compared with traditional models.

Hwang et al.[33] simulated a loan funding in existing P2P platforms adding user's Facebook social information and by collecting user's response to a questionnaire. By integrating graph-based social metrics, they achieved a higher funding rate than the existing P2P lending platforms.

Gao et al.[23] obtained a dataset from a P2P lending platform that has its own embedded social network environment. In this platform, users can invite his/her friends to register whereby they interact with each other by posting comments and discussing in forums. They found evidences, by using signaling theory, that friendship network does provide significant insights in determining is/her loan success in terms of the number and also the percentage of friends with successful loan applications.

In the next section we will look at another area in computer science known as Trust and Reputation systems, which involves creating algorithms and models to assess and quantify the credibility of entities in digital environments, facilitating reliable interactions and informed decision-making.

2.1.3 Trust and Reputation Systems

Online social networks[36] are popular tools for users, as to find new friends who share similar interests, maintain social relationships, and locate various user-generated content. The high popularity of online social networks even leads to a new computer paradigm, that is social computing. Nowadays, more and more people join online social networks for daily communications or even business activities. Many of those activities involve the process of forming opinions on a particular user or product, about whether to trust or not. Interactions among people in online social networks are rather complex, as they involve interactions with others, who may even be strangers.

The notions of trust are the central issues in online social networks. Thus, quantifying trust is a notoriously difficult problem, because of the complexity of the online social network and the trust itself. Studies in trust span multiple disciplines including economics,

sociology, political science, psychology, and computer science. Meanwhile, researchers in online social networks have been conducted from multiple aspects, including network structure, user behaviors, community detection, and so on[36].

One of the most commonly accepted definitions of trust is from the sociologist Diego Gambetta[56] "... trust (or symetrically, distrust) is a particular level of subjective probability with which an agent will perform a particular action..." A trust relationship exists between two agents when one agent has an opinion about the other agent's trustworthiness and a recommendation is an opinion about the trustworthiness from a third-party agent. Reputation is defined as an "expectation about an agent's behavior based on information about or observations of his past actions". Therefore, reputation can be considered as a collective measure of trustworthiness (in the same sense of reliability) based on the referrals or ratings from members in a community.

The trust mechanisms (or trust evaluation) is a tool used to facilitate decision making in diverse applications. Trust and trust-related issues have attracted attention in various networking environments, including online social networks, wireless communication networks, multi-agent systems, and P2P networks.

Trust and reputation systems can use explicit or implicit information for decision making[56]. Example of explicit information derives from let parties generate feedback about each other after completion of a transaction and aggregating the feedback to derive a reputation score. Example of implicit information can be found in social networks such as Facebook or LinkedIn. Entities within a social network can extract some degree of trust from information gathered through friends of friends (FOAF). Another implicit form of trust information is the use of topological analysis in online social networks to determine reputation[2]. In the Google search engine, reputation is determined by the number of links that point to a page, and from where the links originate.

The reputation score is then disseminated to assist others in deciding whether to trust that party in the future. Figure 3 shows a schematic flow of a trust system.

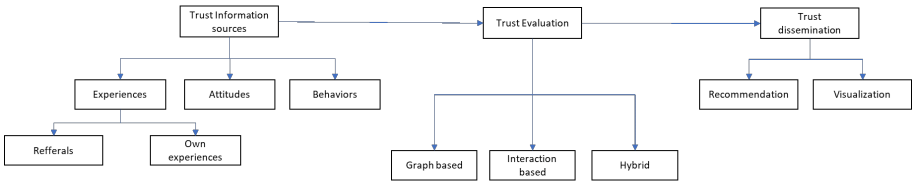


Figure 2.10: Trust evaluation framework

Resnick et al.[49] further define three requirements for a reputation system:(1) to help people decide whom to trust, (2) to encourage trustworthy behavior and appropriate effort, and (3) to deter participation by those who are unskilled or dishonest. Battiti et al.[24] expanded this idea adding the requirements that a reputation system (4) must preserve anonymity associating a peer's reputation with an opaque identifier, and (5) have minimal overhead in terms of computation, storage and infrastructure.

The participation of an individual in a virtual community strongly depends on whether and how much benefit they expect to derive from their participation. If an individual feel that they will not gain anything from joining the community, they are unlikely to participate. Hence, a good reputation system must be designed such that it incentivizes cooperation by making it profitable for users to participate and withhold services that are perceived as benefits[36].

An equally serious challenge to distributed systems and virtual communities comes from users who act in a malicious fashion with the intention of disrupting the system [cheating and other examples]. Therefore, a central objective of all digital reputation schemes is to identify such users and punish them or exclude them from the community. This exclusion is usually achieved by allowing members to distinguish between trustworthy and untrustworthy members. An untrustworthy member will not be chosen for future interactions[24]. DuPont and Karpoff [16] emphasizes that for a good system to work in practice there must be a conjunction of three forces: Law enforcement, reputation and culture. While the impact of loss of reputation is strong in an homogeneous cultural group as consequence of a fraudulent activity, making sure there will be also penalities with respect to the law, it helps to make the system more robust.

Currently, researches on trust evaluation are conducted using the tools in mathematics, statistics, or artificial intelligence. To be specific, many trust models have been proposed using fuzzy theory, subjective logic, machine learning, information entropy, game theory, graph theory, and so on. Table 2.2 shows some of the various trust and reputation systems that have been proposed in the literature by Sherchan et al.[53]. While this list is not exhaustive, it shows the diversity of approaches and perspectives.

Just to illustrate a few, Eigentrust[37] proposed by Kamvar et al. presented a distributed algorithm for the computation of the trust values of all peers in a network. Its algorithm is inspired by the PageRank algorithm used by Google and assumes that trust is transitive. A user weights the trust ratings it receives from other users by the trust it places in the reporting users themselves. Global trust is then computed in a distributed fashion by updating the trust vector at each peer using the trust vectors of neighboring peers. They show that trust values asymptotically approach the eigenvalue of the trust matrix, conditional on the presence of pretrusted users that are always trusted.

In PeerTrust[1], Xiong and Liu define five factors used to compute the trustworthiness of a peer: i) feedback obtained from other peers, ii) scope of feedback such as number of transactions, iii) credibility of the feedback source, iv) transaction context factor to differentiate between mission-critical and noncritical transactions and v) community context factor for addressing community-related characteristics and vulnerabilities.

TidalTrust[25], given two peers in the network, the model generates a calculation about the trust degree that one peer can put on the other, based on the trust paths. The trusted paths are explored by taking breadth-first search from the trustor to the trustee.

Table 2.2: Comparison of Existing Trust and Reputation Systems in the Literature[53]

Method	Origin Discipline	Trust Properties	Trust Computation Model	Trust Information Collection	Trust Evaluation	Trust Dissemination	Malicious Attack Resistance	Application Domains
EigenTrust Kamvar et al. (2003)	C	Pr, Sj	LN	E	G	TR	Y	P2P
PeerTrust Xiong et al. (2004)	C	Dy, Cs, Sj	LN	E	G	TR	Y	P2P
Yu et al. (2004)	C	Pr, Sj	LN	E	G	TR	Y	P2P
TidalTrust Goldbeck at al. (2005)	C	Pr, Sj	Prob.	E	G	TR	N	SN
Josang et al. (2006)	C	Dy, Pr, Sj	Prob.	BI	G	TR	N	P2P
Zhang et al. (2006)	S	Pr, Sj	LN	E	G	TR	N	SN
SUNNY Kuter et al. (2007)	C	Sj	BN	B	G	-	N	SN
Maheswaran et al. (2007)	C	Pr, Cs, Sj	LN	E	G	TR	N	SN
PowerTrust Zhou et al. (2007)	C	Pr, Sj	BN	E	N	-	Y	P2P, SN
Caverlee et al. (2008)	S	Dy	LN	B, E	G	TR	Y	P2P, SN
Liu et al. (2008)	C	Dy	Binary Classification	B	I	-	N	SN, e-commerce
Paradesi et al. (2009/2009)	C	Cp	BN	E	-	-	N	WS
Yan et al. (2009)	C	Sj	LN	B	-	-	N	MA
Zuo et al. (2009)	C	Pr, Cp	LN	TC	G	-	N	SN
Adali et al. (2010)	S	-	Log	B	G	-	N	SN
Nepal et al. (2010)	S	Es	Prob.	B	-	-	N	WS, SN
Trifunovic et al. (2010)	S	Pr, Sj	LN	B, E	H	-	Y	SN
Nepal et al. (2011)	S	Sj, Dy, Cs	LN	B	I	-	N	SN

Origin Discipline	Trust Properties	Trust Computation Model	Trust Information Collection
P: Psychology S: Sociology C: Computer Science	Cs: Context-specific, Dy: Dynamic, Pr: Propagative, Cp: Composable, Sj: Subjective, Es: Event Sensitive	Log: Logarithmic, Prob: Probabilistic, BN: Bayesian Networks, LN: Linear Model (sum or product)	B: Behavior E: Experience, BI: Belief, TC: Trust Certificate

Trust Evaluation	Trust Dissemination	Attack Resistance	Application Domains
G: Graph-based, I: Interaction-based, H: Hybrid, -: Not specified	TR: Trust-based Recommendation, VZ: Visualisation, -: Not specified	Y: Malicious attack considered, N: Malicious attacks not considered	P2P: Peer-to-peer networks, SN: Social Networks, WS: Web Services, MA: Mobile Applications

Therefore, only the shortest strongest trusted paths are used in TidalTrust.

Regret[51] expands the sources coming from i) direct interactions and ii) propagation of experiences from other peers (witnesses) to a third component: information obtained from the analysis of their social network and its topology. The Regret takes a subset of the selected sociogram over the peers that had had interactions with the target peer as the initial graph and apply the following heuristic: identify strongly connected components, find the set of cut-points for each component. The cut-point indicates some kind of local centrality, because if they are eliminated the graph is subdivided in other components. For each component that does not have cut-points, the model chooses the node with the larger degree. The aggregation of these selected peers composes the witness reputation, which took the social relations into account.

Chapter 3

Research Methodology

In the preceding sections, it was introduced the problem of financial exclusion of poor people in the traditional scoring systems. Moreover, it was introduced the idea of using *social capital* to overcome this problem. In fact, it was shown that the idea is not new, many communities around the World have been using informal ways of microfinance during long time. Rotating Savings and Credit Associations (ROSCAs) have its origins in Tanomoshi-Kou in Japan, much before the existence of formal banks. Grameen Bank is another example of how social mechanisms such as "joint-liability" use the social monitoring among the group of people to reduce incentives to moral hazard, and thus contribute to the higher repayment rates of loans.

Social capital is a multidisciplinary concept and have been studied in the early days mainly from a sociological and economical perspective. In both areas, the concept of trust and trustworthiness have emerged as a proxy to measure *social capital* through trust games and attitudinal questions.

More recently, with the emergence of online social networks, *social capital* started to be studied from a computer science perspective. We have seen that through the analysis of the graph structure in the social context, we can identify two kinds of *social capital*. One that comes from the effect of "bonding" of a particular person or node, and other that comes from the effect of "bridging or brokerage".

Another subset of specialization in computer science is known as Trust or Reputation Systems. It tackles the problem of identifying trustworthy nodes in several areas such as peer-to-peer networks, e-commerce, opinion reviews and also social networks.

Together these two specializations can offer a good framework to apply in experiments and provide a good basis for discussion.

3.1 Approach

So, the problem that this research tries to shine a light, is how can we use some of the preceeding ideas to analyze and understand social capital dynamics in social networks that

lead to succesful financial credit transactions. If so, how can we use these learnings to make inferences about identifying possible creditworthy people in a novel social network.

In this context, the research followed the structure:

1. Observation and description of real cases of informal lending arrangements, ROSCA and a peer-to-peer (P2P) loans. Community structure data from these two cases were collected from the field and served the basis for applying some trust and reputation algorithms.

2. Analysis of role of social capital in the success of financial transactions within ROSCAs and P2P loans. This analysis will use techniques from both computer science specializations: Social Network Analysis using graph theory and Trust and Reputation algorithms. The metrics and algorithms chosen will be presented in the next sections.

3. Application of findings to identify creditrustworthy people in a community. For that we have used the social network dataset from the paper "The Diffusion of Microfinance" [4]

4. Conclude by suggesting a theoretical framework to apply in financial institutions.

In the next sections we will show in more details the methods and algorithms used in the analysis of the experiments.

For the quantitative analysis it was used a series of python notebooks. For graph visualization, it was used neo4j.

All the code and datasets is available at https://github.com/robertoshimizu/creditworthiness-socialcapital.git .

3.2 Social Network Analysis

According to Nedham/Holder[40], there are three areas of analysis that are at the heart of graph algorithms. These categories correspond to pathfinding and search, centrality computation, and community detection.

Pathfinding: Paths are fundamental to graph analytics and algorithms. Finding shortest paths is probably the most frequent task performed with graph algorithms and is a precursor for several different types of analysis. As we have seen before, the shortest path is the traversal route with the fewest hops or lowest weight. If the graph is directed, then it's the shortest path between two nodes as allowed by the relationship directions.

Community Detection: Connectedness is a core concept of graph theory that enables a sophisticated network analysis such as finding communities. Most real-world networks exhibit substructures (often quasi-fractal) of more or less independent subgraphs.

Connectivity is used to find communities and quantify the quality of groupings. Evaluating different types of communities within a graph can uncover structures, like hubs and hierarchies, and tendencies of groups to attract or repel others. These techniques are used to study emergent phenomena such as those that lead to echo chambers and filter bubble effects.

Centrality: Centrality algorithms are used to understand the roles of particular nodes in a graph and their impact on that network. They're useful because they identify the most important nodes and help us understand group dynamics such as credibility, accessibility, the speed at which things spread, and bridges between groups. Many of these algorithms were invented for social network analysis[35, 40, 17, 44, 52, 22, 47], and for this reason, some of them we will use in the experiments:

- **Degree centrality**: The degree a given peer is the number of edges (neighboring nodes) that a peer has. Peers who are well-known and highly regarded by other members of the community tend to be easily identified as highly connected node in the social network.

- **Betweenness centrality**, measures the number of shortest paths that pass through a node. It identifies nodes in the network that are in central positions that control information flow and are therefore influential.

  The betweenness centrality of a node is calculated by adding the results of the following formula for all shortest paths:

$$B(u) = \sum_{s \neq u \neq t} \frac{p(u)}{p} \tag{3.1}$$

  where:

  - u is a node.
  - p is the total number of shortest paths between nodes s and t.
  - $p(u)$ it the number of shortest paths between nodes s and t that pass through node u.

- **Closeness centrality**, measures which nodes have the shortest paths to all other nodes. It is a useful measure that estimates how fast the flow of information would be through a given node to other nodes.

  For each node, the Closeness Centrality algorithm calculates the sum of its distances to all other nodes, based on calculating the shortest paths between all pairs of nodes.

The resulting sum is then inverted to determine the closeness centrality score for that node.

It is more common to normalize this score so that it represents the average length of the shortest paths rather than their sum. This adjustment allows comparisons of the closeness centrality of nodes of graphs of different sizes.

$$C_{norm}(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(u,v)} \tag{3.2}$$

where:

- u is a node.
- n is the number of nodes in the graph.
- $d(u,v)$ is the shortest-path distance between another node v and u.

- **Eigenvector centrality** measures a node's importance while giving consideration to the importance of its neighbors. Relationships originating from high-scoring nodes contribute more to the score of a node than connections from low-scoring nodes. A high eigenvector score means that a node is connected to many nodes who themselves have high scores.

  Let $A = (a_{i,j})$ be the adjacency matrix of a graph. The eigenvector centrality x_i of node i is given by:

$$x_i = \frac{1}{\lambda} \sum_k a_{k,i}\, x_k \tag{3.3}$$

  where $\lambda \neq 0$ is a constant. In matrix form we have:

$$\lambda x = xA \tag{3.4}$$

  Hence the centrality vector x is the left-hand eigenvector of the adjacency matrix A associated with the eigenvalue λ. It is wise to choose λ as the largest eigenvalue in absolute value of matrix A. The power method can be used to solve the eigenvector centrality problem.

- **Clustering coefficient**: measures how connected are neighboring nodes. (how connected the friends of a peer are). The close to 1 means that your friends are also connected to each other, close to 0, your friends are not connected to themselves whatsoever. It counts the number of triangles in the ego-network.

  The formula and calculation example have been already described in section 2.

In the experiments we will use these metrics using python package Networkx [31]

3.3 Trust and Reputation Algorithms

From all the algorithms listed in section 2.1.3, Eigentrust[37] and Tidaltrust[25], are indeed valuable foundations for modeling trust in social networks. Both algorithms leverage the network structure and trust relationships to compute trust values. Eigentrust is particularly suitable for the computation of the trust values of all peers in a network, while Tidaltrust focuses on the strongest trust paths between two peers.

In the case of analyzing and compute trust score in the ROSCA, which is a closed system of a very knit group, eigentrust is a relative simple method to identify the most important nodes, either from the perspective of the node's role in the structure of the graph, but also from the perspective of taking into account the transmission and aggregation of opinion about nodes's peers (if only when this information is available).

In the case of the P2P loan, the challenge is to understand how a person who is not directly connected to a borrower, can have enough trust to decide to fund a loan. In this case, tidaltrust algorithm seem to be appropriate, because it calculates trust between a source and a sink node, using the paths that provide highest trust.

The simplicity of both algorithms also was key for selecting them, once as it will be discussed later on, the breadth of data in the experiments dataset were limited.

In the next sections, we will bring more background of both selected algorithms, that supported the decision to use them, as well as provide a good understand later on in the experiments of the insights and results.

3.3.1 Eigentrust

Eigentrust[37] is known as a very important paper that address the computation of a global trust score for all participants of a network. It has been written back in 2003, when there were several platforms to share files directly between peers, including music (mp3). One problem in that set up was to spot unreliable peers that shared low quality files or even malicious files such as malwares. At the same time it was starting to flourish e-commerce platforms such as e-bay, where users could rate other users after a purchase transaction, so it was also important to recognize the reputation (good or bad) of the participants.

With this problems in mind, the researchers proposed eigentrust, a reputation system that aggregates the **local trust values** of all users with minimum overhead. It is based on the concept of transitive trust: A peer i will have a high opinion of those peers who have provided it authentic files. Moreover, peer i is likely to trust the opinions of those peers, since peers who are honest about the files they provided are also likely to be honest about providing opinions about other peers. These opinions are defined in eigentrust as **local trust**. Therefore, the *global reputation* of each peer i is given by the local trust values assigned to peer i by other peers, weighted by the global reputation of the assigning

peers.

For example, consider the folowing social network depicted in Figure 3.1.

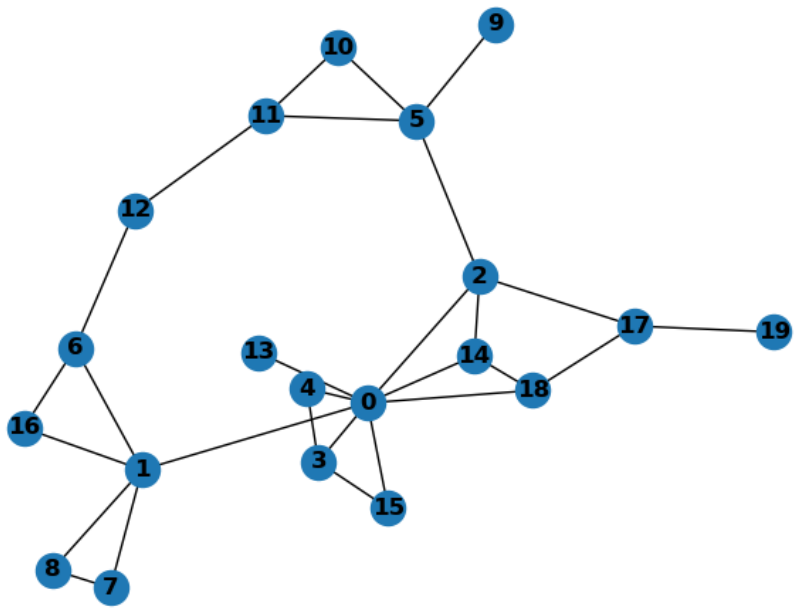


Figure 3.1: Network in eigentrust example

Each time peer i makes a transaction with peer j, it may rate the transaction as:

$$positive : tr(i, j) = +1 \tag{3.5}$$

$$negative : tr(i, j) = -1 \tag{3.6}$$

We may then define the **local trust value** s_{ij} as the sum of the ratings of the individual transactions that peer i has done with peer j:

$$s_{ij} = \sum tr_{ij} \tag{3.7}$$

Now suppose that in an instant time t we have the local trust matrix s as in Table 3.1.

We can see that peer $i = 2$ has a local trust value of peer $j = 5$ by $+3$. Similarly, it has a local trust value of the peer $j = 17$ by -3.

$s_{2,5} = +3$

$s_{2,17} = -3$

$$
\begin{bmatrix}
0 & 5 & 1 & 3 & 11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 6 & 3 & 0 & 0 & 4 & 0 \\
6 & 0 & 0 & 0 & 0 & 0 & 0 & 11 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 & -3 & 0 & 0 \\
5 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
8 & 0 & 0 & 11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 16 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\
0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 5 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 13 & 0 & 0 & 0 & 0 & 3 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 7 & 0 \\
-3 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \\
3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 7 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 7 & 0 & 0
\end{bmatrix}
$$

Table 3.1: Adjacency matrix with local trust values

In order to aggregate local trust values, eigentrust proposed to normalize them (for computation reasons), therefore, they defined a **normalized local trust value**, c_{ij} , as follows:

$$c_{ij} = \frac{max(s_{ij}, 0)}{\sum_j max(s_{ij}, 0)} \tag{3.8}$$

This ensure that all values are between 0 and 1. Table 3.2 shows the normalized local trust from the perspective of node 2.

node	0	1	2	3	4	5	6	7	8	9
adjacency	1	0	0	0	0	1	0	0	0	0
local trust	0	0	0	0	0	3	0	0	0	0
normalized trust	0	0	0	0	0	0.38	0	0	0	0
node	10	11	12	13	14	15	16	17	18	19
adjacency	0	0	0	0	1	0	0	1	0	0
local trust	0	0	0	0	5	0	0	-3	0	0
normalized trust	0	0	0	0	0.62	0	0	0	0	0

Table 3.2: Trust values given by peer 2

Finally they **aggregate the normalized local trust values**. A natural way to do this in a distributed environment is for peer i to ask its acquaintances about their opinions about other peers.

$$t_{ik} = \sum_j c_{ij} c_{jk} \tag{3.9}$$

where t_{ik} represents the trust that peer i places in peer k based on asking his friends.

We can write $t_{ik} = \sum_j c_{ij} c_{jk}$ in matrix notation. If we define C to be the matrix $[c_{ij}]$ and $\overrightarrow{t_i}$ to be the vector containing the values t_{ik}, then $\overrightarrow{t_i} = C^T \overrightarrow{c_i}$. (Note that $\sum_j t_{ij} = 1$).

However, the trust values stored by peer i still reflect only the experience of peer i and his acquantaincs. In order to get a wider view, peer i may wish to ask his friends' friends $\overrightarrow{t_i} = (C^T)^2 \overrightarrow{c_i}$.

If he continues in this manner, $\overrightarrow{t_i} = (C^T)^n \overrightarrow{c_i}$, he will have a complete view of the network after n =large iterations (under the assumptions that C is irreducible and aperiodic.

Fortunately, if n is large, the trust vector $\overrightarrow{t_i}$ will converge to the same vector for every peer i. Namely, it will converge to the left principal eigenvector of C. In other words, $\overrightarrow{t}$ is a global trust vector in this model. Its elements, t_j , quantify how much trust the system as a whole places peer j.

Algorithm 1: Basic Eigentrust algorithm

$\overrightarrow{t}^{(0)} = \overrightarrow{e}$
repeat
  $\overrightarrow{t}^{(k+1)} = C^T \overrightarrow{t}^{(k)}$
  $\delta = || \overrightarrow{t}^{(k+1)} - \overrightarrow{t}^{(k)} ||$
until $\delta < \epsilon$

In our example, running the eigentrust algorithm, we reach to the follow results for $\overrightarrow{t}$ shown in Table 3.3. The pseudocode is in algorithm 1.

node	0	1	2	3	4	5	6	7	8	9
degree	8	5	4	3	2	4	3	2	2	1
eigentrust	0.11	0.05	0.07	0.05	0.05	0.15	0.01	0.02	0.01	0.03
node	10	11	12	13	14	15	16	17	18	19
degree	2	3	2	1	3	2	2	3	3	1
eigentrust	0.08	0.11	0.01	0.00	0.08	0.01	0.02	0.06	0.06	0.03

Table 3.3: Eigentrust global values

In our example, the model places a high global trust in nodes 0, 5 and 11. The results are a very sensitive to the position of the node in the network and its degrees as well as the local trust values.

3.3.2 Tidaltrust

In tidaltrust[25], the approach is based on the assumption that neighbors with higher trust ratings are likely to agree with each other about the trustworthiness of a third party, here

we call this node as sink.

The goal is to infer how much the given node (the source) would trust the sink node. The tidaltrust algorithm starts then from the node source, polling each of its neighbors to obtain their ratings of the sink. If none of them is linked directly to the sink, the algorithm will search for their neighbors, and then recursively until reach the sink. During these iterations, each neighbor keeps track of the current depth from the source and each node will also keep track of the strength of the path to it. The strength of the path to each neighbor is the minimum of the source's rating of the node and the node's rating of its neighbor. The neighbor records the maximum strength path leading to it. Once a path is found from the source to the sink, the depth is set at the maximum depth allowable. Since the search is proceeding in a Breadth First Search fashion, the first path found will be at the minimum depth. The search will continue to find other paths at the minimum depth, and once it is completed, the trust threshold is established by taking the maximum of the trust nodes leading to the sink. The pseudocode is in algorithm 2.

In the example in Figure 3.2, starting from the source it will verify nodes A, B and C. In this first level, each node takes the rating from source to itself, so A takes 10, B takes 8 and C takes 9.

Then node D takes the maximum among the paths coming to it, which is the maximum between 8 and 9. In this case, 8 comes from the minimum between the rating of B which is 8, and B's rating of D which is 10.

Similarly, node E's rating is the maximum of 8, 8 and 9, where the first 8 comes from the minimum between C's rating itself which is 9 and C's rating of E which is 8. The second 8 comes from the the minimum between B's rating itself which is 8 and B's rating of E which is 10. And finally, the 9 comes from the minimum between A's rating itself which is 10 and A's rating of E which is 9

The threshold value is determined once the sink is reached. This value is defined as the maximum rating of the nodes in the last level of the graph that are connected to the sink, that have some trust towards the sink. In this example, both nodes E and D, have trust towards the sink valued 9. The threshold is the maximum between these ratings and in this example, its value is 9.

Once the trust threshold is established, the algorithm continues with its second part where the trust threshold is used. In the second part of the algorithm, trust towards the sink is propagated through all levels back to the source. Every node in the graph, except the nodes in the final level of the graph that have direct trust values towards the sink, calculate their trust values towards the sink by this expression:

$$t_{ik} = \frac{\sum_{j \in adj(i) | t_{i,j} \geq max} t_{ij} * t_{jk}}{\sum_{j \in adj(i) | t_{i,j} \geq max} t_{ij}} \tag{3.10}$$

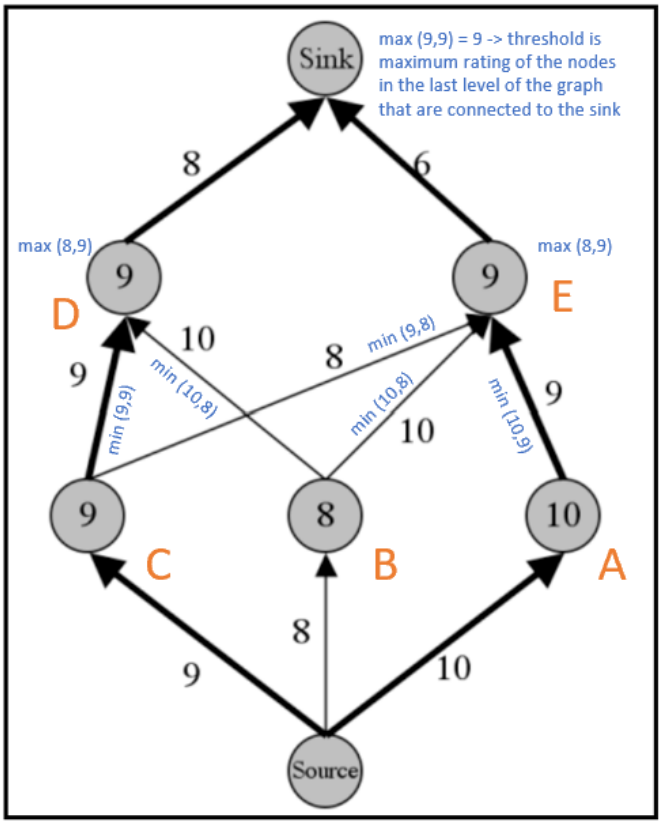


Figure 3.2: Network in tidaltrust example

Notice that in the above expression it will use a variable *max* that represents the largest trust value that can be used as a *minimum* threshold that a path can be found from source to sink. Because the threshold value calculated is 9, then for a given node i, just the paths to j whose strength is higher or equal 9 will be considered in the summation. The bold edges in the Figure 3.2 indicate which paths will ultimately be used in the calculation because they are at or above the max threshold.

The final trust rating from source to sink is the weight average of the two paths, as in the expression above, which leads to 7.

Gathering trust ratings using FOAF

In the tidaltrust paper[25] and another [26] Goldbeck proposes an extension of the ontology Friend-Of-A-Friend (FOAF) to allow users to indicate a level of trust for people they know. They proposed a level of trust on a scale of 1-9. The levels roughly corresponded to the following:

1. Distrusts absolutely
2. Distrust highly
3. Distrust moderately

Algorithm 2: TidalTrust Algorithm

```
foreach n ∈ G do
    color(n) = white
    q = empty
TidalTrust(source, sink)
    push(q, source)
    depth = 1
    maxdepth = infinity
    while d(depth) not empty and depth ≤ maxdepth do
        n = pop(d(depth))
        push(d(depth), n)
        if sink ∈ adj(source) then
            cached_rating(n, sink) = rating(n, sink)
            maxdepth = depth
            flow = min(path_flow(n), rating(n, sink))
            path_flow(sink) = max(path_flow(sink), flow)
            push(children(n), sink)
        else
            foreach n2 ∈ adj(n) do
                if color(n2) = gray then
                    color(n2) = grey
                    push(temp_q, n2)
                if n2 ∈ temp_q then
                    flow = min(path_flow(n), rating(n, n2))
                    path_flow(n2) = max(path_flow(n2), flow)
                    push(children((n), n2))
        if q empty then
            q = empty_q
            depth = depth + 1
            temp_q = empty
    max = path_flow(sink)
    depth = depth − 1
    while depth > 0 do
        while d(depth) not empty do
            n = pop(d(depth))
            foreach n2 ∈ children(n) do
                if rating(n, n2) ≥ max and cached_rating(n2, sink) ≥ 0 then
                    numerator =
                      numerator + rating(n, n2) ∗ cached_rating(n2, sink)
                    denominator = denominator + rating(n, n2)
                if denominator ≥ 0 then
                    cached_rating(n, sink) = numerator/denominator
                else
                    cached_rating(n, sink) = −1
        depth = depth − 1
    return v = 0
```

4. Distrust slightly
5. Trust neutrally
6. Trust slightly
7. Trust moderately
8. Trust highly
9. Trust absolutely

These ratings are fundamental components to weight the strength of trust relationships and allow the tidaltrust algorithm to make its inferences.

However, when it is not possible to gather trust ratings explicitely from users, other kinds of predicates in the FOAF ontology could be used. For example the degree of relationship based on family and friendship ties. For example, the fact that A *is a sibling of* B, we could assign a different trust weight than a case in which C *is an acquaintance of* D.

Chapter 4

Results and Analysis

In this section we will show the results of the analysis of the three proposed experiments:

- ROSCA
- P2P loan
- Database of article "The Diffusion of Microfinance

4.1 ROSCAs and P2P Loans

4.1.1 The ROSCA - Consórcio entre Amigos

During the course of the research, we could observe a real example of a ROSCA - Consorcio entre Amigos and had the opportunity to interview the group's leader and gather data about the structure of the network. These data do not contain any confidential or sensitive information that would reveal identity of the participants. According to the group's leader, ROSCA is a very common practice in the region of North of Minas Gerais State, where most of the participants are from, since long time ago.

People normally join in to form a ROSCA with different goals. Some may participate as **borrowers**, to get some money for immediate use, such as an emergency or to pay outstanding debts. Others just want to participate as **investors**, using a ROSCA as a way to get discipline to save money and get it back by the end with some interest earned.

You may recall from Figure 4.1 how a ROSCA works.

The group's leader described the structure of 4 instances of his ROSCA, which has occurred each one every year for the last 4 years. Each one had an average group size of ten people, in which most of the participants were recurring participants from previous ones. In Figure 4.2 it is depicted the structure of the most recent ROSCA (no. 4).

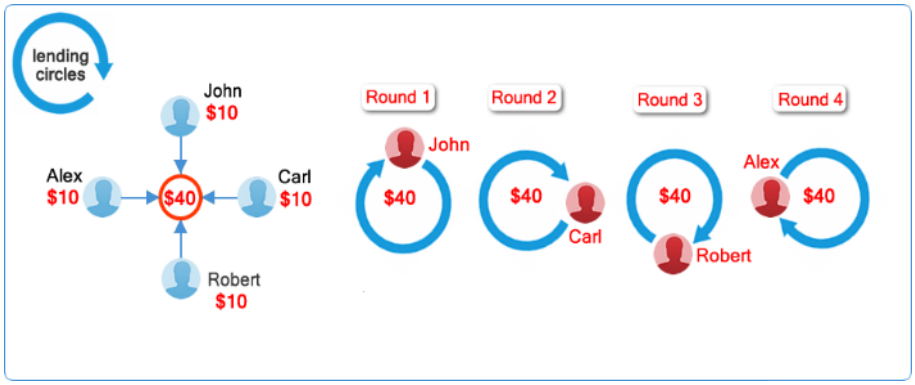


Figure 4.1: Dynamics of a Rotating Savings and Credit Association (ROSCA)

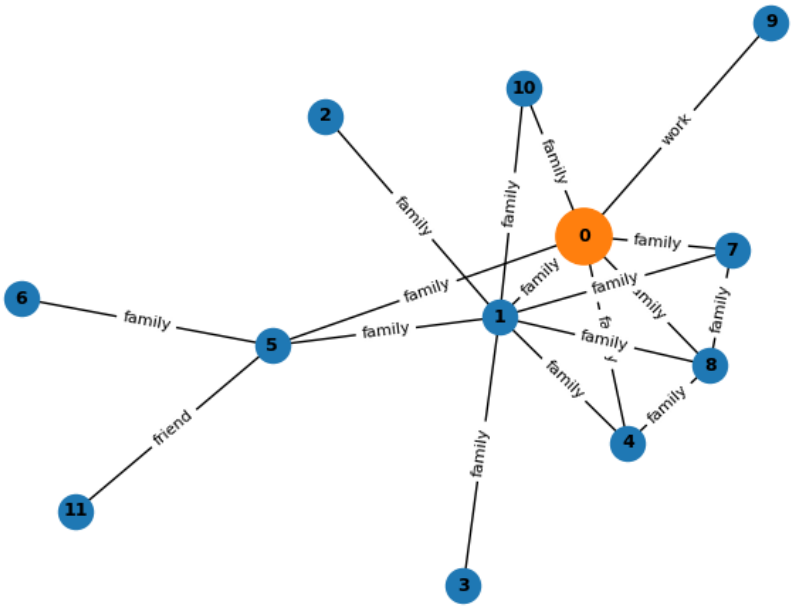


Figure 4.2: Graph structure of ROSCA no 4

Notice that the group's leader (node 0) is highlighted in orange, and most of the connections are between family ties.

Table 4.1 helps to complement the understanding of the structure, using social network analysis centrality metrics.

Table 4.1: Graph metrics of ROSCA no 4

Metric	0	1	2	3	4	5	6	7	8	9	10	11
node degree	7	8	1	1	3	4	1	3	4	1	2	1
clustering coeficient	0.33	0.25	0.00	0.00	1.00	0.17	0.00	1.00	0.83	0.00	1.00	0.00
betweeness centrality	0.32	0.49	0.00	0.00	0.00	0.35	0.00	0.00	0.01	0.00	0.00	0.00
closeness centrality	0.73	0.79	0.46	0.46	0.52	0.61	0.40	0.52	0.55	0.44	0.50	0.40
recurrence	4	4	1	1	4	4	1	4	4	4	1	4

The group's leader (node 0) has a very high degree (degree 7) and connects directly

to other participants including node 1 (degree 8), who has equally lots of common connections.

As a result of having two nodes with high degrees and common connections, it makes a very high clustered group. Most of the connections come from the core of a big extended family (mother, son, daughter, son of daughter-in-law, ...) and therefore they all know each other.

Both the node 0 and node 1 have very high centrality metrics. They both position in the middle of the shortest paths between all members as well as the position as bridge between more peripherical sub groups, such as nodes 5, 6 and 11. Notice that node 5 also is in a privileged position to bridge node 6 and node 11 in the ROSCA. In practice, nodes 0, 1 and 5 work like "high trust entity", the ones that seem to bring all other participants together. They are best positioned to provide "informal recommendations" about creditworthiness based on their knowledge about the people close to them.

In a ROSCA, explicit user's evaluation of others after each money round is not formally collected. Nevertheless, the periodical money transactions in every round during the life cycle and the recurrent participation of some members in previous ROSCAs could be used to evaluate user's local trust. The assumption is that it would be unlikely that a given participant would be invited again by the group's leader if he or she had exhibited a bad behavior in the past.

With this assumption in mind, a possible model could reassign for every past participation in a ROSCA, a positive score: $tr(i, j) = +1$. Therefore, if i and j have participated together in the past in 4 instances of ROSCA, then **local trust value** $s_{ij} = 4$. This assumption enable us to use **Eigentrust** algorithm to calculate the global trust of the members of this ROSCA.

Table 4.2 shows the eigentrust global trust score using the above approach. Notice that we display also pagerank scores, which is part of networkx python package. Pagerank is not much different from eigentrust in a way that it computes the global vector known as the left principal eigenvector, which essentially gives a global trust value from which it can compute each individual score.

Table 4.2: Trust Metrics of ROSCA no 4

Metric	0	1	2	3	4	5	6	7	8	9	10	11
node degree	7	8	1	1	3	4	1	3	4	1	2	1
clustering coeficient	0.33	0.25	0.00	0.00	1.00	0.17	0.00	1.00	0.83	0.00	1.00	0.00
betweeness centrality	0.32	0.49	0.00	0.00	0.00	0.35	0.00	0.00	0.01	0.00	0.00	0.00
closeness centrality	0.73	0.79	0.46	0.46	0.52	0.61	0.40	0.52	0.55	0.44	0.50	0.40
pagerank	0.179	0.208	0.035	0.035	0.078	0.121	0.038	0.078	0.100	0.034	0.056	0.038
eigentrust	0.219	0.202	0.009	0.009	0.105	0.114	0.009	0.105	0.140	0.035	0.018	0.035
recurrence	4	4	1	1	4	4	1	4	4	4	1	4

We can see that eigentrust still computes high scores to the most central and highly connected notes, but also it gives a higher score than pagerank. This is due to the effect of the recurrence and therefore it works well to reward participants past behaviors from

a completely new comer.

4.1.2 Peer-to-Peer Loans - Empréstimos entre Pessoas

Normally in the ROSCA, people contribute with more or less R$ 250,00 to R$ 300,00 per round, so overall the kind of amount people can get are in the range of R$ 2.500,00 to R$ 3.000,00, which is about a month salary of a person in this social-economic group. However, when they need a substantial amount, for example R$ 10.000,00 to R$ 20.000,00, they find it extremelly difficult to find a bank that lends with a decent interest rate. Therefore, some of them end up asking money from "family and friends", establishing a kind of peer-to-peer loan.

During this research period, there were two cases of such peer-to-peer loan, which we could capture the structure of the loan and the social network around it. It was great opportunity to analyze how a closed knit arrangement such as ROSCA could evolve to another form of credit arrangement. In both examples, the loans were succesfully repaid in the expected timeframe.

P2P number One

The first P2P case, which we shall call P2P no 1, shown in Figure 4.3, was composed by 5 people (nodes 1 to 5), each one commiting amounts shown in Table 4.3, to a person (node 0 in orange), totalizing a loan of R$ 10.000,00. The loan was repaid in 12 monthly installments.

Table 4.3: Structure of Loan - P2P no 1

node 0	R$ 10.000,00
node 1	R$ (1.000,00)
node 2	R$ (2.500,00)
node 3	R$ (1.500,00)
node 4	R$ (2.500,00)
node 5	R$ (2.500,00)

Nodes 0, 1, 2, 3 and 4 were participants of the same ROSCA analyzed earlier, however the latter (node 4) did not have a family tie with the other four. They all had in common a track record of financial transactions in the ROSCA.

The interesting fact is that this P2P received a newcomer (node 5), a person who did not participate in any previous ROSCA whatsoever, and neither had a previous relationship with any member, except with node 4.

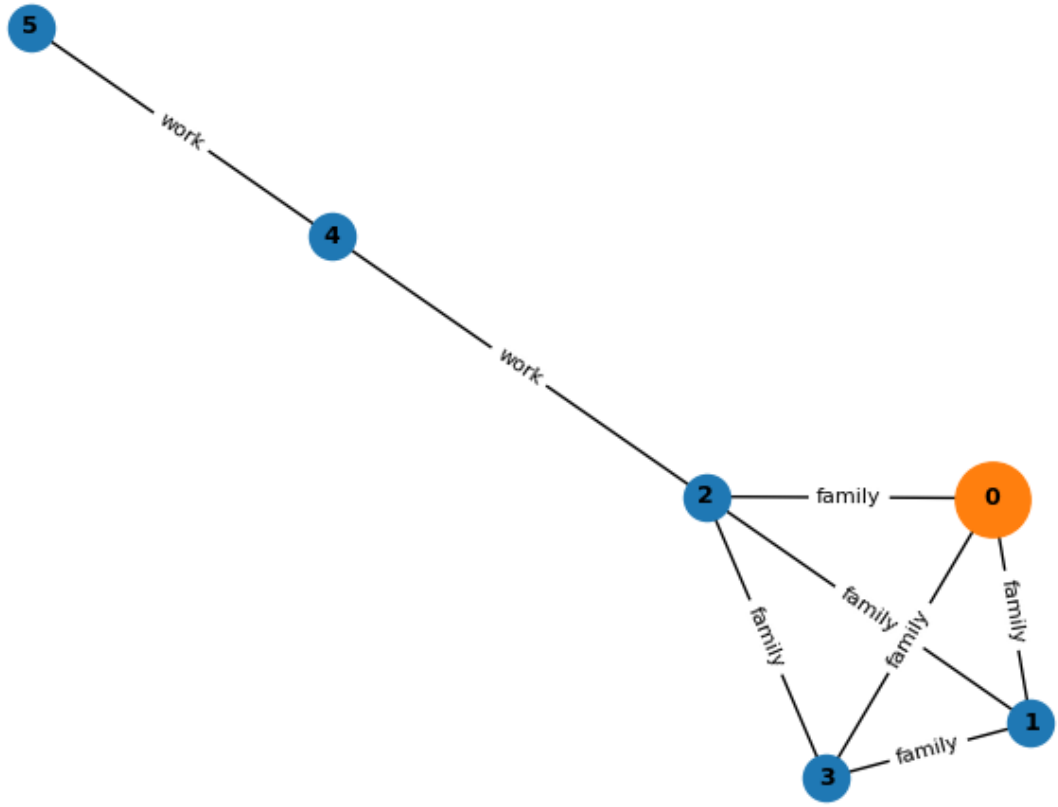


Figure 4.3: Graph structure of P2P no 1

Table 4.4 shows some metrics of the members of the network. The traditional centrality metrics and even the eigenvector centrality (which is another version of the eigenvector) characterizes well the importance of each node in the network. Node 2 and 4 are very important to bridge other people to join the loan, More specifically, node 2 brings node 4, which in turn brings node 5.

Table 4.4: Graph metrics of P2P no 1

Metric	0	1	2	3	4	5
node degree	3	3	4	3	2	1
clustering	1.00	1.00	0.50	1.00	0.00	0.00
eigenvector centrality	0.48	0.48	0.52	0.48	0.19	0.06
betweeness centrality	0.00	0.00	0.60	0.00	0.40	0.00
closeness centrality	0.63	0.63	0.83	0.63	0.63	0.42

From the point of view of node 5, how come he/she could commit a substancial amount of money R$ 2.500,00 without knowing node 0?

For this setting, we decided to use the tidaltrust algorithm. As we saw in section 3.3.2, we could use tidaltrust to infer how much node 5 (source) trusts node 0 (sink).

Similar on what we did in eigentrust, we have assumed relative weights based on the

nature of the relationship. Table 4.5 displays the proposition of relative weights, which we use to rewrite the network structure shown in Figure 4.4, based on the inputs of the P2P organizer.

Table 4.5: Relative weight of relationship

	1	3	6
friends	acquaintance	friend	close friend
family	distant relation	second degree	intimate relation
work	affiliated	colleague	reporting line

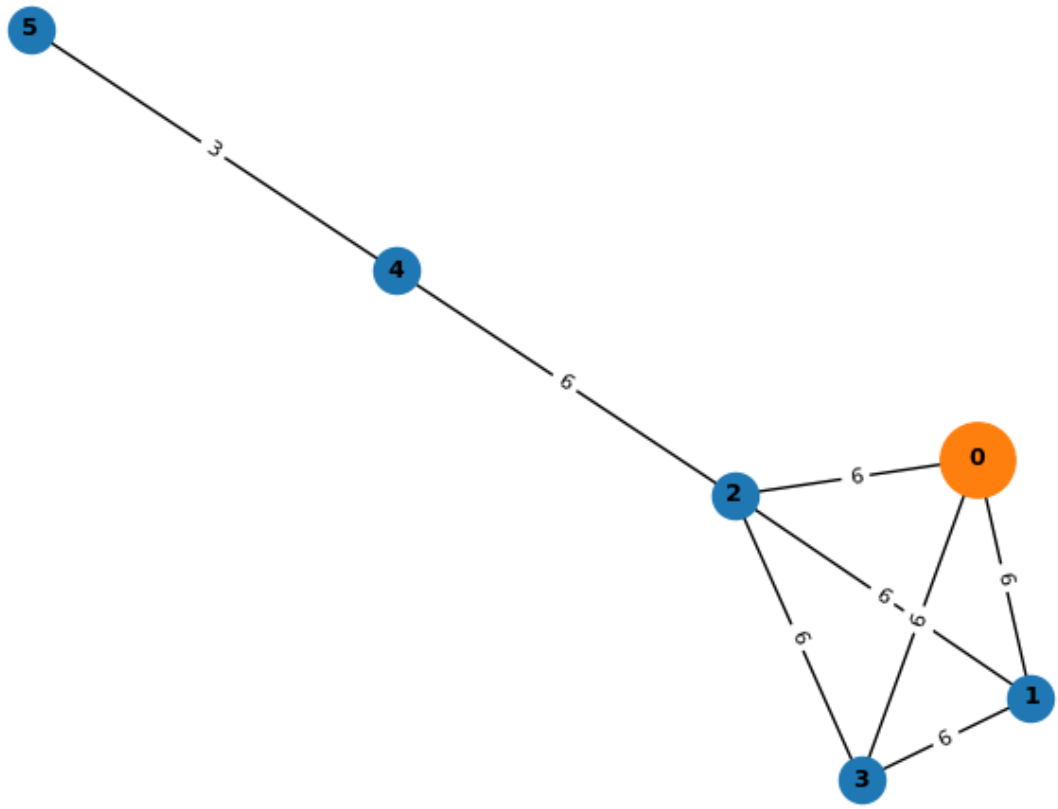


Figure 4.4: Graph structure of P2P no 1 using relative weights

This allows us to run the tidaltrust algorithm, giving us the results displayed in Table 4.6.

Table 4.6: Outputs from Tidaltrust for P2P no 1

trust	6.0
threshold	3
paths used	[[5, 4, 2, 0]]
nodes used	[0, 2, 4, 5]
nodes unused	[1, 3]
source	5
sink	0
tag	weight

Along the shortest path [5, 4, 2, 0], the minimum trust is 3.0, which is used as a threshold. Because the strength of trust between node 4 and 2 is 6, and the strength between 2 and 0 is 6, by applying the recurrent formula of tidaltrust, the trust strength from 4 to 0 yields 6. Therefore, between 5 and 0, from the perspective of 5, it relies on node 2's opinion (a trusted node who has an opinion of node 0).

P2P number Two

The second P2P case, which we shall call P2P no 2, shown in Figure 4.5, was composed by 8 people (nodes 1 to 8), each one commiting amounts shown in table Table 4.7, totalizing a loan of R$ 7.000,00 to node 0.

Table 4.7: Structure of Loan - P2P no 2

node 0	R$ 7.000,00
node 1	R$ (500,00)
node 2	R$ (1.000,00)
node 3	R$ (500,00)
node 4	R$ (500,00)
node 5	R$ (2.500,00)
node 6	R$ (500,00)
node 7	R$ (500,00)
node 8	R$ (1.000,00)

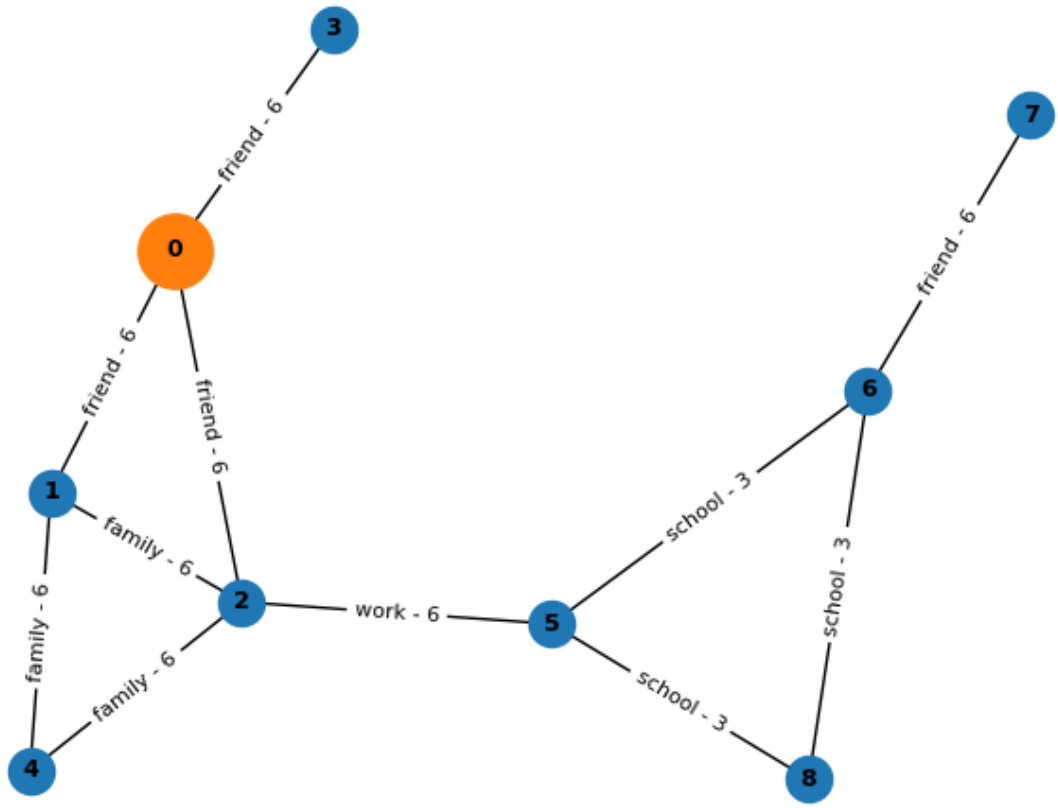


Figure 4.5: Graph structure of P2P no 2

Nodes 1, 2, 4 and 5 were participants of the same ROSCA analyzed earlier. In particular, nodes 1 and 2 of this P2P are the same nodes 1 and 2, respectively of the ROSCA in Table 4.1.

Table 4.8 shows nodes 2 and 5 having the highest betweeness and closenest centrality respectively. They were best positioned to have opinion about the people in their respective tightly-knit group, and through the bridge (edge 2 - 5), reputation can flow from one group to another. This shows a powerful combination of the two kinds of social capital (bonding and bridging) operating.

Table 4.8: Graph metrics of P2P no 2

Metric	0	1	2	3	4	5	6	7	8
node degree	3	3	4	1	2	3	3	1	2
clustering	0.33	0.67	0.33	0.00	1.00	0.33	0.33	0.00	1.00
eigenvector centrality	0.40	0.46	0.54	0.14	0.35	0.33	0.21	0.07	0.2
betweeness centrality	0.25	0.04	0.61	0.00	0.00	0.54	0.25	0.00	0.00
closeness centrality	0.47	0.47	0.62	0.33	0.42	0.57	0.44	0.32	0.42

As a result, this kind of social structure allows a distant node like 7 to be confident enough to lend money to node 0. Table 4.9 shows the trust calculated using tidaltrust, from source node 7 to sink node 0 as 6. The high trust levels between node 2 and 5, allows a high opinion of node 0 by node 2, to be transitioned highly to node 5 acquaintances. They can only know through node 5, a potential creditworthy borrowers like node 0, that they can lend money and obtain a good return for their investment.

Table 4.9: Outputs from Tidaltrust for P2P no 2

trust	6.0
threshold	3
paths used	[[7, 6, 5, 2, 0]]
nodes used	[0, 2, 5, 6, 7]
nodes unused	[1, 3, 4, 8]
source	7
sink	0
tag	weight

4.2 Social Networks and Microfinance Database (The Diffusion of Microfinance)

This database contains data from a survey of social networks in 75 villages in rural southern Karnataka, a state in India. A census of households was conducted, and a subset of individuals was asked detailed questions about the relationships they had with others in the village. About half of households received detailed surveys in which individuals were asked to list the names of people with whom they shared a certain relationship. Households were randomly sampled and stratified by religion and geographic sub-region. For each variable, an individual matrix and a household matrix were constructed. A relationship between households exists if any household members indicated a relationship

with members from the other household. These questions were asked in the individual survey. Individuals were asked who they:

- borrow money from
- give advice to
- borrow kerosene or rice from
- lend or kerosene or rice to
- lend money to
- obtain medical advice from
- engage socially with
- are related to
- go to temple with
- invite to one's home
- visit in another's home

With each one of these questions/relationships, it is possible to construct a graph. Table 4.10 shows the size and sparcity of each graph for village no 1.

Table 4.10: Graph properties for each relationship - Village no 1

Relationship	Nodes	Edges	No. Connected Components
Borrow money from	838	2149	40
Give advice to	838	2010	65
Help with a decision	837	2067	64
Borrow kerosene or rice from	836	2219	41
Lend or kerosene or rice to	837	2213	40
Lend money to	838	2125	45
Obtain medical advice from	836	2045	75
Engage socially with	838	2296	28
Are related to	838	2288	37
Go to temple with	835	1872	153
Invite to one's home with	837	2216	37
Visit in another's home	837	2252	34

For example, relationships that involve a higher degree of intimacy such as 'Go to temple with', 'Give and ask for advice including medical', tend to occur in a more closed

circle (within same household), therefore, the number of connected components are larger. On the other hand, relationships that involve some degree of social cooperation (lend or borrow money, kerosene, rice and social invitations), transcend the closed circle of household, and therefore increases the level of connectivity.

Figure 4.6 illustrates visually the connectivity and sparcity looking from the **Lend money to** relationship. Normally for all the relationships, the resulting graph has lots of small clusters and one giant connected component. In practice, there are several scattered households or families, and eventually a larger community of linked households. This big connected component is the object of study to explore the communities and interelationships within it.

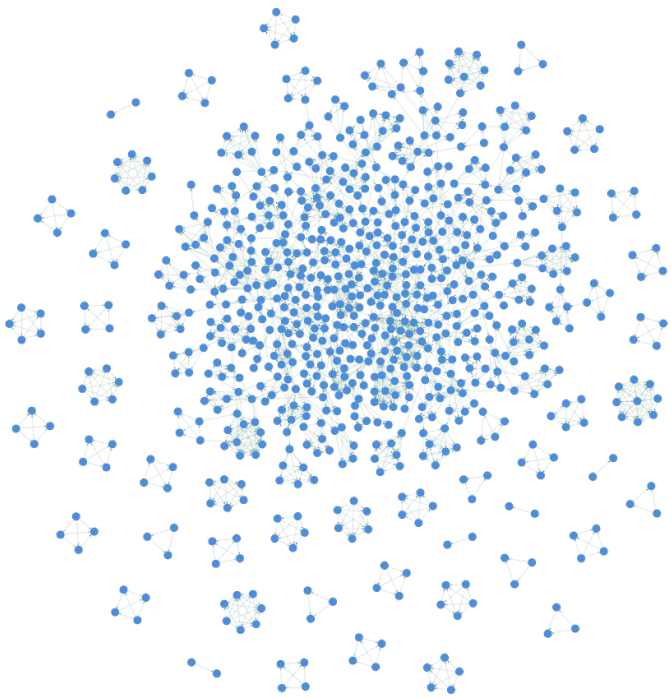


Figure 4.6: Graph of"'Lend Money To" relationships - village 1

In Table 4.11, we can see some metrics that help us to understand how the community is structured looking at its largest connected component. Overall, all relationships exhibit characteristics expected by a social network: an average shortest path in the range of 6 (Small World 6-degrees of separation) and a high level of clustering.

Table 4.11: Graph properties of the biggest connected component of each relationship

Relationship	Diameter	Avg. Shortest Path	Avg. Clustering
Borrow money from	13	6.6	0.8
Give advice to	19	8.0	0.8
Help with a decision	17	6.5	0.8
Borrow kerosene or rice from	19	6.8	0.8
Lend or kerosene or rice to	14	6.6	0.7
Lend money to	18	7.1	0.8
Obtain medical advice from	16	7.3	0.8
Engage socially with	13	5.7	6.3
Are related to	13	6.3	0.8
Go to temple with	7	4.0	0.9
Invite to one's home with	12	6.0	0.7
Visit in another's home	15	5.8	0.7

One of the reasons that this database was chosen, is that it captures the "credit relationship" by the properties **lendmoney** and **borrow money from**. If a given person express likelihood to **lend money** to another one, it is assumed that this person might have some information or some degree of **trust** to make the assertion. Therefore, for the purposes of this research, we will use the relationship "lendmoney" for the analysis of the graph structure and application of trust algorithms to find the most trusted nodes.

4.2.1 "Analysis of "Lend Money To" relationship

Initially, we would like to understand some characteristics of the graph of this relationship, to check if it fulfills some properties of a typical social network.

As we could see from Figure 4.6 the overall structure of the **lendmoney** graph is composed of a giant component and some scattered small clusters. When we zoom out the giant component on Table 4.11, we find it to have a very high average clustering of 0.8, a diameter of 18 and an average shortest path of 7.1.

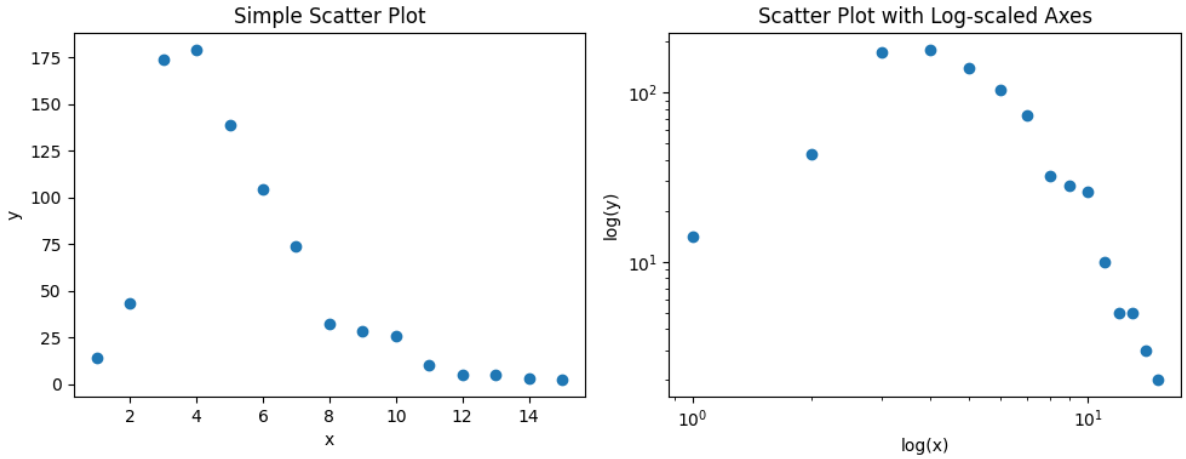


Figure 4.7: Degree distribution "Lend Money To" relationship - village 1

In fact, when we look at the degree Figure 4.7 and shortest path Figure 4.8 distributions, we can see that this graph does fit into a typical social network, exhibiting characteristics of a "small World". In general nodes are quite connected within 4 to 6 nodes (which may suggest the average size of a typical household) and few nodes with very high degree (8 or more). This arrangement may produce a very interconnected network Figure 4.9, especially in a village, where the degree of separation (shortest path) between every node follows closely the same shape of Milgrom "Small World" experiment, exhibiting a 7 degrees of separation.

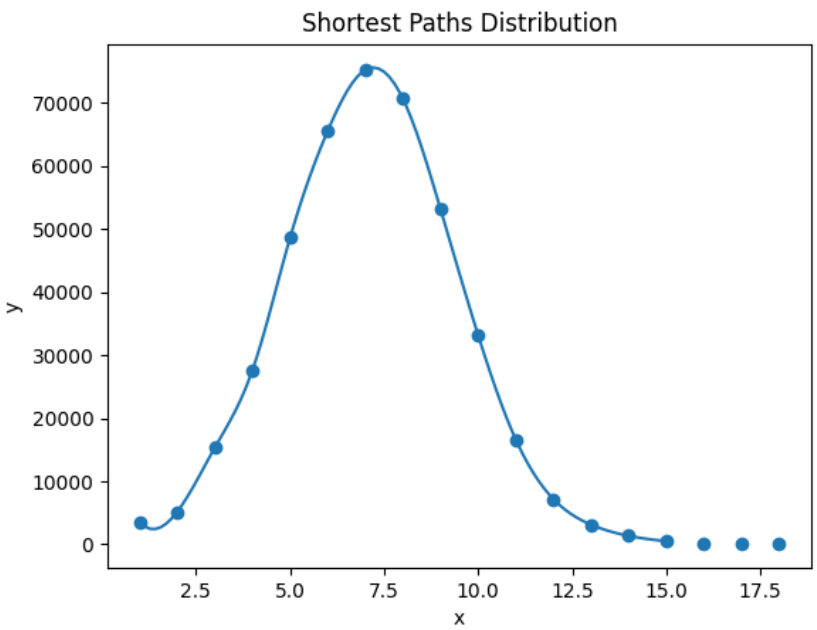


Figure 4.8: Shortest Paths distribution "Lend Money To" relationship in the giant component - village 1

Figure 4.9: Clustering distribution "Lend Money To" relationship in the giant component - village 1

The dataset offers very few socio-economic indicators, however when we look at the distribution of ages of the people who lend and that of the people who is recommended (potential borrowers), we see a statistic significant difference. Sample t-test using scipy [59] reveals a $pvalue = 3.347e - 05$.

Figure 4.10 suggests that in this village, lenders tend to be older than potential borrowers. Table 4.12 shows average age of lenders as 46 years old while age of potential borrowers as 33 years old. This insight tend to make sense, if we imagine that in rural villages normally older people have more wealth than younger people, and therefore they are the ones more likely to finance loans to younger ones, which may be their sons, siblings or young brothers.

Having made sure the graph follows a typical social network structure, allows us to start applying trust algorithms and draw some conclusions based on the social context discussed in section 2 with more confidence.

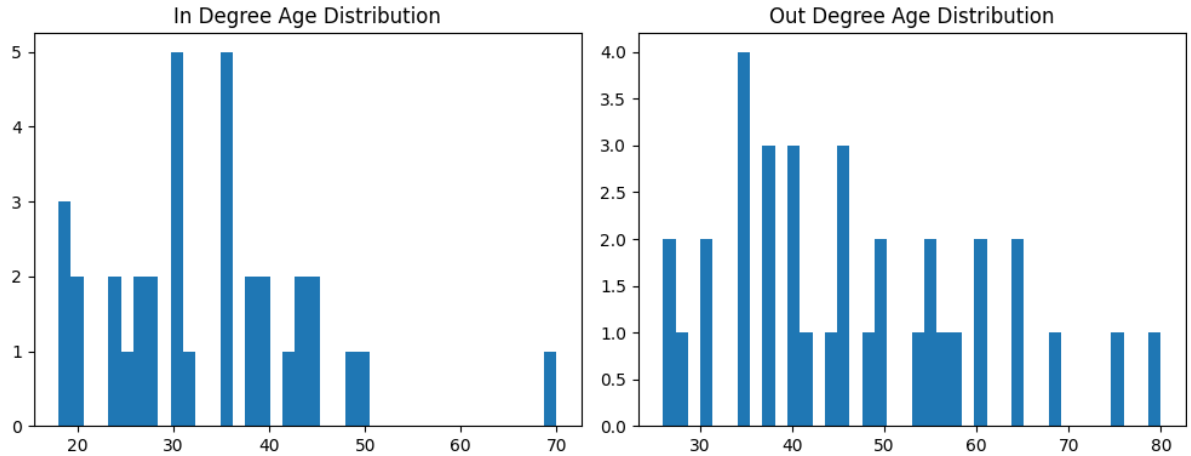


Figure 4.10: Age distributions "Lend Money To" relationship - village 1

Table 4.12: Descriptive Statistics for In_Degree and Out_Degree ages

statistic	In_Degree	Out_Degree
min	18	26
max	70	80
mean	33.5	46.7
variance	119	191

Finding trustworthy nodes

In this multitude of nodes, eigentrust was used to spot trustworthy nodes, those nodes that are more likely to be a target to "lend money to".

In order to use the eigentrust algorithm, a basic assumption is to assign a positive local trust $tr(i, j) = +1$ for every relationship "lend money to" between two nodes i and j. The concept of "lend money" here should be understood as a propensity to lend money to a given person, it does not mean necessarily that money has been lent. Therefore, if one may say that would lend money to somebody, we will assume that there must be some kind of trust between them. We could of course, assign a qualitative measure of trust, for example in a given scale from 1 to 5. Or, as in the original paper suggests, assign a post-fact evaluation of the transaction, i.e., if the borrower paid the loan on time or not. Unfortunately, there are no such information in the original database, therefore assuming a local trust $tr(i, j) = +1$ is reasonable. This simplify the results, because if every relationship has an equal local trust of 1, then the overall calculation will be a consequence only due to the structure of the graph.

Running the eigentrust algorithm with the above assumption, we calculated the trust score for all nodes. Table 4.13 shows the top 10 nodes ranked by eigentrust scores as well as its in_degree, out_degree and clustering coeficients.

Table 4.13: Top 10 eigentrust for "Lend Money to"

pid	eigentrust value	in_degree	out_degree	clustering
109805	1.16e-04	5	0	0.6
118005	3.52e-05	4	0	1.0
116312	2.52e-05	11	0	1.0
109804	2.46e-05	4	1	0.6
106307	1.16e-05	6	0	1.0
118004	1.03e-05	3	1	1.0
111008	9.10e-06	7	0	1.0
109803	8.10e-06	2	2	1.0
109507	7.49e-06	6	0	1.0
116311	5.50e-06	12	1	0.73

We can see that eigentrust values are not driven neither by in_degree neither or by high clustering coeficients alone.

For a better understanding of the reasons for high scores, it is necessary to understand the context of the network that the node is. Figure 4.11 illustrates the structure of the network around node 109805 by a maximum path length of 6-nodes away.

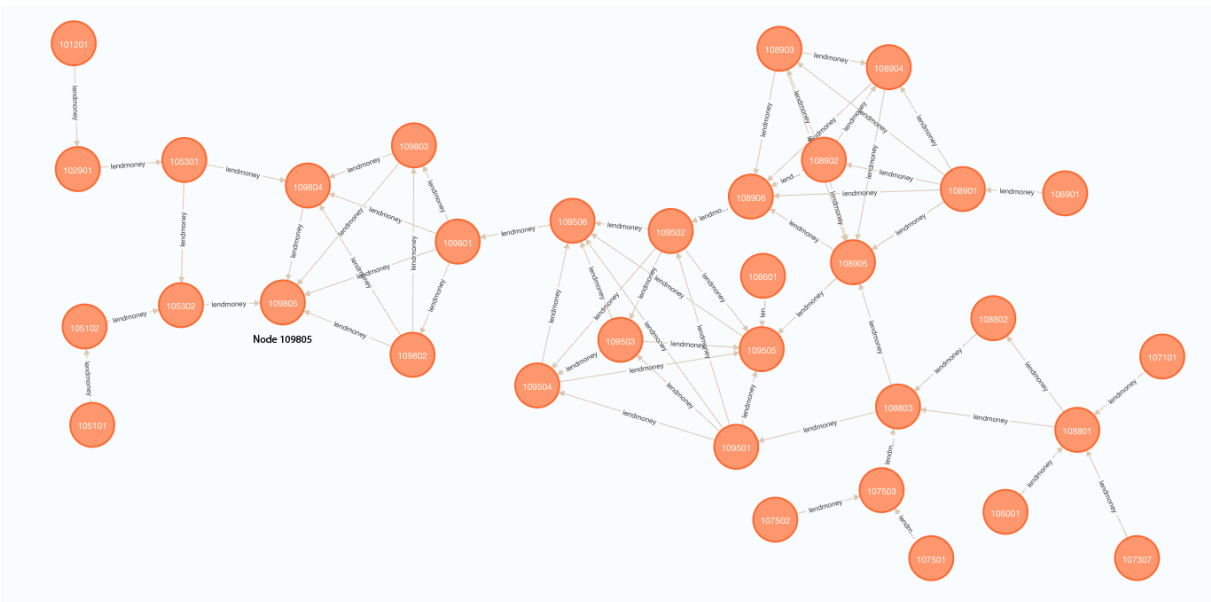


Figure 4.11: Neighborhood structure of node 109805

We can see that node 109805 reputation comes from two streams of endorsements, one coming from 4 peers of its tightly-knit group, and another from node 106302, which bridges to other two different nodes (105102 and 105301) that recommends it.

Moreover, there is one particular node 109801 from the node 10980's tightly-knit group, who has a bridge to node 109506, which is a highly recommended node of another tightly-knit group. In this case, we can see a clear pattern of concatenation of a highly recommended member of a tightly-knit group bridging to other recommended member of other tightly-knit groups and so on successively. In other words, the high engentrust scores is a result of the effect of a chain of nodes that has one of the two kinds of social capital, either a bonding type or a bridging type.

This result shows a certain robustsness to eigentrust, because it dilutes the bias of endorsements coming solely from peers from its tightly-knit group alone. It gives importance to sources of trust coming from other parts of the social network. Node 109805 is not the one that has the highest in_degree and maximum clustering coefficient of 1.

Chapter 5

Discussion

During the course of the research, it was possible to gather real life cases, whose analysis and conclusions supported many of the theoretical concepts seen in the previous sections. Nevertheless, it is essential to recognize the potential limitations and challenges of the research in using such datasets, as the sample size is very small, and therefore we need to be careful in generalization of findings given the complexity of trust dynamics.

In addition, there were limitations also in the breadth of the data, meaning that the dataset lacked more personal information that could help a better characterization of the nodes. It also lacked a evaluation of of every node about the strength of their respective personal relationships with other nodes, either in a form of a **pre-trust** value or a **post-transaction score**. These limitations, as already eluded, guided the selection of simpler and more basic algorithms, eigentrust and tidaltrust. It also due to the fact that such information is private and confidential. In fact, people in closed-knit groups such as family don't feel confortable making a relative rating of their peers, as it is a sensitive information that if not well managed, can potentially erode the relationship. Therefore, there is an ethical consideration that needs to take into consideration in making experiments in this domain, and also with consequences of the design of a framework for determining trustworthiness using social capital.

In the three experiments, ROSCA, peer-to-peer loans and the Indian village trust analysis, in order to identify sources and kinds of social capital, we had to consider not only the graph metrics of centrality or clustering or scores from the algorithms, but also the social context of the people represented in the network. And the social context is gathered by interviewing people and understand the social dynamics. As also eluded earlier, this situation corroborate once more on the fact that social network sits in an interface of computer science, sociology and economy, and therefore neither of the three domains alone is enough to characterize what is happening.

Group or community structure and the personal connections are very important to provide insights about the outcome of the economic transactions that took place. There is a trust dynamics that takes place between the individuals, and it seems that this dynamics

is specific in the case of the ROSCA and in the case of the peer-to-peer loan.

One of the first questions to start this discussion could be how these two arrangements (ROSCA and peer-to-peer) worked out fine, i.e, participants stuck until the end.

Starting from the ROSCA, the economic objectives of this arrangement is to have people commited to save money together during some months, so in each month, one of the participants can take a lump sum. In order to this arrangement succeed until the last round, is the fact that people who had received in earlier months do not cheat, by leaving the group earlier. Because there is no law enforcement involved, the only incentive for people to stay tight is to maintain their reputation. The consequence of let the group down is of a social shame.

So, when we look at the graph of the ROSCA in Figure 4.2, we saw a very tightly-knit social arrangement. The average clustering coefficient is high, and in fact, there is a core sub group in where almost every person is connected to each other.

The leader is at the center of gravity of the group, in this case having a high closeness centrality of 0.79, he or she is well positioned to monitor the behavior of everyone else, and keep track of the history. A good participation, give to each participant the chance to be invited again in subsequent ROSCAs, and so the recurrence becomes also a metric of reputation and consistency.

In successful ROSCAs, it seems that the leader has a position of respect by all, earned by its long standing or close relationship with all. This is supported in by having the second highest eigentrust score of 0.202. From the conversations of some members of the ROSCA, they regard the group leader as generally to be a more senior person in the sense that everybody respect him/her, or alternatively a more proficient person in the sense of having more financial literacy, organization and coordination. According to them, having a highly respectable and proficient leader is one of the key success factors. The leader can in the worse case, by identifying some potential menace, replace a person, or even assume his position and commitments, to keep the group running until the end. In this particular group that we studied, the only event that the leader recalled that a ROSCA broke, was due to the leader of that case being responsible for the breakout. One of the reasons for the breakout was his lack of track record in leading groups and volunterism to assume the role in that particular one, as he secretely had personal financial problems that later on became public.

Next, in the peer-to-peer (P2P) experiments, the economic objective is to enable a person to borrow an amount of money significantly higher than the amount the same person could get from a ROSCA round. In the cases seen in this research, the P2P loan is an extension of a loan among family and friends to other people (lenders) who are not necessarily from the tightly-knit group of the borrower. In fact, as we have seem from Figure 4.3 and Figure 4.5, these extra lenders have no connection whatsoever with the borrower.

Here it is noticed another sociologic concept applied to social networks which is *homophily*, which describes the tendency of individuals to associate and bond with similar others, as in the proverb *"birds of a feather flock together"*. It appears that the structure of the peer-to-peer experiments enabled the connection between to different group of individuals, in this case groups associated with different social and economic class. Table 4.7 show that the amount funded by the group of lenders that belong to the other group other than the borrower's, was significantly higher.

Homophily works very well in the case of ROSCA, giving a common sense of belonging, in this case, a common family heritage and same social and economic environment. However, in the case of peer-to-peer, it worked partially. In order for the loan amount to achieve a higher amount, it was necessary to complement it through participation of a different kind of lenders. A lender that had more wealth than the people from the borrower's social circle, and who could see the peer-to-peer loan as an investment.

Investing in a pool of money to lend to somebody else, especially to a person that the lenders do not know, it is a risky investment. However, it seemed to worked out, because in the group's structure, nodes 2 and 5 seem to have considerable *social capital*, as also supported by their high betweeness and closeness centrality. The reputation and trust measured by the weight 6 of the relationship between nodes 2 and 5, and their respective reputaion in their respective groups, helped to reduce the risk to a point that make the loan possible.

In addition, from the eyes of these other investors, having seen that people from the same group of the borrowers also participated in the pool, was a good sign. The fact that borrower's closest friends put their "skin in the game", could have also contributed to give more reassurance, providing less incentive to the borrower to default the loan.

In this sense, the tidaltrust model seemed to have captured well these trust transitivity from high trusted nodes such nodes 2 and 5, which were passed to other distant nodes, such as node 7.

Contrasting the two experiments and the algorithms used, it seems that eigentrust helps to identify high trusted nodes, that in the perspective of the credit, could be ideal candidates for:

- be a creditworthy borrower
- be a recommender for potential borrowers

Tidaltrust helped to identify and quantify trust path between a source (investor) and a sink (borrower). In the experiments we saw that in those paths there were high trust nodes that gave high recommendations, which were passed on, therefore resulting in a high score path. The absence of high trusted nodes, would have perhaps resulted in higher uncertainty for the investor which could result in not be willing to participate in

the funding. This quantitative metric seem to support very well the correlation between P2P lending outcomes and friendship discussed in section 2.1.2.

The experiments ROSCA and P2P were good because we could analyze them from the perspective of the network structure, application of trust algorithms and social dynamics. It helped us to understand the role of some nodes, identify their type of social capital and insights that support the succesful outcomes.

In the experiment of village1 of the "The Diffusion of Microfinance" article, on the other hand, we simply got a social network databset composed by relationship types such as "lend money to". Recall, that this information comes from a questionnaire, in which people were asked to appoint the likelihood to "lend money" to people from his/her social circle. Therefore, we assumed it as an intention to lend money, not necessarily a register that money has been lent in the past. In addition, no relative weight in the "lend money" relationship were provided, all have equal weights. Despite the limitations, we suspect that these kind of dataset resembles what would be available in real life.

So a question to explore is how a financial institution could use this dataset together with the insights of the other two experiments, identify potential candidates from the trust perspective, to receive loan offers?

Here again we chose to use eigentrust algorithm because it is robust for large networks and relatively simple and we can apply with the assumption of equal weights for relationships. In this way, eigentrust scores are reduced to the structure of the graph, more specifically to the node position in the graph. The highest eigentrust scores rewards nodes that exhibit one of the two kinds of social capital, bonding or bridging, and are surrounded by nodes with same pattern. Figure 4.11 shows the position of node 109805.

This result lead us to conclude that at the earliest, position of the node in the graph which may characterize it having one of the two types os social capital is very fundamental. Algorithms such as eigentrust and its derivations (socialtrust, etc) adds up a second layer that takes into consideration other factors such as past opinions about peers, pre trusted scores, affinity relations, etc. These factors may enhance or degrade the trust score and are important depending on the social context.

However, in real life, especially in traditional financial institutions, there is no mapping of relationship of clients in a form of a graph, let alone, evaluations or opinions between clients regarding trust. In order to come with such information, it would certainly touch sensitive and confidential information. In the old days, especially in small towns, we could recognize the role of a bank branch manager, a person who knew and had relationships with citizens, and therefore he or she would have first hand social information that was used to approve or deny loans or capital limits. Nowadays, there is barely no human interaction in banks, and therefore social information has disappeared.

A possible institution that congregates people in a social network with financing purposes are cooperatives. Finance cooperatives like Sicoob or Sicredi are member-owned

and operated institutions. Members, who are also customers, own shares and have a say in the governance of the cooperative. These institutions primarily serve specific communities or groups sharing common interests, ensuring that financial services are tailored to the needs of their members. Potentially, they could build a social graph, that represent the relationships and interactions among members, highlighting key connectors, influential members, and community clusters. Using social network analysis and integrating with trust and reputation algorithms, they could then identify patterns and strengths of relationships, predict trustworthiness, and assess creditworthiness based on social capital within the network. Nevertheless, data collection would still be an issue as they should comply with privacy laws and ethical standards and members should consent to the use of their data for creating this social graph.

Fintechs are promising institutions that can use forms of social capital in their business models. In fact, Nubank, a leading fintech in the credit card business, in its initial years, only accepted applications of new clients through recommendation of existing clients, and quite surely it would have used the quality of the existing client to weight its recommendation. This led to a relatively smaller delinquency rates compared to the incubents.

Fintechs are normally very niche financial platforms or a marketplace of innovative financial products. Given their disruptive business models and cutting edge technology, they could test some ideas of this work.

For example, the concept of seeding a marketplace with pre-existing social relationships presents a compelling approach. In this model, the marketplace begins with a core group of individuals who share prior connections (for example from a ROSCA), possibly extending from real-world interactions. New members join through referrals, harnessing the power of existing trust bonds. This approach naturally creates a sense of community and trust, potentially leading to lower default rates as members might feel a stronger sense of accountability towards individuals within their network. However, this model faces challenges in scalability and diversity. Growing the network could be slower, and there's a risk of creating an environment with limited diversity in risk profiles.

Alternatively, an open marketplace model invites anyone to participate, relying on the platform's mechanisms to build and foster trust. This model leverages in its ability to scale rapidly and incorporate a diverse range of borrower and lender profiles. The key challenge here lies in establishing initial trust levels, given the absence of pre-existing relationships. Such a marketplace would need robust systems for identity verification, transparent transaction histories, and a reliable user rating system to mitigate the risks of defaults and fraudulent activities.

Finally, a hybrid model, combining elements of both approaches, could potentially offer the best of both worlds. This model allows for the organic growth of the platform through new memberships while also encouraging and rewarding referrals and network building within the platform. It endeavors to balance scalability with trust-building,

albeit requiring sophisticated algorithms to manage this duality effectively.

In terms of performance evaluation, developing metrics to quantify social capital becomes crucial, looking at the strength and number of connections, as well as referrals and endorsements. User engagement and the quality of interactions would offer insights into the platform's social dynamics, while trust metrics derived from surveys and behavioral data could gauge the trust levels within the platform. Balancing social data with traditional credit assessment methods is essential. Also, vigilance against the manipulation is necessary, ensuring that no single connection or referral unduly influences the platform's dynamics.

This point is very important, because while the empirical examples presented in this dissertation had succesfull outcomes, users have reported examples of failures in the past, whose lessons learned led them to adopt small changes or take extra care with some nuances. For example, when a member of the ROSCA is under a very financial stress situation such as high indebtness or unemployment. Another example is when no assessment of the creditor's cash flow is made, so he or she may face difficulties to save money given the level of household expenses. And lastly, from a social perspective, allowing new members, with no much stock of social capital within the group to take risk positions, for example borrowing large amounts or be allowed to take the grants of the ROSCA in earlier rounds.

Another possible are of research could to address the challenge the social risk. In social arrangements, there is the possibility of collusion, where people may not tell the truth in order to gain advantages as a group. Or they may not be honest when asked to evaluate someone in their social network, either out of fear of retaliation, shame, or cultural factors. To address this, an area of research could employ advanced analytical tools capable of detecting patterns indicative of collusion. Machine learning algorithms could be employed to analyze transaction data for anomalies that suggest coordinated behavior. Moreover, implementing strict verification processes for new members and monitoring the frequency and nature of interactions between members could help in identifying suspicious activities.

Chapter 6

Conclusion

In this research we posed the question of whether social capital could enhance credit scoring and through usage of computer science methods and algorithms, we could propose a theoretical method to accomplish this task.

We have got inspiration from informal lending arrangements that takes place in communities throughout the World for centuries. In this informal arrangements, there are financial transactions such as loans that rely on trust and reputation between lenders and borrowers. In particular we looked at two informal arrangements: ROSCAs (Rotating Savings and Credit Associations) and peer-to-peer loans.

Peer-to-peer loans institutions such as Prosper, Lending Club, both in the USA, were extensively studied by the academic community. In some of those papers, there are evidences that peer-to-peer loans when there are friends involved, perform better than when all participants are anonimous.

Social Capital is a area of study that has roots in economy, sociology and more recently in computer science, especially after the advent of social networks. There are, indeed, many definitions of social capital, however in this research we have adopted the definiton that comes from any economic benefit derived through the result of social relationships that are possible because there is trust. Social networks are represented by graphs, and therefore the concepts and metrics used to describe graphs helps to understand social capital in social contexts. In the literature, we can identify two types of social capital: one that comes from bonding like those of highly knit groups, in where members have a close relationship between them and a common sense of trust; another one that comes from broking, normaly a very trustworthy person who bridges two distinct groups of people and make economic transactions possible.

Trust and Reputation Systems is also a emerging field that has been applied in a variety of applications such as e-commerce and p2p networks. Since them several algorithms have been introduced, in this research we chose to use eigentrust and tidaltrust.

As a use case to apply the knowledge and to help to answer the research posed question, we have gathered real cases of ROSCA and peer-to-peer loans. In those cases we could

verify that in a ROSCA prevails the bonding type of social capital, while in P2P loan a mix of bonding and bridging. In particular, the bridging type in P2P is key to find more wealth investors and fund the loan. Centrality metrics are useful to identify important nodes and correlate with the social context. In terms of the algorithms, eigentrust is well suited to calculate trust of all members, and in the case of the ROSCA, the group leaders were the ones of highest score. Tidaltrust is more well suited to P2P loan, because it helps to calculate trust between two nodes, more especifically we were interested in calculate the trust in the path between the most distant investor and the borrower. In our case, the score was high, because there were very high trusted nodes along the path, and this is perhaps the reason why the loan was funded.

Although these experiments supported the theoretical concepts and applicability of the algorithms, we aknowledged the fact that the sample was small therefore we need to take care of generalizing the results. Both ROSCAs and P2P in the experiments had a happy outcome, however there is still possibility that in certain cases, and due to many reasons, the loan could not be repaid or the ROSCA broken.

In addition to the experiments, we have also gathered a public database from the article the Diffusion of Microfinance. The idea was to test how some of the approaches can be used in a real and scalable network. We have picked a social network of a village in india that mapped the intention or likelihood to lend money to its friends or connections. Again, we have chosen to use eigentrust, which generated a ranking score of the highest trusted people in this village, that we could potentially use as a short list of high trust borrowers.

All in all, we are convinced that analyzing the network structure using the centrality metrics and the usage of trust algorithms can reveal trusted nodes that either be consider candidates for loan offers, or they could be good recommenders. This can constitute a framework to complement with traditional scoring systems, and in the aggregate provide a better assessment of risk.

The only caveat, is that the usage of such framework implies that we have to have the social network mapped, and unfortunately this mapping is still very rare in traditional financial institutions and banks. In these institutions there are privacy and ethical concerns when modeling trust using social networks.

We are yet to see also other forms of arrangements that overcome the privacy and ethical concerns, among those blockchain looks promising. Blockchain platforms such as Etherium advocates the usage of encryption to collect social information anonymously, and through smart contracts could incorporate algorithms like the ones we use in this research to generate a score, in which an individual could opt to make it available to financial institutions.

Bibliography

[1] Li Xiong 0001 and Ling Liu 0001. "PeerTrust: Supporting Reputation-Based Trust for Peer-to-Peer Electronic Communities". In: *IEEE Trans. Knowl. Data Eng* 16.7 (2004), pp. 843–857. URL: http://doi.ieeecomputersociety.org/10.1109/TKDE.2004.1318566.

[2] S. Agreste et al. "Trust networks: topology, dynamics, and measurements". In: (June 2015), pp. 26–35.

[3] Carlos Sanchís-Pedregosa Antonio Moreno-Moreno Emma Berenguer. "Reputation Systems". In: *Economics and Sociology* 11.4 (2018), pp. 19–79. DOI: 10.14254/2071-789X.2018/11-4/4.

[4] Abhijit Banerjee et al. *The Diffusion of Microfinance*. Version V9. Harvard Dataverse, 2013. DOI: 10.7910/DVN/U3BIHX. URL: https://doi.org/10.7910/DVN/U3BIHX.

[5] Gopal Bao and Mohlemann. "Throwing Good Money After Bad: Risk Mitigation Strategies in the P2P Lending Platforms". In: *Information Systems Frontiers* (2018).

[6] T Bastelaer. "Does Social Capital Facilitate The Poors' Access To Credit? A Review Of The Microeconomic Literature. Word Bank Working Paper No. 8". In: (Jan. 2002).

[7] T. Besley, S. Coate, and G. Loury. *The Economics Of Rotating Savings And Credit Associations*. Working papers 556. Massachusetts Institute of Technology (MIT), Department of Economics, 1990. URL: https://ideas.repec.org/p/mit/worpap/556.html.

[8] Stephen P. Borgatti, Candance Jones, and Martin G. Everett. "Network measures of social capital". In: *Connections* (1998).

[9] Ronald S. Burt. "Structural holes and good ideas". In: *American Journal of Sociology* (2004).

[10] Ronald S. Burt. *Structural Holes: The Social Structure of Competition*. Harvard University Press, 1992.

[11] Xiangru Chen, Lina Zhou, and Difang Wan. "Group social capital and lending outcomes in the financial credit market: An empirical study of online peer-to-peer lending". In: *Electronic Commerce Research and Applications* 15 (2016), pp. 1–13.

[12] J.S. Coleman. "The Foundation of Social Theory". In: (Jan. 2000).

[13] James S Coleman. *Foundations of Social Theory*. Harvard University Press, 1990.

[14] James S Coleman. "Social capital in the creation of human capital". In: *American Journal of Sociology* (1988).

[15] James S Coleman. *The Adolescence Society*. Free Press, 1961.

[16] Dupont and Karpoff. "The Trust Triangle: Laws, Reputation, and Culture in Empirical Finance Research". In: *Journal of Business Ethics* (2018).

[17] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cmabridge University Press, 2010.

[18] *Economics, organization, and management / Paul Milgrom, John Roberts*. English. Prentice-Hall. ISBN: 0132239671.

[19] Jose Scheinkman Edward Glaeser. "Measuring Trust". In: *The Quarterly Journal of Economics* 115.3 (2000), pp. 811–846.

[20] Michel Ferrary. "Confiance et accumulation de capital social dans la régulation des activités de crédit". fre. In: (1999). ISSN: 0035-2969. DOI: 10.2307/3322828. URL: https://www.persee.fr/doc/rfsoc_0035-2969_1999_num_40_3_5189.

[21] Seth Freedman and Ginger Zhe Jin. *The Information Value of Online Social Networks: Lessons from Peer-to-Peer Lending*. NBER Working Papers 19820. National Bureau of Economic Research, Inc, 2014. URL: https://EconPapers.repec.org/RePEc:nbr:nberwo:19820.

[22] Noah E. Friedkin and Eugene C. Johnsen. *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*. Structural Analysis in the Social Sciences. Cambridge University Press, 2011. DOI: 10.1017/CBO9780511976735.

[23] Hongming Gao. "Prediction Re-loan Success based on the Friendship Network Characteristics in the Online Micro-loan Marketplace". In: *Twenty-fourth Americas Conference on Information Systems, New Orleans* (2018).

[24] Anurag Garg and Roberto Battiti. *Digital Reputation for Virtual Communities*. Tech. rep. Oct. 2005.

[25] Jennifer Golbeck. "Computing and Applying Trust in Web-Based Social Networks". In: (Jan. 2005).

[26] Jennifer Golbeck, Bijan Parsia, and James Hendler. "Trust Networks on the Semantic Web". In: *Cooperative Information Agents VII*. Ed. by Matthias Klusch et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 238–249. ISBN: 978-3-540-45217-1.

[27] Mark Granovetter. "Economic Action and Social Structure: The Problem of Embeddedness". In: *American Journal of Sociology* 91 (Nov. 1985), pp. 481–510. DOI: 10.1086/228311.

[28] Mark Granovetter. "Economic action and social structure: The problem of embeddedness". In: *American Journal of Sociology* (1995).

[29] Mark Granovetter. "Getting a Job: A Study of Contacts and Careers". In: *Chicago Press* (1974).

[30] Mark Granovetter. "The Strength of Weak Ties". In: *The American Journal of Sociology* 78 (May 1973), pp. 1360–1380. DOI: 10.1086/225469.

[31] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. "Exploring Network Structure, Dynamics, and Function using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Ed. by Gaël Varoquaux, Travis Vaught, and Jarrod Millman. Pasadena, CA USA, 2008, pp. 11–15.

[32] Li-An Zhou Hanming Fang Rongzhu Ke. "Rosca Meets Formal Credit Market". In: *NBER Working Paper* 21863 (2015).

[33] Jun-Fei Wan Hwang Ting-Kai Yung-Ming Li. "A Social Recommendation Mechanism for Peer-to-Peer Lending". In: *Twenty-fourth Americas Conference on Information Systems, New Orleans* (2018).

[34] Yoichi Izumida. "The Kou in Japan: A Precursor of Modern Finance". In: *The Ohio State University and The Agency for International Development – The World Bank* (1989).

[35] Mathew Jackson. *Social and Economic Networks*. Princeton University Press, 2008.

[36] Wenjun Jiang et al. "Understanding Graph-Based Trust Evaluation in Online Social Networks: Methodologies and Challenges". In: *ACM Computing Surveys* 49.1 (July 2016), 10:1–10:?? ISSN: 0360-0300 (print), 1557-7341 (electronic). DOI: https://doi.org/10.1145/2906151.

[37] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. "The Eigentrust algorithm for reputation management in P2P networks". In: *WWW*. Ed. by Gusztáv Hencsey et al. ACM, 2003, pp. 640–651. ISBN: 1-58113-680-3.

[38] Dean Karlan. "Social Connections and Group Banking". In: *The Economic Journal* 117.517 (2007), F52–F84.

[39] De Liu et al. "Friendships in Online Peer-to-Peer Lending: Pipes, Prisms, and Relational Herding". In: *MIS Quarterly* 39 (Sept. 2015), pp. 729–742. DOI: 10.2139/ssrn.2251155.

[40] Amy Holder Mark Needham. *Graph Algorithms - Practical Examples in Apache Spark and Neo4j*. O'Reilly, 2019.

[41] Stanley Milgram. "The Small World Problem". In: *Psichology Today* (1967).

[42] Siva Viswanathan Mingfeng Lin Nagpurnanand Prabhala. "Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending". In: *Western Finance Association* (2014).

[43] Gardiner Morse. "The Science Behind Six Degrees". In: *Harvard Business Review* (Feb. 2003).

[44] Mark E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010, pp. 1–740. ISBN: 978-0-19920665-0.

[45] Mark E. J. Newman. *The structure and function of complex networks*. SIAM Review, 1993.

[46] Alejandro Portes. "Social capital: Its origins and applications in modern sociology". In: *Annual Review of Sociology* (1998).

[47] Josep M. Pujol, Ramon Sangüesa, and Jordi Delgado. "Extracting reputation in multi agent systems by means of social network topology". In: *AAMAS*. ACM, 2002, pp. 467–474.

[48] Robert Putnam. "The Prosperous Community: Social Capital and Public Life". In: *Am. Prospect.* 13 (Jan. 1997).

[49] Paul Resnick and Hal R. Varian. "Recommender Systems". In: *Commun. ACM* 40.3 (Mar. 1997), pp. 56–58. ISSN: 0001-0782. DOI: 10.1145/245108.245121. URL: http://doi.acm.org/10.1145/245108.245121.

[50] Ricardo Abramovay Rodrigo Gravina Prates Junqueira. "A sustentabilidade das microfinanças solidárias". In: *Revista de Administração* 40.1 (2005), pp. 19–33.

[51] Jordi Sabater and Carles Sierra. "Reputation and social network analysis in multi-agent systems". In: *AAMAS*. ACM, 2002, pp. 475–482.

[52] John Scott. *Social network analysis: a handbook*. SAGE Publications, 2011.

[53] Wanita Sherchan, Surya Nepal, and Cécile Paris. "A survey of trust in social networks. ACM Comput". In: *ACM Computing Surveys* 45 (Aug. 2013), pp. 1–33. DOI: 10.1145/2501654.2501661.

[54] Dyah Sukmaningsih. "A Model for Lender-Borrower Trust in Peer-To-Peer Lending". In: *ComTech: Computer, Mathematics and Engineering Applications* 9 (July 2018), p. 15. DOI: 10.21512/comtech.v9i1.4287.

[55] Tianhui Tan and Tuan Phan. "Social Media-Driven Credit Scoring: the Predictive Value of Social Structures". In: *ICIS*. Ed. by Pär J. Ågerfalk, Natalia Levina, and Sia Siew Kien. Association for Information Systems, 2016. ISBN: 978-0-9966831-3-5. URL: http://aisel.aisnet.org/icis2016/.

[56] Mozhgan Tavakolifard and Kevin Almeroth. "A Taxonomy to Express Open Challenges in Trust and Reputation Systems". In: *Journal of Communications* 7 (July 2012), pp. 538–551. DOI: 10.4304/jcm.7.7.538-551.

[57] "The Initiative on Defining Monitoring and Measuring Social Capital – Overview and Program Description". In: *Social Capital Initiative* Working Paper No.1 (1998).

[58] Santiago Varella, Silvana Parente, and Conceição Faheina. "The CrediAmigo program: urban productive inclusion through microcredit". In: (Dec. 2017).

[59] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[60] Duncan J. Watts and Steven H. Strogatz. "Collective dynamics of 'small-world' networks". In: *Nature* (1998).

[61] Yanhao Wei et al. "Credit Scoring with Social Network Data". In: *SSRN Electronic Journal* 35 (Jan. 2014). DOI: 10.2139/ssrn.2475265.

[62] Haewon Yum, Byungtae Lee, and Myungsin Chae. "From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms". In: *Electronic Commerce Research and Applications* 11.5 (2012), pp. 469–483.