

TI 2011-072/1  
Tinbergen Institute Discussion Paper



# Risk Aversion and Social Networks

*Jaromir Kovarik<sup>1</sup>*

*Marco J. van der Leij<sup>2</sup>*

*<sup>1</sup> University of the Basque Country;*

*<sup>2</sup> Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

# Risk aversion and social networks\*

Jaromír Kovářik<sup>†</sup>

Marco J. van der Leij<sup>‡</sup>

26th April 2011

## Abstract

Agents involved in the formation of a social or economic network typically face uncertainty about the benefits of creating a link. However, the interplay of such uncertainty and risk attitudes has been neglected in the network formation literature. We propose a dynamic network formation model that builds on standard microeconomic concepts of utility maximization, incomplete information, and risk aversion. The model predicts that an agent's risk aversion is correlated with her network clustering coefficients, but not with her degree. We discover a mechanism that generates a correlation between network position and payoffs of individuals. Moreover, we show how the generated network architecture depends on the uncertainty in the environment it is embedded in.

Keywords: Network formation, risk aversion, clustering coefficient, local/global search

---

\*We thank Michèle Belot, Pablo Brañas, Arthur Campbell, Pascal Courty, María Paz Espinosa, Marcel Fafchamps, James Fowler, Edoardo Gallo, Jayant Ganguli, Sanjeev Goyal, Matthew Jackson, Friederike Mengel, Paolo Pin, Andrea Prat, Giorgio Topa, Fernando Vega-Redondo, Adrien Vigier, Andrew Zeitlin and audiences at the European University Institute, University of Amsterdam, University of Cambridge, University of Málaga, University of the Basque Country, University of Oxford, University of Warwick, SAE 2009 in Zaragoza, ICFSN 2009 in Paris and EAA-ESEM 2009 in Barcelona. Marco van der Leij would like to thank the Spanish Ministry of Science and Innovation (SEJ2007-62656) and Jaromír Kovářik acknowledges Spanish Ministry of Science of Innovation (ECO2009-09120) and the Government of the Basque Country (IT-223-07) for financial support.

<sup>†</sup>*Corresponding author.* Dpto. Fundamentos del Análisis Económico I, University of the Basque Country. Email: [jaromir.kovarik@ehu.es](mailto:jaromir.kovarik@ehu.es).

<sup>‡</sup>Center for Nonlinear Dynamics in Economics and Finance (CeNDEF), University of Amsterdam. E-mail: [m.j.vanderleij@uva.nl](mailto:m.j.vanderleij@uva.nl)

# 1 Introduction

Social networks play an important role in many socio-economic settings, and it has been documented that the particular network architectures influence both global and individual economic outcomes.<sup>1</sup> It is then important to understand both how and why networks are formed.

Risk and uncertainty are ubiquitous and individual attitudes toward risk constitute a fundamental element of decision-making in economics and finance (Pratt, 1964; Arrow, 1965). Risk is also present while people form ties with others, but so far the role of risk and risk aversion in network formation has been neglected in the literature.

Economic agents form relationships with others in order to generate benefits. Network links provide access to information, ideas, different markets, or represent joint creation of innovations.<sup>2</sup> However, in many situations the potential benefits from linking to a certain agent are not fully observable before the establishment of the relationship. Hence, creating connections can be risky. In network contexts, people can avoid such risky decisions by connecting to individuals about whom they have more information. In this way, risk aversion may play an important role in network formation processes.

In this paper, we propose a network formation model, building on standard microeconomic concepts of utility maximization, incomplete information, and risk aversion. In the model, the linking decision is endogenized and individuals can create links in two ways. They can link up either locally to neighbors of their neighbors or globally, using random search in the population. Naturally, creating links randomly is more uncertain and aversion to risk influences which type of link a node is willing to create.

The introduction of risk aversion and risk exposure to network formation generates several questions. Can different network positions of individuals be traced back to heterogeneity in risk attitudes? If so, how does it affect the payoff distribution? How does higher exposure to risk and uncertainty influence linking decisions and which effect does it have on generated architectures? Given the relevance of networks for economic outcomes, answering these questions can contribute to a better understanding of wealth distribution in networked populations. It also provides an opportunity to evaluate the impact of exogenous environmental changes, such as natural disasters, financial crises, emergence of epidemics or policy interventions, onto networked societies.

In our model, individuals sequentially enter the population. Initially, an entering individual, say,  $A$  does not know anyone, and acquaints a random individual  $B$ . By interacting with  $B$ ,  $A$  learns the benefits from the interaction with  $B$ , but she also learns about the potential benefits

---

<sup>1</sup>See Goyal (2007) and Jackson (2008) for reviews.

<sup>2</sup>See Jackson (2010) for a survey of economic applications.

from the neighbors of  $B$ .<sup>3</sup> With this information,  $A$  then has to make a choice: does she acquaint one of  $B$ 's neighbors about whom she is now informed, or does she leave the "circle of friends" of  $B$  and meet some random stranger?

We argue that  $A$  connects to a neighbor of  $B$  if at least one of them is attractive enough for  $A$ . Otherwise, she will create a link to a random individual in the population. There are two crucial factors for such a decision: first, the beliefs about the benefits one could get from neighbors of neighbors versus that from strangers, and second, the amount of risk aversion of the individual, because creating a link to a random node is much riskier than creating a link to a neighbor of a neighbor, about whom one has more information.

We show that in the resulting network, risk averse individuals tend to have neighborhoods that are highly clustered; that is, neighbors of a risk averse individual are more likely to be neighbors as well. Thus, we are partly able to trace back heterogeneity in local clustering to heterogeneity in risk aversion. On the other hand, there is no relation between risk attitudes and degree. Since the link between the network position and risk aversion is new in the literature, in Section 5 we empirically test the predictions of the model. We find that these predictions are confirmed in the data.

The main findings of the model arise from the microfoundation of the network formation process. First, we show that there is a positive relation between the clustering coefficient and expected payoffs of agents. There are two reasons that induce people to link through network-based meetings and, consequently, increase their clustering coefficient: risk aversion and attractive neighborhoods. In the former case, risk averse people accept relatively lower payoffs and drive the payoff of highly clustered individuals down, while the latter effect makes people link through the network, because they link to attractive neighbors of their neighbors, rising the payoffs of highly clustered agents. We show that the latter effect outperforms the former.

Second, relating the economic fundamentals of the environment to linking decision allows to make predictions about the shape of generated networks. We show that the linking decisions of individuals and, consequently, the whole network architecture depend on the riskiness of the environment. More precisely, the volatility of benefits from linking to "strangers" generates more clustered networks and allows us to rank the in-degree distribution in the sense of second-order stochastic dominance. If the environment becomes more risky, the number of links of highly connected individuals increases, while the number of links of less connected agents decreases. Hence, more risky environments generate more unequal networks.

Our paper makes three important contributions to social network analysis. First, the proposed network formation process has solid behavioral foundations as in economic models, but at

---

<sup>3</sup>The idea that agents learn more about their local neighborhood than about nodes outside the local network is natural (Galleoti et al., 2010; Gallo, 2009) and has been confirmed empirically by Fafchamps, Goyal and Van der Leij (2010).

the same time our model is empirically verifiable as in statistical mechanics models, synergizing the economic and physics approach to network formation (see Jackson, 2006, for a discussion of this issue).<sup>4</sup>

Second, we discover a new mechanism that relates network structure to individual payoffs. This mechanism differs starkly from the more traditional, sociological explanations (Granovetter, 1973; Coleman, 1988; Burt, 1995). In those sociological theories, the social network is assumed to be rigid, and behavioral processes taking place on the network allow some individuals to benefit more from their (given) network position than others. In contrast, in our model it is the *network formation process* that creates a relation between clustering and payoffs. Individuals learn about their network neighborhood during the network formation process. If this information or experience is particularly positive, then the individual does not leave her neighborhood, and therefore any additional tie will be created within her local network. Only those who happen to get into a bad neighborhood have incentives to leave and try to form “random” ties. We may therefore call this mechanism a “good neighborhood effect”.<sup>5</sup>

The last contribution lies in the linkage between the environment the network is embedded in (characterized by the distribution of payoffs from linking), on the one hand and the resulting network architecture and expected payoffs, on the other. Jackson and Rogers (2007) show that more unequal networks lower the welfare of the population. However, their model is mechanical and cannot predict under which conditions the network is to be expected more or less unequal. We show that shocks increasing the volatility of potential benefits from relationships make people rely more on local neighborhoods, affecting the clustering and degree distribution of the global architecture.<sup>6</sup> This may help to evaluate policies in contexts, where networks play a role. A particular intervention will affect both behavior and the underlying social organization and, therefore, final policy implications cannot be correctly evaluated without a proper understanding of how the network itself will react to the proposed policy. Hence, abstracting away the endogenous network formation process may result in undesirable policy decisions.

---

<sup>4</sup>Campbell (2008) also microfoundates the decision to link globally or locally. He argues that people link to neighbors of their neighbors in order to signal their willingness to cooperate. See also Vigier (2009) or Bramoullé & Rogers (2009). These models are purely theoretical, but we provide empirical evidence that risk aversion plays an important role in the network formation process. Babus & Ule (2008), Gallo (2009) and König, Tessone & Zenou (2009) start from the standard game theoretical models, but propose versions that are empirically more realistic.

<sup>5</sup>Naturally, this mechanism does not preclude the existence of the other mentioned mechanisms. However, we do point out that the empirical *identification* of those sociological theories becomes much more challenging in the presence of the “good neighborhood effect”. Data from a simple non-repeated survey measuring networks and individual performance, such as has been common in the sociological and management science literature (e.g. Burt, 1995), would not allow the researcher to separate the effect of, say, structural holes from our mechanism. These problems are in line with the identification problems recognized in the social interaction/peer effects literature (Manski, 1993; Moffitt, 2001).

<sup>6</sup>The present model does not allow for relinking, but we believe that the same mechanism would also operate when agents decide which links to sever or to whom to redirect their links.

We proceed our paper as follows. In Section 2 we present the model and establish the relation between network structure and risk aversion. In Sections 3 and 4 we continue the theoretical analysis linking the network structure and payoffs, and analyze the effect of contextual variables on the resulting architecture. Section 5 tests the theoretical predictions, regarding network structure and risk aversion. In Section 6 we discuss potential extensions and applications of the model. Section 7 concludes. An illustration of how the model changes if we relax one of the model's assumption, model simulations and proofs are relegated to Appendix.

## 2 The Model

In this section, we propose a network formation model. The model is a variant of the models in Jackson and Rogers (2007) and Vázquez (2003), but unlike those paper, we base the model on standard economic assumptions. We formally characterize the network structures generated by the model and relate individual network characteristics to their risk attitudes.

Let  $N(t)$  be the population of agents existing at time  $t$ . The directed network among those agents is denoted by  $G(t)$ , and  $g_{ij}(t) = 1$  denotes a directed link from  $i$  to  $j$  at time  $t$ . Define  $N_i(t) = \{j : g_{ij}(t) = 1\}$  as the outdegree neighborhood of individual  $i$  at time  $t$ . For notational convenience, dependence on  $t$  will be dropped if confusion is unlikely.

Network formation occurs through the following dynamic process. Each period one new player enters the population. This player is identified by its entrance period  $i$ . We assume that individuals have a capacity constraint with respect to outdegree, and agent  $i$  is only able to have  $m$  links pointing outwards, which it creates when entering the network. On the other hand, we assume that individuals do not have capacity constraints with respect to the in-degree, and therefore no individual  $j$  refuses a link  $ij$  if offered by  $i$ . Links are only created by the entering node  $i$  at time of entrance. Afterwards links cannot be changed.<sup>7</sup>

The benefits node  $i$  gets from linking with an existing node  $j \in N(i) \equiv \{0, 1, \dots, i - 1\}$  is denoted as  $u_{ij}$ , which is drawn from an i.i.d. distribution  $F$  having support on the interval  $[a, b]$ . This distribution (but not the realizations) is common knowledge and has mean  $\bar{u}$ . Naturally, the assumption that  $u_{ij}$  is independently distributed for each  $i$  and  $j$  is a very strong assumption. For example, it excludes the possibility that some nodes have intrinsic traits  $v_j$  that make them more beneficial for any node  $i$ . It also excludes any (indirect) network effects once the network is in place. As we will see, even this simplest case gives rise to a relation between network structure, risk aversion and payoffs, such that it is best to focus on this case first. Later, in Section 6, we explore deviations from the independence assumptions.

---

<sup>7</sup>The network formation process is initialized by letting the first  $m + 1$  agents create a link with all their predecessors, that is, each agent  $k \in \{0, \dots, m\}$  creates  $k - 1$  links, such that  $g_{ki} = 1$  if  $i < k$ , and  $g_{ki} = 0$  if  $i > k$ .

Let  $U_i = U_i(\sum_{N_i} u_{ij})$  be the (Bernoulli) utility function of  $i$ . This utility is strictly increasing and concave with a constant risk premium  $r_i$ , such that node  $i$  is indifferent between a sure benefit of  $\bar{u} - r_i$  against a random benefit drawn from  $F$ , that is  $U_i(x + \bar{u} - r_i) = E[U_i(x + u_{ij})]$  for all  $x$ . Let individuals differ in their risk attitudes. In particular, we assume that there are two levels of risk aversion with risk premium,  $r_H$  and  $r_L$ , with  $\bar{u} - b < r_L < r_H < \bar{u} - a$ . A new node  $i$  has risk premium  $r_i = r_H$  with probability  $\theta$ , and  $r_i = r_L$  with  $1 - \theta$ .<sup>8</sup>

The decision of node  $i$  to link with  $m$  nodes  $j_1, \dots, j_m$  goes as follows. When entering, individual  $i$  initially does not have information on the benefits nor on the number of links of the other individuals. Nonetheless, individual  $i$  may obtain information on  $j$  by connecting to a friend of  $j$ , say  $k$ , who is connected to  $j$ ,  $g_{kj} = 1$ .<sup>9</sup> The new node  $i$  first connects with one randomly drawn existing node  $j_1$ . We assume that by connecting to  $j_1$ , individual  $i$  obtains a perfect signal on the benefits of the out-degree neighbors of  $j_1$ .<sup>10</sup> Individual  $i$  then makes a decision on whom to connect next. If

$$\max_{k \in N_{j_1}} u_{ik} > \bar{u} - r_i \quad (1)$$

then  $i$  connects to the node  $k \in N_{j_1}$  that maximizes  $u_{ik}$ , otherwise  $i$  connects to a random node outside  $N_{j_1}$ . Let this second node to which  $i$  links be denoted by  $j_2$ . By connecting to  $j_2$  it again learns about the benefits of the nodes in  $N_{j_2}$ . Starting from  $j_2$  the process is repeated, that is, if

$$\max_{k \in N_{j_2}} u_{ik} > \bar{u} - r_i,$$

then  $i$  connects to the node  $k \in N_{j_2}$  that maximizes  $u_{ik}$ , otherwise  $i$  connects to a random node outside  $N_{j_2}$ .<sup>11</sup> The linking process of  $i$  stops when  $i$  has formed  $m$  links, after which agent  $i + 1$  enters the network, and starts to create links. Figure 1 illustrates one step of the network formation process when  $m = 3$ .<sup>12</sup>

---

<sup>8</sup>We would like to stress here that all the results of this paper would still work if we assumed more complex distributions of risk aversion.

<sup>9</sup>An alternative approach is to consider a process that creates homophily with respect to quality, such as in Montgomery (1991). In that case, linking to, say, a high type would give us information on the type of the neighbors.

<sup>10</sup>As in Jackson and Rogers (2007) we assume that network search is directed, in particular, channeled through out-degree links. Allowing for network search through in-degree links would complicate the analysis significantly. See Jackson and Rogers (2007) for details.

<sup>11</sup>We assume that  $i$  does not recall (or is unable to contact) the neighbors of previously visited links, that is, at the  $s$ -th step, the agent  $i$  is unable to recall the benefits from linking to agents in neighborhoods of  $N_1, \dots, N_{s-1}$ . This is only in order to keep the model tractable, but has no major implications on the results. If we would allow for aggregation of information on neighbors, the probability to link a friend would steadily increase during the  $m$  linking steps.

<sup>12</sup>It is worth noting that with the decision rule in (1), we assume a certain bounded rationality of agents. More precisely, suppose that  $i$  connects to  $j$  in the first linking decision and  $k$ , a neighbor of  $j$ , afterwards. If  $j$  linked up locally to a neighbor of  $k$  in one of the previous rounds, then there exists a node  $l$  who is an out-degree neighbor of both  $j$  and  $k$ . Hence,  $i$  observed  $l$  after the first linking, but has not connected to him (since he linked up to  $k$ ), and observes  $l$  again after linking to  $k$ . This means that  $i$  observes again the same node, about whom he has full



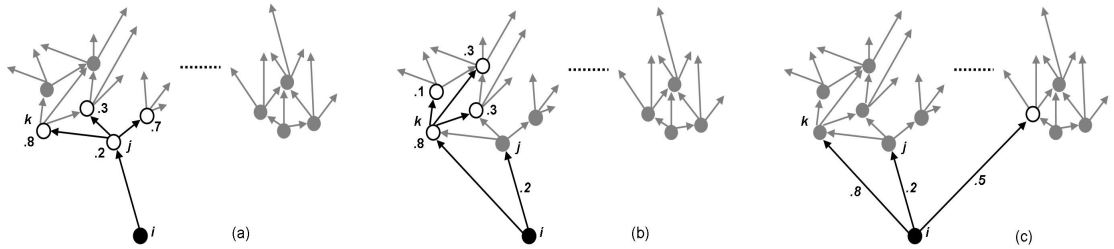


Figure 1: Example of link formation of a new node  $i$ , when  $m = 3$ ,  $\bar{u} = .5$  and  $r_i = .1$ . (a) Node  $i$  creates a random link to  $j$ , and learns about the benefits of linking to  $j$  and  $j$ 's neighbors. A link to  $k$  gives the highest benefit to  $i$ , and since  $u_{ik} = .8 > \bar{u} - r_i = .4$ , a link to  $k$  is preferred to a random link. (b) Node  $i$  creates a link to  $k$  and learns about the benefits of linking to  $k$ 's neighbors. (c) Since the benefits of linking to a neighbor of  $k$  are all lower than  $\bar{u} - r_i$ , node  $i$  creates a link to a random node in some other part of the network.

We first analyze the statistical properties of the network that is generated by the above network formation process, and we compare it to the structural properties commonly observed in social networks (see Jackson and Rogers, 2007). In particular, we will look at degree distribution and local clustering. Given the complexity of the problem (especially due to the dependence of meetings on the network structure), we rely on mean-field analysis of the model. The mean-field approach approximates the complex evolution of a stochastic system by a simpler deterministic system, in which the evolution is determined by the expected change. In order to check the precision of the approximations, we provide simulation results for particular cases.

The first step is to derive the probability that an agent entering at time  $t+1$  is willing to create a link with a particular node  $i$  having in-degree  $d_i(t) = \sum_j g_{ji}(t)$ . Define  $p(r)$  as the probability that an entering node  $t+1$  of type  $r$ , having already linked to  $j_1, \dots, j_s : 0 < s < m$  agents, decides to link to a friend of  $j_s$  instead of linking to someone randomly. That is,  $p(r)$  measures the tendency to search locally through the network, instead of randomly. This probability equals the probability that at least one of  $m$  friends of  $j_s$  is more attractive than the benefits of linking randomly, and is therefore given by:

$$p(r) = 1 - F(\bar{u} - r)^m. \quad (2)$$

Note that this probability depends on the risk aversion of agent  $t+1$ , such that  $p(r_H) > p(r_L)$ .

---

information. Since this occurs with positive (non-negligible) probability, there is a certain (expected) utility loss from linking to a neighbor of a neighbor anytime agents link through the network. Completely rational agents should take this potential utility loss into account while deciding whether to connect locally or through random search. In the main model, we abstract from this possibility, but in the appendix, we illustrate how taking this into account affects the linking decision of individual for  $m = 3$  and how it adds substantial complexity to the model.

At entrance of agent  $t + 1$  the process is as follows: agent  $t + 1$  first links up randomly. Thus, for a particular agent  $i < t + 1$  the probability of receiving this link is  $\frac{1}{t}$ . Once this link has been created,  $t + 1$  faces  $m - 1$  decisions between linking locally through the network (by observing the neighbors of his neighbors), or linking to a randomly chosen agent from the population. In this case, the probability of  $i$  to increase its degree in one of these decisions is approximately

$$\frac{1 - p(r_{t+1})}{t} + p(r_{t+1}) \frac{d_i(t)}{t} \frac{1}{m}, \quad (3)$$

where the first part corresponds to the probability that  $t + 1$  decides for a random search and links up to  $i$ . The second part of the expression is the joint probability of three events: (i)  $t + 1$  finds it attractive to connect through the network structure,  $p(r_{t+1})$ , (ii) she has connected to one of the  $d_i$  (in-degree) neighbors  $j$  of  $i$  in the previous decision,  $d_i(t)/t$ ,<sup>13</sup> and (iii)  $i$  has the largest gain for  $t + 1$  out of the (outdegree) neighbors of  $j$ ,  $1/m$ .

Given that each link  $i < t + 1$  can receive at most one link in each period, that is, multiple links are ruled out, we can write the deterministic change of  $i$ 's in-degree in period  $t$  as

$$\frac{dd_i(t)}{dt} = \frac{1}{t} + (m - 1) \left[ \frac{1 - p_\theta}{t} + p_\theta \frac{d_i(t)}{t} \frac{1}{m} \right]. \quad (4)$$

where  $p_\theta = \theta p(r_H) + (1 - \theta)p(r_L)$  is the expected probability that an agent finds it optimal to follow a network-based meeting.

The resulting differential equation (4) deserves some attention. Expression (4) shows that the proposed linking process leads to a similar evolution of the degree distribution in the present setting, as in Jackson and Rogers (2007).<sup>14</sup> The right-hand side of (4) is a weighted sum of uniform random matching and linear preferential attachment.<sup>15</sup> The equation (4) can be easily solved and, using that the initial in-degree of entering agents is 0, the in-degree of an agent  $i$  at period  $t$  is

$$d_i(t) = \left[ \frac{m(m + p_\theta - mp_\theta)}{(m - 1)p_\theta} \right] \left( \frac{t}{i} \right)^{\frac{(m-1)p_\theta}{m}} - \frac{m(m + p_\theta - mp_\theta)}{(m - 1)p_\theta}. \quad (5)$$

A second important feature of socially generated social networks is the clustering coefficient. The clustering coefficient of individual  $i$  is the fraction of  $i$ 's direct neighbors that are neighbors themselves, thus measuring how much overlap there is in friendship circles. There are several definitions of the clustering coefficient depending on the way one keeps track of the direction of

---

<sup>13</sup>This approximation does not take into account that there is positive assortativity in the network. If indegree neighbor  $j$  is found through local network search, then the probability that  $j$  is found increases in the degree of  $j$ ,  $d_j$ , and given that there is a positive degree correlation, in the degree of  $i$ ,  $d_i$ , as well. Simulation results obtainable from the authors suggest that this ignorance does not have major implications on the results.

<sup>14</sup>The idea behind (3) goes into the direction of Vázquez (2003), where in each moment the entering agent decides whether to follow the network or to search randomly, but as we show here this can be easily merged into the process proposed by Jackson and Rogers (2007).

<sup>15</sup>Hence, our model also encompasses the properties of Pennock et al. (2002), who use a similar equation to show that it can generate a large set of empirically plausible unimodal distributions of nodes observed in Internet.

the links.<sup>16</sup> We focus on one measure, called the fraction of transitive triples. In the terms of our model, it measures the fraction of times, in which an agent  $i$  connects to agent  $k$ , who has an in-going link from  $j$  who at the same time has an in-going link from  $i$ . Formally:

$$C_i(g) = \frac{\sum_{j \neq i; k \neq i, j} g_{ij} g_{jk} g_{ik}}{\sum_{j \neq i; k \neq i, j} g_{ij} g_{jk}}. \quad (6)$$

Again there are several definitions of the average clustering coefficients in the population. Here, we focus on the definition that measures the total fraction of transitive triples in the network. Formally,

$$C(g) = \frac{\sum_{i; j \neq i; k \neq i, j} g_{ij} g_{jk} g_{ik}}{\sum_{i; j \neq i; k \neq i, j} g_{ij} g_{jk}}. \quad (7)$$

We first show that the proposed network formation process, based on the utility maximization of agents, exhibit typical features of empirical social networks:<sup>17</sup>

**Theorem 1** *Under mean field approximation,*

(i) *if  $m > 1$  and  $p(r_L) > 0$ , the (complementary) cumulative distribution function of in-degrees in period  $t$  can be characterized as*

$$1 - F_t(d) = \left( \frac{\frac{m(m+p_\theta - mp_\theta)}{(m-1)p_\theta}}{d + \frac{m(m+p_\theta - mp_\theta)}{(m-1)p_\theta}} \right)^{\frac{m}{(m-1)p_\theta}}$$

(ii) *the average clustering coefficient in the network satisfies*

$$C(g) \geq \frac{p_\theta}{m^2} (m - 1). \quad (8)$$

Theorem 1 shows that - as long as there is a positive probability of low-risk individuals to find an attractive agent through network -  $f(d) \propto d^{-\left(\frac{m}{(m-1)p_\theta} + 1\right)}$  for large  $d$ ; that is, the in-degree distribution of agents for large  $d$  has a power-law distribution in the tail, and the average clustering coefficient will be strictly positive independently of other characteristics of the model.

We now turn to the relation between an agent's risk aversion and her position in the network. We have the following proposition:

**Proposition 2** *The in-degree of  $i$ ,  $d_i(t)$  is independent of  $r_i$ , the degree of risk aversion of  $i$ , while  $r_H$  types have a higher fraction of transitive triples than  $r_L$  types.*

<sup>16</sup>See Wasserman and Faust (1994) or Newman (2003) for discussions on this issue.

<sup>17</sup>All proofs can be found in Appendix.

By assumption of the model the outdegree of all individuals is identical,  $m$ , and therefore independent of risk aversion. Expression (5) shows that the same also holds for in-degree. On average, the in-degree of both  $r_H$  and  $r_L$  types depends on the distribution of risk aversion in the population, determined by  $\theta$ , rather than on  $i$ 's type. This is due to two facts; first, because link formation is one-sided, that is, only the entering agent  $t + 1$  decides on the formation of a link, and second, because the distribution of risk aversion in the population is independent of the distribution of utilities that  $i$  conveys for entering agent  $t + 1$ . As a result, entering node  $t + 1$  does not take into account the risk attitude of agents while deciding whether to connect to them or not.

Proposition 2 also shows that in our model more risk averse types with  $r_H$  have a higher fraction of transitive triples than  $r_L$  types. The intuition is that random links do not contribute to closing triads, whenever the network is large. The fact that  $r_H$  types have fewer random links implies that they have larger clustering.

To complement the results of this section, we provide simulation results of the model in the appendix. The motivation is twofold. First, we aim to check the precision of the mean-field approximations in previous theoretical results. The simulations confirm the findings from Theorem 1 and Proposition 2. Second, we would like to show that the proposed network formation process generates architectures, which exhibits other typical properties of empirical social networks, namely short network distances, assortativity, and the negative clustering/degree correlation. The first states that the average network distances and the largest distance between two (reachable) nodes in real-life networks are in general low in relation to the size of the network. The second property, assortativity, is a tendency such that high (low) degree nodes are more likely linked to high (low) degree nodes. Last, negative clustering/degree correlation simply suggests that the larger the degree of the node the lower its clustering coefficient. The simulated networks exhibit these network properties (see Figures 4 and 5 in Appendix).

### 3 Network Position and Payoffs

One of the main interests in the study of social and economic networks is the relation between individual network position and individual economic outcomes. In particular, the relation between clustering coefficient and payoffs has raised some debate. On the one hand, the theory of network closure (Coleman, 1988) argues that local clustering is beneficial, because it allows for better monitoring, which enforces more cooperation and higher trust levels (see also Granovetter, 1985). On the other hand, the theory of structural holes (Burt, 1995) argues that network positions that bridge different groups allow for better information access and control. These structural hole positions are typically characterized by low local clustering.

Given that our model builds on standard economic assumptions of utility maximization, we are able to give an alternative view on the relation between clustering and payoffs. To this aim, we first show that the expected monetary payoff depends on the type. Define the monetary payoff of  $i$  as  $\sum_{j \in N_i} u_{ij}$ .<sup>18</sup>

**Proposition 3** *Suppose that  $r_H > r_L > 0$ . The expected monetary payoff of an individual of type  $r$ ,  $E[\sum_{j \in N_i} u_{ij} | r_i = r]$ , is decreasing with the risk premium  $r$ .*

Proposition 3 shows that, in our model, individuals with larger risk premium tend to earn less. There is a standard economic interpretation behind this result: risk averse individuals accept sure relatively low payoffs from second-order neighbors in order to avoid risky decisions.

Now we proceed with the analysis of the relation between the individual network position and payoffs. By the construction of the model, there is no relation between in-degree (outdegree and, hence, degree) and payoffs, since only the quality of outdegree neighbors are relevant for payoffs of agents and outdegree is the same for all nodes. Therefore, we focus on clustering coefficient.

So far, we have derived two results regarding risk aversion, Propositions 2 and 3. The former establishes a positive relation between risk aversion of individuals and their clustering coefficient, while the latter proposition proves that the expected payoff is negatively affected by risk aversion. This might suggest that if there is any relation between clustering coefficient and payoffs it should be negative. However, Theorem 4 shows that this is not the case.<sup>19</sup>

**Theorem 4** *For  $m < 5$ , the expected monetary payoff of individual  $i$ , conditional on her clustering coefficient  $c$ ,  $E[\sum_{j \in N_i} u_{ij} | C_i = c]$ , is (weakly) increasing in  $c$ .*

At first sight, Theorem 4 seems to contradict Propositions 2 and 3. Nevertheless, a closer look at the forces behind the formation of transitive triples (that determine the level of clustering) reveals a more complex relation between payoffs and clustering.

There are two forces that influence this relation. Propositions 2 and 3 captures one direction: Risk averse individuals pay a risk premium for sure payoffs from network-based linking, which leads to larger clustering and this drives payoffs of more clustered individuals down.

However, there is a “neighborhood effect,” which goes into the opposite direction. More precisely, people, whose neighborhoods are attractive, tend to stay within their neighborhoods, i.e. link to the neighbors of their neighbors, increasing the individual clustering coefficient and increasing the average payoff of highly clustered nodes.

---

<sup>18</sup>Utilities of individuals  $U(\sum_{j \in N_i} u_{ij})$  are non-comparable, and therefore not considered.

<sup>19</sup>We were only able to prove the theorem for  $m < 5$ . The matters are complex for larger  $m$ . However, we can easily show that there are upper and lower bounds on the payoff conditional on clustering, both strictly increasing in the level of clustering. Therefore, we conjecture that the theorem holds for any  $m$ .

Theorem 4 shows that the neighborhood effect always dominates the influences from Propositions 2 and 3. Note that if relatively less risk averse individuals adhere to the clustering, then their neighbors have to be really attractive. This effect drives up the average payoffs conditional on the level of clustering.

## 4 Risk Preferences and Contexts

Recent empirical literature presents evidence that who links up with whom is an endogenous process influenced by the socioeconomic environment (de Weerdt, 2004; Krishnan & Sciubba, 2009). Other streams of literature document how existing networks reshape when the environmental conditions change. For instance, Goyal, van der Leij & Moraga-González (2006) report how the structure of scientific collaboration has changed over past decades in parallel with the burst of communication technologies, and Eeckhout & Munshi (2010) - while analyzing an informal financial institution that brings agents together in groups - observe that participants rematch immediately following an unexpected exogenous regulatory change. To provide an example outside the domain of economics, it has been documented that the emergence of HIV epidemics has considerably affected the architecture of needle-sharing among drug users (Rothemberg et al., 1998). Hence, networks endogenously reorganize in presence of exogenous shocks. This generate many new questions. Which aspects of the environment trigger the endogenous adaptation of social organization? Why and how do network architectures react to these variables?

The present framework allows us to relate how network properties depend on the economic and social context in which the network formation takes place. If risk attitudes or the distribution of benefits are different in one environment compared to the other, then individual decisions are different, and so is the network formation process and the eventual network structure. Hence, different social and economic contexts lead to different network architectures, and this may have implications on eventual social and economic outcomes as well.

We illustrate formally how the change of the context, characterized by the distribution function of benefits,  $F$ , interacts with risk preferences of individuals. To this aim, we assume that agents have constant relative risk aversion utility functions with risk aversion coefficient  $\rho_i$  and the payoff distribution of linking to individual  $j$ ,  $u_{ij}$ , is normally distributed with mean  $\bar{u}$  and variance  $\sigma^2$ .

Let the (Bernoulli) utility function of individual  $i$  be

$$U_i(x) = -\frac{1}{\rho_i} e^{-\rho_i x}. \quad (9)$$

With this utility function, the risk premium of an individual will be a function of her risk-aversion coefficient and characteristics of the payoff distribution  $F$ :

$$r = r(\rho_i, \bar{u}, \sigma^2)$$

**Proposition 5** *Suppose that individual  $i$  has utility function (9) with coefficient of absolute risk aversion  $\rho_i$ . Let  $F$  be a cumulative distribution function of a normal distribution with mean  $\bar{u}$  and variance  $\sigma^2$ . Then,*

- (i)  $p[r(\rho_i, \bar{u}, \sigma^2)]$  does not depend on  $\bar{u}$ , and
- (ii)  $p[r(\rho_i, \bar{u}, \sigma^2)]$  increases with  $\sigma^2$ .

This result has strong implications for the model. It shows that we can observe the same individuals in very different network position (in terms of their random vs. local search), depending on the riskiness of the environment, in which a particular network is embedded. More precisely, we show that more risky contexts will drive people to link up to neighbors of their neighbors more often. On the other hand, a simple increases or decrease of the average benefits will not affect the decisions of agents, as long as preferences for absolute riskiness are preserved.

Proposition 5 also has direct implications for the global structure of the model:

**Theorem 6** *Let  $F$  and  $F'$  be two normal distribution functions with means  $\bar{u}$  and  $\bar{u}'$  and variances  $\sigma^2$  and  $\sigma'^2$  respectively. Consider the networks  $g$  and  $g'$  associated with linking benefit distributions  $F$  and  $F'$ .*

- (i) *If  $\sigma^2 > \sigma'^2$ , then the degree distribution of  $g'$  second order stochastically dominates the degree distribution of  $g$ , and  $C(g) > C(g')$ .*
- (ii) *If  $\sigma^2 = \sigma'^2$ , then the degree distribution of  $g$  and  $g'$  are identical and  $C(g) = C(g')$ , independently of  $\bar{u}$  and  $\bar{u}'$ .*

This result shows how a change of context affects the network properties. A mean preserving-spread of the payoff distribution has a direct effect on whether the network will be more or less random, since more risky environment enhances local, non-random search. This at the same time affects the probabilities of incumbent nodes to receive a link. More precisely, less connected agents, who mostly rely on global search, become less likely to receive a link, while agents above a certain degree, whose main source of new connections is to be found through the network, are now more likely to receive new incoming links. The global affect, stated in Theorem 6, is a shift of degree distribution in terms of second-order stochastic dominance; a riskier environment creates more inequality in terms of connectivity.

Concerning the local clustering of the network, riskier environments generate more clustered network architectures. Location shift alone will affect neither the degree distribution nor local clustering of the network.

These findings illustrate how network architectures endogenously adapt to changes of environmental variables. For instance, people might be more careful choosing close friends than

mere acquaintanceship, leading to more clustered networks in the former case. Other examples might be that firms' position in technological networks can differ according to the riskiness of the innovation in progress, and that people will search for new sexual partners more locally after the start of the HIV epidemics. Despite that our network model does not allow for relinking, we believe the same logic - which determines which type of connection will establish - will also operate when individuals sever or redirect their links.

## 5 Empirical Analysis

In this section, we report an elicited real-life social network, and relate the social network characteristics to risk attitudes of individuals.

The data presented here were collected within a sequence of surveys and experiments in the spring semester of 2005 at the University of Granada, Spain.<sup>20</sup> First-year undergraduate students of Economics were invited to participate in the sessions. In order to stimulate the participation, subjects were rewarded with classroom points that served to increase the final grade in the course of Microeconomics I. Subjects were informed that the number of points obtained during the sessions in which they would participate, contributed to their final grade in the course in the following way: the student who obtained the highest number of points would add three extra points (out of ten) to her final grade. Other subjects' grade depended on how close their performance was to the winner's. Subjects were not informed of their own and others' performances in any of the sessions and tasks until all sessions had ended. In this document, we combine the data from two particular sessions: one in which the network was elicited and one in which risk aversion was elicited.

This data set is the only one we know that contains both data on risk aversion and on network structure, and therefore gives us a unique opportunity to directly test the predictions from our theoretical model.

### 5.1 Network elicitation

All first-year students (more than 300) were invited to participate in a network elicitation survey and a subsequent game serving to elicit their sharing preferences. The aim of the survey to elicit relationships among subjects was to be able to study experimentally the effects of social distance on sharing behavior, rather than the analysis of the actual underlying network structure (Brañas et al., 2006). Nevertheless, as shown below, the network structure exhibits features traditionally observed empirically in social and economic networks.

---

<sup>20</sup>The sequence of sessions carried out by the pool under scrutiny was as follows: a dictator game (March), a GRE-type math test (beginning of April), risk aversion elicitation (end of April), network elicitation jointly with a two-stage variation of the dictator game (May), and session containing Traveler's dilemma and self-confidence questionnaire (June).



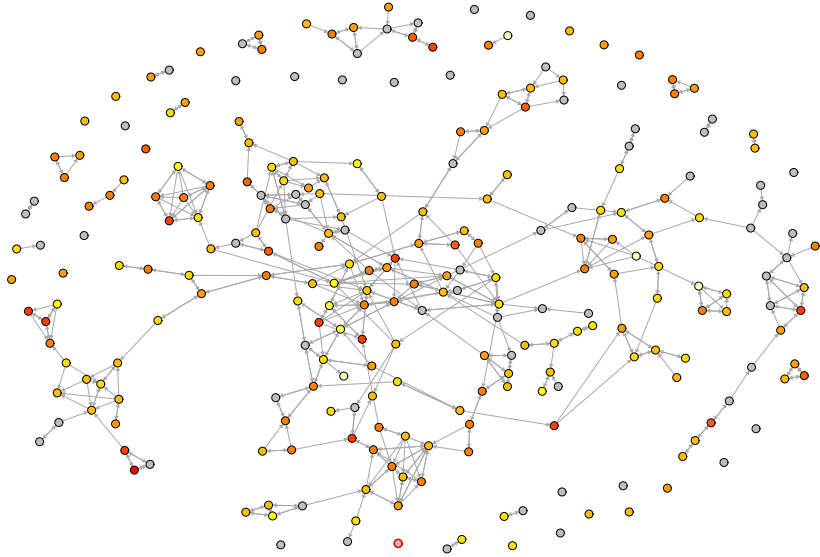


Figure 2: The network of friendship relationships between 256 undergraduate students at the University of Granada. The colors of the nodes denote the amount of risk aversion (white: very risk loving. red: very risk averse. grey: risk aversion not elicited).

The elicitation protocol was very simple. Subjects had to write the names of their friends in class on a piece of paper. There was *no* restriction on the number of friends people could have named. On the other hand, a special emphasis was put in the instructions to make clear to the subjects that in the next stage of the experiment they would be given a chance to benefit one of the friends they named. To this aim, the instructions stated: “the more friends you will name, the lower the probability to be able to reward a particular friend of yours.” For more details on the experimental protocol used in the network elicitation stage, we refer the reader to Brañas et al. (2006).

Figure 2 provides a global view of the elicited social network of the 256 individuals, who participated in the network elicitation part of the experiment. There is a giant component in the network containing 183 nodes and thanks to the highly interconnected cluster in the center of the graph the average distance between reachable individuals (i.e. between agents in the giant component) is quite small (on average 7.5).

Table 1 provides a summary of the statistics of the network variables of interest: in-degree, outdegree, degree, and clustering coefficient. In-degree of individual  $i$  gives the number of individuals  $j \neq i$  that name  $i$  as a friend of theirs; outdegree equals the number of individuals  $i$  names as her friends; degree is the number links of  $i$ , abstracting from the direction of the link; clustering coefficient gives the fraction of pairs of  $i$ 's neighbors that are neighbors themselves (ignoring direction of the links). Thus, the first three variables are measures of connectivity of individuals, while the clustering coefficient reflects the density or interconnectivity of agents' neighborhoods. We can see that the average values of connectivity characteristics are low in

	Obs.	Mean	St.Dev.	Min	Max
indegree	256	2.031	1.671	0	8
outdegree	256	2.031	1.444	0	6
degree	256	2.750	1.887	0	9
clustering	181	0.397	0.360	0	1
female	256	0.534	0.500	0	1
age	176	19.53	2.272	18	35
risk	187	5.39	1.702	0	10

Table 1: Summary statistics for the friendship network among 256 undergraduate students at the University of Granada. Individual clustering coefficient is only defined and computed for those nodes with degree 2 or higher and therefore contains only 181 observations.

	A	B
A	4,4	8,3
B	3,8	0,0

Table 2: Game from the risk aversion elicitation.

comparison with other studies, as for example Leider et al. (2009). We believe that this is due to the elicitation mechanism, which motivates to solely name close friends, rather than acquaintances. Moreover, the value of the average clustering, .397, is an order of magnitude larger than in a comparable randomly generated network.<sup>21</sup>

## 5.2 Risk aversion elicitation

The risk aversion test was performed in a session different from the network elicitation one. There were 187 subjects who coincided in both sessions, allowing us to relate their attitudes toward risk to their position in the class network.

To elicit risk attitudes, a variation of the standard Holt and Laury’s (2002) protocol was used. In the first stage, subjects were asked to choose between a “two-player” game in Table 2 and a fixed amount of money. Students were told (i) they would be the row players and (ii) the play of the column player would be simulated by a computer, which would play A with 50% probability or B with the same probability. The aim of this exposition was to somehow check the rationality of subjects, as playing A always leads to a higher payoff. In fact, only one subject, who was removed from the analysis, chose to play B. Moreover, given that the column player was a computer, neither social nor efficiency concerns were at work here. Hence, the presented game actually represents a lottery where a rational player, i.e. player who chooses A, obtains either 4 or 8 experimental points both with the same probability 1/2.

<sup>21</sup>The clustering of a randomly generated network with the same size and average degree would be 0.011.

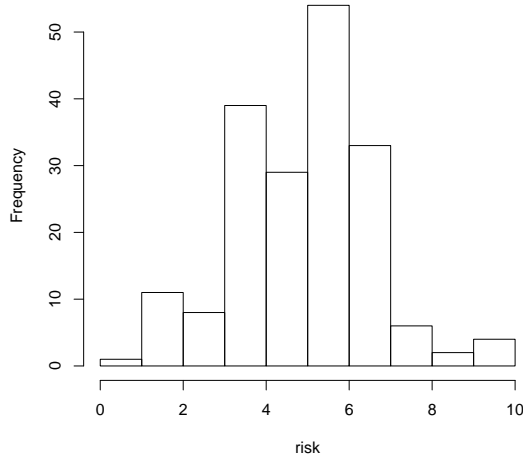


Figure 3: A histogram of the risk variable elicited from 187 students.

Each subject faced a series of 10 decisions, in which he had to decide between the game described above or a fixed amount of money. Decision 1 was to decide between either 1 experimental point for sure or to play the game; Decision 2 between 2 points for sure or the game, and so on, until Decision 10 between 10 points for sure or the game. The number of decisions in which the participant chooses to play the game is the elicited risk attitude we use in the present paper. Hence, note that the variable of interest measures risk lovingness: the larger its value the lower risk aversion.

Figure 3 shows a histogram of the risk variable. We observe a striking heterogeneity of choices; some students behave extremely risk-averse (risk=0), and others extremely risk-loving (risk=10). This suggests that risk aversion is not a homogenous principle, but that there is a substantial individual heterogeneity with respect to risk attitudes.

### 5.3 Results

Table 3 lists the correlation coefficients of the variable measuring risk attitudes with the measures of connectivity and clustering coefficient.<sup>22</sup> The correlations of connectivity measures are calculated for all students participating in both the network and risk attitude elicitation, but the correlation of the clustering coefficient with risk attitudes is calculated only for subjects, whose degree is larger than one.<sup>23</sup> The Table shows that there is no significant relation between

<sup>22</sup> $p$ -values are obtained from a random permutation test on zero correlation, in which the distribution under the null hypothesis is simulated by permuting the observations of the risk variable 100000 times and recalculating the correlation coefficient for the permuted sample.

<sup>23</sup>Since the clustering coefficient is defined as a number of friendships between the neighbors of a node divided by all possible friendships among them, it is not determined for people, whose degree is lower than two. Some scholars set the value of the coefficient in such cases to zero (see, for instance, Vega-Redondo (2007)). To check the robustness of our findings, we also performed the present analysis under this definition. The obtained results lead to even stronger relation between the variables of interests, without affecting the relation between connectivity

	indegree	outdegree	degree	clustering
risk				
correlation	-0.036	-0.002	0.003	-0.226***
<i>p</i> -value	0.629	0.975	0.965	0.006

Table 3: Correlation between risk attitudes and network characteristics. \*\*\*significant at 1% level

risk attitudes of individuals and their connectivity measures, while risk and clustering coefficient are significantly correlated. Recall that high values of risk denote a *low* risk aversion and the negative correlation of the two variables thus suggests that neighbors of risk averse individuals are more likely to be neighbors themselves.

The correlations suggest that heterogeneity in risk attitudes is related to heterogeneity in clustering. However, it is known that network degree is negatively related to clustering as well. To be sure that the results are not driven by this degree/clustering relation, we regress clustering over risk attitudes, gender, degree measures, and a constant, and we report the OLS coefficients and standard errors in Table 4. Since the network formation process creates an unspecified autocorrelation in the dependent variable, we cannot interpret the OLS standard errors directly. We therefore also perform a random permutation test on the null hypothesis that risk attitudes are partially uncorrelated with clustering coefficient (controlling for other regressors), and we report the corresponding *p*-value.<sup>24</sup>

The estimation results confirm the negative relation between clustering coefficient and the variable *risk*. More risk averse individuals tend to have larger probability that their neighbors are also connected, even after controlling for gender and connectivity. As seen in the table, the effect of risk aversion is robust against the inclusion of basic connectivity measures into the regression analysis. As a final check we perform a regression that includes the age of the participants as well. Not all participants provided their age, and therefore we lose some observations. Nonetheless, the results remain unaffected.

To conclude, the regression results support our theoretical predictions that risk attitudes and network clustering are related. We would like to be cautious regarding the interpretation, though. The data were collected for other purposes than testing our theory. That is, we did not have any control over the friendship formation process of the students, and it is obvious that many factors contributed to that process. We cannot exclude that some of these factors were confounding

---

measures and risk aversion.

<sup>24</sup>The null distribution of the random permutation test is simulated by repeatedly permuting separately the male observations of the risk variable, and the female observations of the risk variable, in total a 100000 times, and reestimating the regression under the obtained null hypothesis. Doing so, we maintain the difference in risk attitudes between male and female in the simulations of the null hypothesis. The network variables are not permuted in order to maintain the network structure and the implied correlation in the network variables.

	(1)	(2)	(3)	(4)	(5)
risk	-0.039** (0.016)	-0.038** (0.016)	-0.038** (0.016)	-0.034** (0.016)	-0.037** (0.018)
gender	0.193*** (0.056)	0.197*** (0.056)	0.204*** (0.057)	0.195*** (0.056)	0.140** (0.060)
indegree	-	-0.022 (0.017)	-	-	-
outdegree	-	-	-0.012 (0.023)	-	-
degree	-	-	-	-0.042** (0.018)	-0.047** (0.020)
age	-	-	-	-	-0.002 (0.012)
constant	0.499*** (0.099)	0.553*** (0.107)	0.525*** (0.112)	0.625*** (0.111)	0.751*** (0.283)
# obs.	146	146	146	146	117
$R^2$	0.12	0.13	0.11	0.16	0.14
$p$ -value	Random permutation test. $H_0 : \beta_{risk} = 0$				
	0.020	0.027	0.022	0.040	0.046

Table 4: Estimation results from regressions of clustering on a risk variables and degree variables. OLS standard errors in parentheses. \*\*OLS coefficient significant at 5% level. \*\*\*at 1% level.

with our mechanism, but on the other hand, in this setting of friendship formation between undergraduate college students, we find it very hard to think of an alternative explanation for the relation between risk aversion and clustering. On the contrary, we believe that the most likely explanation for this relation is that students that are less risk averse, are less shy and more open to exploring new friendships outside their current group of friends. It is exactly this intuition that we envision, and that we model more formally here in this paper.

## 6 Extensions and Discussion

Our model is built on strong assumptions in order to keep it tractable. In particular the assumption that benefits are identically and independently distributed is unlikely to be true. These assumptions help us to focus on the role of risk aversion in network formation, the main objective of this paper, but it is important to understand what happens if we relax some of these assumptions. In this section, we therefore discuss variations on the standard model.

### 6.1 Common Benefits from Linking

The assumption that the benefits that agents derive from their neighbors, are idiosyncratic makes the proposed model applicable to only a few contexts. There is a large number of applications, where the potential benefits of a particular node are the same for all the members of the population. Examples of such applications can be labor market connections, where some individuals have better access to job opportunities, coauthorship networks, research networks among companies etc. In terms of our model, this would make  $u_{ij} = u_j$  for each  $i \in N \setminus \{j\}$ . The effect of this specification is that network-based linking would become more frequent, because anytime an entering node  $i$  links up to  $j$ , who formed at least one link through the network (say to a node  $k$ ),  $i$  will create a link to  $k$  with probability one if  $i$ 's risk premium is equal or larger than  $j$ 's one, since, given that  $g_{jk} = 1$  and  $u_{ik} = u_{jk} = u_k$ ,  $u_k > \bar{u} - r_i$  if  $r_i \geq r_j$ . Then, network-based search is enhanced under such a specification.

However, the main results of this paper remain unchanged. Note that the relation between clustering and risk aversion holds, since the above argument does not hold for  $r_i < r_j$ , that is if  $i$  is less risk averse than  $j$ , he does not necessarily create links to neighbors of  $j$  found through the network. As a result, there still is a positive relation between risk aversion and clustering coefficient, while in-degree would still be independent of risk attitudes. Similar considerations hold for the relation between clustering and expected payoffs. The benefits an agent earns, conditional on the way of linking are larger when linking through the network.

Since in such a specification the network-based search is enhanced the only effect would be larger clustering coefficient and a different, more unequal degree distribution. The qualitative features of generated networks would be unchanged. Therefore, our model also applies

to situations where the benefits of a particular node are the same for all the members of the population.

## 6.2 Public Knowledge

In our model we assumed that an entering agent initially has zero information about the benefits it can obtain by connecting to other agents. In reality, this is not always realistic. For example, in the academic world it is always possible to find information on other scientists by looking up their C.V.

The assumption of no prior information on the benefits of linking is easily relaxed. For example, suppose instead that an individual  $i$  has an imperfect signal about the benefits of linking to  $j$ , say  $\tilde{u}_{ij} = u_{ij} + \epsilon_{ij}$  where  $\epsilon_{ij}$  is unobserved i.i.d. noise with zero expectation. Initially, the entering agent  $i$  links to the agent  $j$  about whom it has the best signal,  $\max_j \tilde{u}_{ij}$ . Next, the agent receives a better, perhaps perfect, signal about the neighbors of  $j$ , and again the agent decides to link the best neighbor of  $j$ , or to link to the best outside option, that is the node  $k$  with the best signal that is not  $j$  or a neighbor of  $j$ ,  $\max_{k \in N \setminus \{j \cup N_j\}} \tilde{u}_{ik}$ . Agent  $i$  chooses to link to a neighbor of  $j$  if and only if

$$\max_{k \in N_{j_1}} u_{ik} > \max_{k \in N \setminus \{j \cup N_j\}} \mathbb{E}[u_{ik} | \tilde{u}_{ik}] - \tilde{r}_i$$

Naturally,  $\tilde{r}_i$  is smaller than  $r_i$ , because given that there is some initial information on non-neighbors, the risk of linking to a non-neighbor is smaller. Nonetheless, given that  $i$  still has a better signal about the neighbors of  $j$  than about non-neighbors, risk aversion again plays a role; the more risk-averse agent  $i$ , the more likely  $i$  links to a neighbor of  $j$ . This implies the same positive relation between clustering coefficient and risk aversion. Here it is irrelevant that the choice of non-neighbor is not random anymore, that is, agent  $i$  would choose the non-neighbor with the best signal, but given that the benefits and the signals are still randomly distributed, for the outside observer the choice of agent  $i$  is observably equivalent to random linking if it is not a neighbor of  $j$ .

Note that conditional on the signal  $\tilde{u}_{ik}$ , the expected benefit of linking to  $k$ ,  $\mathbb{E}[u_{ik} | \tilde{u}_{ik}]$  will be between the true benefit  $u_{ik}$  and the average benefit  $\bar{u}$  with the expected value closer to the former when the quality of the signal is better. Therefore,

$$\max_{k \in N \setminus \{j \cup N_j\}} \mathbb{E}[u_{ik} | \tilde{u}_{ik}] > \bar{u}.$$

When agents have prior information on non-neighbors, it is therefore more likely that they choose a “random” link than a friend of a friend, compared to the case where agents do not have such prior information. Moreover, the better the signal on the benefits of linking to non-neighbors,

the more likely it is that the agent chooses to link to a non-neighbor, which for the outsider is a “random” link.

This observation allows for a direct application towards the impact of internet on network formation. The emergence of internet has made it much easier to publish and obtain information on other individuals. For example, it is now standard that scientists put their C.V. on their homepage, which is then publicly available. In our model this implies that individuals have some prior knowledge on the benefit of linking to individuals, and therefore they are much less likely to link to a friend of a friend. That is, random linking should have become much more prominent than local network-based linking. Evidence provided in Fafchamps, Goyal & Van der Leij (2010) indeed suggests that this is the case.

## 7 Conclusion

This paper contributes to synergizing the game-theoretic and statistical mechanical models of network formation. We introduce a simple economic reasoning into the models of Vázquez (2003) and Jackson and Rogers (2007) and show that all the stylized facts of socially generated networks can be derived from standard microeconomic concepts.

Moreover, we show that inherent characteristics of individuals may play an important role in network formation and explaining empirical regularities of networks. Ex-ante individual heterogeneity is an issue that has been underexplored; partly in order to keep models tractable, and partly due to the believe the network formation is an endogenous process, and that understanding this endogenous process is what is most important.<sup>25</sup> However, recent work of Fowler, Dawes & Christakis (2009) suggests that ex-ante individual differences are very important as well. Comparing the network positions of identical and non-identical twins they find that about 45 % of the variation in in-degree *and* clustering coefficient can be traced back to variation in genes. They do not explore what behavioral heritable aspects lead to this variation, though. With respect to in-degree it has been argued before that ex-ante heterogeneity in technology/potential benefits may lead to a stronger attractiveness of some nodes, and therefore a higher in-degree.<sup>26</sup> Since Cesarini et al. (2009) report that a non-negligible part of risk-taking preferences of people are due to genes, we believe that we uncover a possible heritable aspect explaining social network positions of individuals: the variation in clustering can partly be traced back to variation in risk

---

<sup>25</sup>Exceptions are Galeotti, Goyal & Kamphorst (2006), and Jackson & Rogers (2005) in the economics literature, Bianconi & Barabasi (2001) and Kong et al. (2008) in the physics literature, and Burt et al. (1998) in the sociology literature.

<sup>26</sup>For example, Google has been able to outcompete other search engines and become a star on the WWW, because of its superior search technology, see Barabasi (2003) and Kong et al. (2008).



aversion among individuals.<sup>27</sup> We hope our results will enhance the exploration of the relevance of heterogeneity in social networks.

Other contribution of the paper is the welfare impact of social network position. Microaspects of social positions have for long been discussed in the sociological literature (Coleman, 1988; Granovetter, 1973 ; Burt, 1995) and only recently drawn attention from theoretical scholars (see Ballester et al. (2006) for a recent influential reference). We adhere to this discussion by providing a linkage between the profits of agents and their clustering coefficient.

As the last contribution, we provide an argument for why different network topologies arise in different socioeconomic contexts, and why they may be affected by changing environment, such as lower cost of communication and link maintenance, or external interventions or shocks that influence the benefits from linking opportunities. Hence, our model or its variations might provide an interesting tool for the evaluation of policies in networked contexts. Nevertheless, whether networks indeed react this way to external shocks is an empirical question, which we leave for future research.

## References

- [1] Arrow, K.J. "The theory of risk aversion," in *Aspects of the Theory of Risk Bearing*, Y.J. Saatio, Helsinki (1965).
- [2] Babus, A. and Ule, A. *Limited Connections: Economic Foundations for a Preferential Attachment Model*, mimeo (2008).
- [3] Ballester, C., Calvó-Armengol, A. and Zenou, Y. "Who is Who in Networks. Wanted: The Key Player." *Econometrica* 75(4): 1403-1418 (2006).
- [4] Barabasi, A.L. *Linked: The New Science of Networks*, Perseus Books (2003).
- [5] Bianconi, G. and Barabasi, A.L. Bose-Einstein Condensation in Complex Networks, *Physical Review Letters* 86(24), 5632-5635 (2001).
- [6] Bramoullé, Y. and Rogers, B. *Diversity and Popularity in Social Networks*, mimeo (2009)
- [7] Brañas, P., Cobo-Reyes, R., Jimenez, N. and Ponti G. *Eliciting social networks: an experimental guide tour*, mimeo, University of Granada (2006).
- [8] Burt, R.S. *Structural Holes: The Social Structure of Competition*, Harvard University Press (1995).

---

<sup>27</sup>A similar relation is also indirectly present in Dohmen et al. (2008). In a sample of tens of thousands of observations, they report that less risk averse individuals are more likely to migrate for work than more risk averse people. One can expect then that the social ties of the more risk averse people will more likely know each other, i.e. will be more clustered.

- [9] Burt, R.S., Jannotta, J.E., and Mahoney, J.T., Personality Correlates of Structural Holes, *Social Networks* 20, 63-87 (1998)
- [10] Campbell, A. Signalling in Social Networks, mimeo (2008).
- [11] Cesarini, D., Dawes, C., Johannesson, M., Lichtenstein, P. and Wallace, B. Genetic Variation in Preferences for Giving and Risk Taking, *Quarterly Journal of Economics*, forthcoming (2009).
- [12] Coleman, J. Social Capital in the Creation of Human Capital, *American Journal of Sociology*, 94, 95-120 (1988).
- [13] de Weerd J. Risk-sharing and Endogenous Network Formation, in Dercon, S. (ed.): *Insurance Against Poverty*, Oxford University Press 197-216 (2004).
- [14] Dohmen, T., Huffman, D., Falk, A., Jaeger, D.A., and Sunde, U. Direct Evidence on Risk Attitudes and Migration, forthcoming in *Review of Economics and Statistics* (2008).
- [15] Eeckhout J. and Munshi K. Matching in Informal Financial Institutions, *Journal of the European Economic Association*, forthcoming (2010).
- [16] Fafchamps, M., Goyal, S., and Van der Leij, M. Matching and Network Effects, *Journal of the European Economic Association*, forthcoming (2010).
- [17] Fafchamps M., and Gubert, F. Risk Sharing and Network Formation, *American Economic Review, Papers & Proceedings*, 97(2), 75-79 (2007).
- [18] Fowler, J., Dawes, C., and Christakis, N. A Model of Genetic Variation in Human Social Networks, *Proceedings of the National Academy of Sciences*, 106(6), 1720-1724 (2009).
- [19] Galeotti, A., Goyal, S., Jackson, M., Vega-Redondo, F. and Yariv L. Network Games, *Review of Economic Studies* 77(1), 218-244 (2010).
- [20] Galeotti, A., Goyal, S., and Kamphorst, J. Network Formation with Heterogeneous Players, *Games and Economic Behavior*, 54(2), 353-372 (2006).
- [21] Gallo, E. A Small World with Brokers: Heterogeneous Information in Network Formation, mimeo (2009).
- [22] Goyal, S. *Connections. An Introduction to the Economics of Networks*, Princeton Univ. Press (2007).
- [23] Goyal, S., van der Leij M.J. and Moraga-Gonzalez J.L. Economics: An Emerging Small World, *Journal of Political Economy* 114(2), 403-412 (2006).

- [24] Graham, B. Identifying Social Interactions through Conditional Variance Restrictions, *Econometrica* 76(3), 643-660 (2008).
- [25] Granovetter, M. The Strength of Weak Ties, *American Journal of Sociology* 78(6), 1360-1380 (1973).
- [26] Granovetter, M. Economic Action and Social Structure: The Problem of Embeddedness, *American Journal of Sociology* 91: 481-510 (1985).
- [27] Holt, C.A and and Laury, S.K. Risk Aversion and Incentive Effects, *American Economic Review* 92, 1644-1655 (2002).
- [28] Jackson, M. The Economics of Social Networks, in: Blundell, R. Newey, W., and Persson, T. (eds.), *Proceedings of the 9th World Congress of the Econometric Society*. Cambridge: Cambridge University Press (2006).
- [29] Jackson, M. "An Overview of Social Networks and Economic Applications," in the *The Handbook of Social Economics*, J. Benhabib, A. Bisin, and M.O. Jackson (eds.), North Holland Press (2010).
- [30] Jackson, M. *Social and Economic Networks*, Princeton University Press (2008).
- [31] Jackson, M. and Rogers, B. The Economics of Small Worlds, *Journal of the European Economics Association*, 3(2-3), 617-627 (2005).
- [32] Jackson, M. and Rogers, B. Meeting Strangers and Friends of Friends: How Random Are Social Networks, *American Economic Review* 97(3), 890-915 (2007).
- [33] Kong, J., Sarshar, N., and Roychowdhury, V. Experience versus Talent Shapes the Structure of the Web, *Proceedings of the National Academy of Sciences* 105(37), 13724-13729 (2008)
- [34] König M., Tessone C., and Zenou, Y. Games of Dynamic Network Formation, mimeo (2009)
- [35] Krishnan P. and Sciubba E. Links and Architecture in Village Networks, *Economic Journal* 119, 917-949 (2009).
- [36] Leider, S., Mobius, M., Rosenblat, T., and Quoc-Anh, D. Directed Altruism and Enforced Reciprocity in Social Networks, *Quarterly Journal of Economics*, forthcoming (2009).
- [37] Manski, C. Identification of Endogenous Social Effects: The Reflection Problem, *Review of Economic Studies* 60(3), 531-542 (1993).
- [38] Montgomery, J. Social Networks and Labor Market Outcomes: Toward an Economic Analysis, *American Economic Review* 81(5), 1408-1418 (1991).

- [39] Moffitt, R., Policy Interventions, Low-Level Equilibria, and Social Interactions, in *Social Dynamics*, edited by Steven Durlauf and Peyton Young, MIT press.
- [40] Myers C.A. and Shultz G.P. *The Dynamics of a Labor Market*, New York: Prentice-Hall (1951).
- [41] Newman, M. The Structure and Function of Complex Networks, *SIAM Review* 45(2), 167-256 (2003).
- [42] Pennock, D.M., Flake W.G., Lawrence, S., Gover, E., and Giles, C.L. Winners Don't Take All: Characterizing the Competition for Links on the Web, *Proceedings of the National Academy of Sciences* 99(8), 5207-5211 (2002).
- [43] Pratt, J. W. Risk aversion in the small and in the large, *Econometrica* 32, 122-136 (1964).
- [44] Rothenberg R.B., Potterat J.J., Woodhouse D.E., Muth S.Q., Darrow W.W. and Klondahl A.S. Social network dynamics and HIV transmission, *AIDS* 12(12), 1529-36 (1998).
- [45] Vázquez, A. Growing Network with Local Rules: Preferential Attachment, Clustering Hierarchy, and Degree Correlations, *Physical Review E* 67, 056104 (2003).
- [46] Vega-Redondo, F. *Complex Social Networks*, *Econometric Society Monograph Series*, Cambridge University Press (2007).
- [47] Vigier, A. *Network Topology: Extracting Economic Content*, mimeo (2009).
- [48] Wasserman, S. and Faust, K. *Social Network Analysis: Methods and Applications*. Cambridge, MA: Cambridge Univ. Press (1994).

## 8 Appendix

### 8.1 Perfect rationality for $m = 3$

There is an important issue concerning the rationality of players for  $m > 2$ . In particular, the clustering of the network can lead to a situation, such that anytime a node  $i$  links through network there is a positive probability that one or more of newly observed neighbors of neighbor have already been observed and not chosen in previous linking stages. A completely rational individual should take this into account. Since the entering node has not linked up to such node(s),  $i$  has information about them. This affects the mean field analysis, because the expected payoff from linking to such neighbors of neighbors is lower than the expected payoff  $i$  gets from observing and linking to someone  $i$  has no information about. In this section, we illustrate this argument formally.

Denote the nodes that  $t + 1$  connects in each linking stage as  $j_1$ ,  $j_2$  and  $j_3$ . First, note that this issue never concerns the first and last linking decision, since the first is always random, while in the last linking decision the entering nodes do not care about who they observe afterwards. Hence, for  $m = 3$  the only linking decision, in which he may observe someone, whom he has already observed, is the second one. Suppose that  $t + 1$  decides to link to a node  $j_2 \in N_{j_1}(t + 1)$  such that  $j_2 \in \arg \max_{j \in N_{j_1}(t+1)} u_{ij}$ . If so, then there is a positive probability that  $j_1$  is connected to a neighbor of  $j_2$ . It this occurs there exist a node  $l$ , an out-degree neighbor of both  $j_1$  and  $j_2$ , who  $t + 1$  observes after linking to  $j_1$  and will observe after linking to  $j_2$ . Furthermore, there is an important information in the fact that  $t + 1$  observed  $l$ , but has not connected to him.

Formally,  $t + 1$  links to a  $j_2 \in \arg \max_{j \in N_{j_1}(t+1)} u_{ij}$  if

$$\max_{j \in N_{j_1}(t+1)} u_{ij} - \sum_{s=1}^{m-1} \left\{ \frac{C(g)^s (m-s)}{m} \left[ \bar{u} - \frac{m-s}{m} \bar{u} - \frac{s}{m} \int_a^{\max_{j \in N_{j_1}(t+1)} u_{ij}} u dF(u) \right] \right\} > \bar{u} - r_{t+1} \quad (10)$$

where  $C(g)$  is the fraction of transitive triples in the population and measures the average probability that a triangle exists.  $C(g) = \frac{p_\theta(m-1)}{m^2 - \frac{p_\theta}{m}}$  for  $m = 3$  and reflects the average probability that  $t + 1$  observes  $m - 1$  new individuals and 1 individual  $t + 1$  has already been observed and have not chosen because the utility he would reported to  $t + 1$  was lower than  $\max_{j \in N_{j_1}(t+1)} u_{ij}$ . The second expression,  $\bar{u} - \frac{m-1}{m} \bar{u} - \frac{1}{m} \int_a^{\max_{j \in N_{j_1}(t+1)} u_{ij}} u dF(u)$ , reflects the expected utility loss due the fact that  $t + 1$  observes only  $m - 1$  new individuals (instead of  $m$ ), taking into account the expected utility from the individual observed and unchosen in the previous linking stage.

After some simplification of (10), we get

$$\max_{j \in N_{j_1}(t+1)} u_{ij} - \sum_{s=1}^{m-1} \frac{C(g)^s s^2}{m^s} [\bar{u} - \int_a^{\max_{j \in N_{j_1}(t+1)} u_{ij}} udF(u)] > \bar{u} - r_{t+1} \quad (11)$$

$$\max_{j \in N_{j_1}(t+1)} u_{ij} - \left[ \sum_{s=1}^{m-1} \frac{C(g)^s s^2}{m^s} \right] \left[ \int_{\max_{j \in N_{j_1}(t+1)} u_{ij}}^b udF(u) \right] > \bar{u} - r_{t+1}. \quad (12)$$

As a result, the probability that  $t + 1$  links through network search in its second linking decision is

$$p^{2nd}(r_{t+1}) = 1 - F \left[ \bar{u} - r_{t+1} + \sum_{s=1}^{m-1} \frac{C(g)^s s^2}{m^s} \int_{\max_{j \in N_{j_1}(t+1)} u_{ij}}^b udF(u)^m \right] > p(r_{t+1}).$$

Then, the expected probability of node  $i < t + 1$  to receive a new link in  $t + 1$ , analogous to expression (4), is

$$\frac{dd_i(t)}{dt} = \frac{1}{t} + \left[ \frac{1 - p^{2nd}(r_{t+1})}{t} + p^{2nd}(r_{t+1}) \frac{d_i(t)}{t} \frac{1}{m} \right] + \left[ \frac{1 - p(r_{t+1})}{t} + p(r_{t+1}) \frac{d_i(t)}{t} \frac{1}{m} \right]. \quad (13)$$

The only difference between (4) and (13) is the intermediate term. The effect of perfect rationality is to enhance global search. In (13), it increases the probability of receiving a random link and decreases the probability of receiving a link from a neighbor of a neighbor. The overall effect, hence, depends on the current in-degree of each node. In particular, nodes with large in-degrees will be negatively affected by perfect rationality, because a large fraction of nodes they receive is through local linking. Nodes with low connectivity, on the other hand, benefit from the form of rationality we model here, since they receive almost no links through network anyway. Formally,  $\frac{\partial dd_i(t)}{\partial p(r_{t+1})} = \frac{d_i(t)}{t} \frac{1}{m} - \frac{1}{t} > 0$  if  $d_i(t) > m$ . Hence, the effect of the rationality discussed here is the following:

If  $d_i(t) > m$ ,  $i$  receives a link with lower probability than in the original specification,

If  $d_i(t) = m$ ,  $i$  is unaffected by the new specification,

If  $d_i(t) < m$ ,  $i$  receives a link with higher probability in the new specification.

In sum, the effect of the perfect rationality considered here is to enhance random search. This will affect the in-degree of each agent as a function of his connectivity. From the global point of view, the tails of the degree distribution shift down, more frequent global search lowers the clustering coefficient, and the distances would shrink.

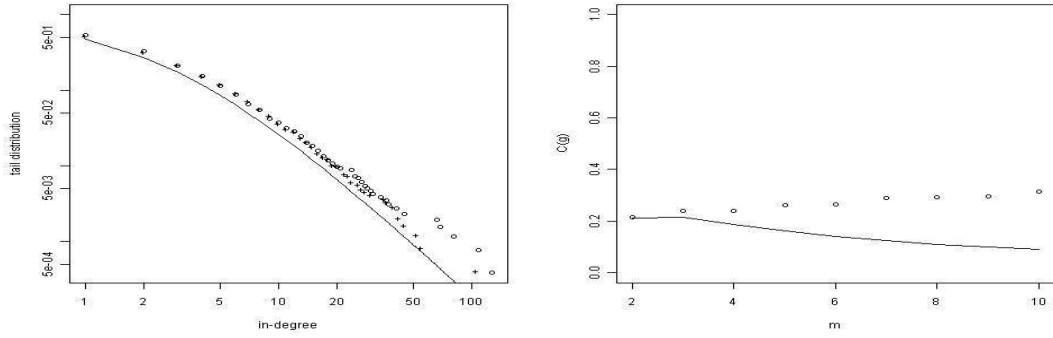


Figure 4: (a) The predicted (solid line) and simulated (crosses and circles) degree distributions. (b) The average clustering coefficient. The solid line is the predicted lower bound from Proposition 4.

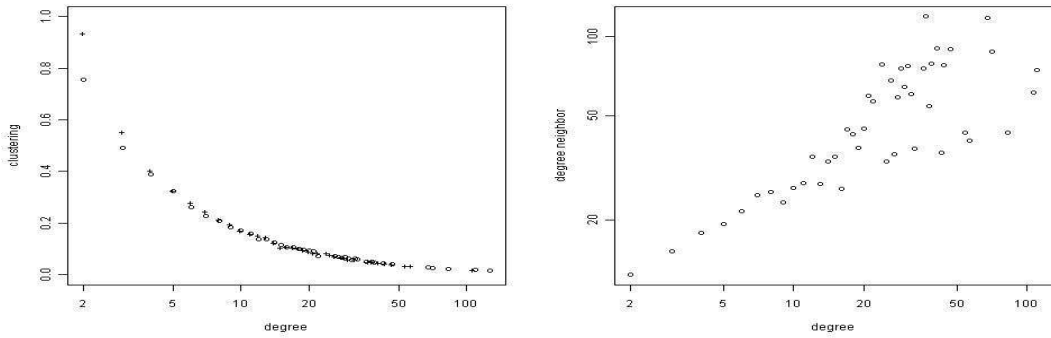


Figure 5: (a) Relation between degree and clustering., (b) Assortativity

## 8.2 Simulations

Results in the main text are based on the mean field approach. Therefore, it only provides approximations. In order to check the accuracy of the predictions, we also run simulations of the model and match them with the approximations. Moreover, there are three other stylized facts of empirical social networks, that we wish to verify in our model through means of simulations: short network distances, positive assortativity and the negative clustering/degree correlation. The reader is referred to Goyal (2007) or Jackson (2008) for formal definitions and evidence.

The simulation assumes that  $u_{ij}$  is drawn from a standard uniform distribution. Agents have a risk premium of either  $r_L = 0$  and  $r_H = .25$  with equal probability ( $\theta = .5$ ). We initially set  $m = 2$  and we generate a network of 5000 nodes. Figures 4a through 5b contain various plots for four of the five stylized facts of observed social networks that our model predicts. The distances are only discussed at the end of this section.

Figure 4a contrasts the predicted in-degree distribution with the simulated one. They seem to be very similar. Hence, we can conclude that the mean field approach approximates very well the degree distribution generated by the model. Moreover, we distinguish between the high (crosses) and low (circles) risk premium types. Figure 4b plots the average clustering coefficient (in terms of fraction of transitive triples) for several values of  $m$ . The figure shows that the clustering coefficient is indeed positive and lies above the lower bound derived in Proposition 1. In fact, the simulated values of the average clustering coefficient are well above and increase over  $m$ , suggesting that the more connections the agents of our model form, the more clustered the network becomes.

Figure 5a shows that our model also generates the negative clustering/degree correlation. (Here the clustering coefficient is measured ignoring the direction of the links.) The x- and y-axes plot the degree and clustering, respectively, and there is an obvious negative relation between the two variables in the graph. Moreover, in this plot we also make a distinction between the clustering coefficient of high risk-averse agents and low risk-averse agents. The plot shows that the clustering coefficient is substantially higher for high-risk averse agents, in particular for low degree values, where the majority of the nodes lies (see Figure 2a).

To check for assortativity, we draw a plot with the degree of a node on the x-axis and the average degree of an out-neighboring node on the y-axis. This plot shows a positive correlation. Nodes with high degree have also high degree neighbors, indicating positive assortativity. We also compute the degree correlation, which is .260, well above zero.

To check for network distances, we compute the average networks distance and the largest distance between two nodes in the resulting simulated network, again ignoring directions. The obtained values are 5.74, and 13, respectively, thus of the order of  $\ln(n)$ .

### 8.3 Proofs

**Proof of Theorem 1.** Concerning part (i), (4) can be rewritten as  $\frac{dd_i(t)}{dt} = a\frac{d_i(t)}{t} + \frac{b}{t} + c$ , where  $a = \frac{(m-1)p_\theta}{m}$ ,  $b = [1 + (m-1)(1-p_\theta)]$ , and  $c = 0$ . Given that  $m > 1$  and that  $p(r_L) > 0$  ensures  $p_\theta > 0$ , the first part of Lemma 1 in Jackson and Rogers (2007) applies.

As for the clustering coefficient, consider an agent  $i$ . Each agent initially creates 1 random link and afterwards faces  $m - 1$  decisions to either link locally or search randomly. The first case occurs with probability  $p(r_i)$ . Thus the agent has on average  $p(r_i) \times (m - 1)$  links that are based on network search. If  $k$  is found through network search, then it must be through  $j$  to whom  $i$  is also linked. So we have  $g_{ij} = g_{jk} = g_{ik} = 1$ . Each such network-searched link creates *at least* one transitive triple. Given that the amount of triples for which  $g_{ij} = g_{jk} = 1$  equals  $m^2$  and  $E[p(r_i)] = p_\theta$ , we obtain (8). ■



**Proof of Proposition 2.** The first part follows directly from (5), which does not depend on the risk aversion of individual  $i$ . The second follows from that  $p(r_H) > p(r_L)$ , thus high types are more likely to search through the network. Each time an agent  $i$  decides to link through the network at least one transitive triple is created in its neighborhood, whereas the probability that a transitive triple is created after a random linking decision converges to 0 for large  $t$ . The proposition directly follows. ■

**Proof of Proposition 3.** Note that

$$E\left[\sum_{j \in N_i} u_{ij} | r_i = r\right] = \bar{u} + (m-1) \left\{ [1 - p(r)]\bar{u} + p(r)E\left[\max_{j \in N_i} u_{ij} \mid \max_{j \in N_i} u_{ij} > \bar{u} - r\right]\right\}.$$

Since

$$\begin{aligned} & [1 - p(r_H)]\bar{u} + p(r_H)E\left[\max_{j \in N_i} u_{ij} \mid \max_{j \in N_i} u_{ij} > \bar{u} - r_H\right] \\ &= \int_a^{\bar{u}-r_H} \bar{u} dF(u)^m + \int_{\bar{u}-r_H}^b u dF(u)^m \\ &= \int_a^{\bar{u}-r_H} \bar{u} dF(u)^m + \int_{\bar{u}-r_H}^{\bar{u}-r_L} u dF(u)^m + \int_{\bar{u}-r_L}^b u dF(u)^m \\ &< \int_a^{\bar{u}-r_H} \bar{u} dF(u)^m + \int_{\bar{u}-r_H}^{\bar{u}-r_L} \bar{u} dF(u)^m + \int_{\bar{u}-r_L}^b u dF(u)^m \\ &= [1 - p(r_L)]\bar{u} + p(r_L)E\left[\max_{j \in N_i} u_{ij} \mid \max_{j \in N_i} u_{ij} > \bar{u} - r_L\right], \end{aligned}$$

it directly follows that  $E[\sum_{j \in N_i} u_{ij} | r_i = r_L] > E[\sum_{j \in N_i} u_{ij} | r_i = r_H]$ . ■

**Proof of Theorem 4.** For any existing link  $ij$ , define  $L_{ij}$  an indicator that is 1 if  $i$  found  $j$  through local network search, and 0 if found by random search. For any existing link  $ij$ , let

$$\bar{u} \equiv E[u_{ij} | L_{ij} = 0] = \int_a^b u dF(u)$$

and

$$\begin{aligned} \tilde{u} &\equiv E[u_{ij} | L_{ij} = 1] \\ &= P[r_i = r_H | L_{ij} = 1]E[u_{ij} | L_{ij} = 1, r_i = r_H] + P[r_i = r_L | L_{ij} = 1]E[u_{ij} | L_{ij} = 1, r_i = r_L] \\ &= P[r_i = r_H | L_{ij} = 1] \int_{\bar{u}-r_H}^b u dF(u)^m + P[r_i = r_L | L_{ij} = 1] \int_{\bar{u}-r_L}^b u dF(u)^m \end{aligned}$$

denote the expected payoff of linking up to a random individual and a neighbor of a neighbor, respectively. Let  $L_i = \sum_{j \in N_i} L_{ij}$ . Then

$$E\left[\sum_{j \in N_i} u_{ij} | C_i = c\right] = \bar{u}E[m - L_i | C_i = c] + \tilde{u}E[L_i | C_i = c].$$

Naturally,  $\tilde{u} > \bar{u}$ . Hence, nodes who search more often locally will tend to earn higher payoffs.

To complete the proof, we now show that  $E[L_i|C_i = c]$  is weakly increasing in  $c$  for  $m < 5$ . Note that each node  $i$  can close at most  $\sum_{j=1}^{m-1} j = \frac{m(m-1)}{2}$  triples in period  $i$  since links are directed and entering nodes can only link up to older nodes.

Suppose  $m = 2$ . Then  $C_i$  is (approximately) 0 or positive, depending on whether the second link was random or via a friend of friend. Hence,  $E[L_i|C_i = 0] = 0$  and  $E[L_i|C_i > 0] = 1$ .

Next, suppose  $m = 3$  and denote the outdegree neighbors of  $i$  as  $j_1, j_2, j_3$ . Then  $L_i$  is at most 2, and at most 3 triples can be closed. The first link does not close triples, the second link closes one triple if  $L_{ij_2} = 1$ , and the third link closes one or two triples if  $L_{ij_3} = 1$ . Hence,  $E[L_i|C_i = 0] = 0$ ,  $E[L_i|C_i = 1/9] = 1$ ,  $E[L_i|C_i \geq 2/9] = 2$ .

Finally, suppose  $m = 4$  and let  $i$  have outdegree neighbors  $j_1, \dots, j_4$ . Then  $L_i$  is at most 3, and at most 6 triples can be closed. We have,  $E[L_i|C_i = 0] = 0$ ,  $E[L_i|C_i = 1/16] = 1$ ,  $E[L_i|C_i = 2/16] = 2$ ,  $E[L_i|C_i = 3/16] \in (2, 3)$ , and  $E[L_i|C_i \geq 4/16] = 3$ . ■

**Proof of Proposition 5.** With utility (9), the certainty equivalent  $y$  of linking to a random agent solves

$$U_i(y) = E_F[U(x)].$$

Hence, with normal distribution of payoffs,  $y$  is given by

$$y = \bar{u} - \frac{\rho_i}{2}\sigma^2$$

The risk premium is defined as  $r = \bar{u} - y = \frac{\rho_i}{2}\sigma^2$ , leading to

$$p\left(\frac{\rho_i}{2}\sigma^2\right) = 1 - F\left(\bar{u} - \frac{\rho_i}{2}\sigma^2\right)^m.$$

With a normal distribution,  $F(\bar{u} - \frac{\rho_i}{2}\sigma^2) = \Phi(-\frac{\rho_i}{2}\sigma)$  where  $\Phi(\cdot)$  is the cumulative distribution function of standard normal distribution. Since it is decreasing with  $\sigma^2$  and independent of  $\bar{u}$ , the proposition directly follows. ■

**Proof of Theorem 6.** Since  $p_\theta = \theta p(r_H) + (1-\theta)p(r_L)$ , it follows from Proposition 5 that ratio of the global and local search probabilities  $\frac{1+(1-p_\theta)(m-1)}{p_\theta(m-1)}$  decreases with  $\sigma^2$  and is independent of  $\bar{u}_F$ . It then directly follows from Jackson and Rogers (2007), Theorem 6, that the degree distribution of  $g'$  second order stochastically dominates the degree distribution of  $g$  whenever  $\sigma^2 > \sigma'^2$ , independently of  $\bar{u}$  and  $\bar{u}'$ . Moreover, since  $p_\theta$  increases with  $\sigma^2$  and remains constant with  $\bar{u}$ , the results on the clustering coefficient directly follow. ■