serie ad

WP-AD 2010-02

Strong ties in a small world

Marco J. van der Leij and Sanjeev Goyal

Ivie

Working papers

Los documentos de trabajo del Ivie ofrecen un avance de los resultados de las investigaciones económicas en curso, con objeto de generar un proceso de discusión previo a su remisión a las revistas científicas. Al publicar este documento de trabajo, el Ivie no asume responsabilidad sobre su contenido.

Ivie working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication. Ivie's decision to publish this working paper does not imply any responsibility for its content.

La Serie AD es continuadora de la labor iniciada por el Departamento de Fundamentos de Análisis Económico de la Universidad de Alicante en su colección "A DISCUSIÓN" y difunde trabajos de marcado contenido teórico. Esta serie es coordinada por Carmen Herrero.

The AD series, coordinated by Carmen Herrero, is a continuation of the work initiated by the Department of Economic Analysis of the Universidad de Alicante in its collection "A DISCUSIÓN", providing and distributing papers marked by their theoretical content.

Todos los documentos de trabajo están disponibles de forma gratuita en la web del Ivie http://www.ivie.es, así como las instrucciones para los autores que desean publicar en nuestras series.

Working papers can be downloaded free of charge from the Ivie website http://www.ivie.es, as well as the instructions for authors who are interested in publishing in our series.

Edita / Published by: Instituto Valenciano de Investigaciones Económicas, S.A.

Depósito Legal / Legal Deposit no.: V-725-2010

Impreso en España (enero 2010) / Printed in Spain (January 2010)

WP-AD 2010-02

Strong ties in a small world*

Marco J. van der Leij and Sanjeev Goyal**

Abstract

This paper examines the celebrated "Strength of weak ties" theory of Granovetter (1973). We formalize the theory in terms of two hypotheses: one, for any three players with two links present, the probability of a third link being present is increasing in the strength of the two ties, and two, the removal of a weak tie increases average distance in the network more than the removal of a strong tie.
We test these hypotheses using data on the network of coauthorship among economists.
We find support for the hypothesis of transitivity of strong ties, but we reject the hypothesis that weak ties reduce distance more than strong ties do.
We then identify two general features of networks which explain these findings: significant inequality in the distribution of connections across individuals and stronger ties among individuals who have more connections.

Keywords: network; strength of weak ties; core-periphery; co-authorship network

JEL Classification: A14, D85, Z13

---

* We thank Francis Bloch, Vincent Buskens, Antoni Calvó-Armengol, Dennis Fok, Matthew Jackson, Willemien Kets, Marcel Fafchamps, Alan Kirman, Michael Kosfeld, Richard Paap, Werner Raub and seminar participants at NAKE Research Day 2004, Erasmus University, Marseille, Utrecht, Mannheim and ICTP (Trieste), for helpful discussions and comments. Marco van der Leij would like to thank the Spanish Ministry of Science and Innovation (grant SEJ2007-62656) for financial support.
** M.J. van der Leij: Universidad de Alicante. Contact author: vanderleij@merlin.fae.ua.es. S. Goyal: University of Cambridge and Christ's College.

1 Introduction

This paper examines the celebrated strength of weak ties hypothesis. In a seminal paper, Granovetter (1973) argued that weak ties in a social network (one's acquaintances) are more important for information dissemination than strong ties (one's close friends). Consequently, individuals and societies with few weak ties are disadvantaged. Not only do they receive news or important information later than others, but they are also less able to organize themselves.

In short, Granovetter's argument proceeds as follows: strong ties are *transitive*; this means that if two individuals have a common close friend, then it is unlikely that they are not related at all. Therefore, strong ties cover densely knitted networks, where a 'friend of my friend is also my friend'. On the other hand, weak ties are much less transitive, and therefore weak ties cover a larger but less dense area. Weak ties are more likely to be *bridges*: crucial ties that interconnect different subgroups in the social network. The suggested network structure implies that information from a strong tie is likely to be very similar to the information one already has. On the other hand, weak ties are more likely to open up information sources very different from one's own. Also, a society with few weak links is likely to be scattered into separate cliques with little communication between cliques. Granovetter's arguments may be viewed as an aspect of a more general theory of social structure: the idea that the social world is a collection of groups which are internally densely connected via strong links, and there are a few weak links connecting the groups.

In his original paper Granovetter (1973) provides some evidence in support of his theory. He finds that in a survey of recent job changers living in a Boston suburb 27.8% of the respondents who found their new job through a contact said they rarely saw this contact, while only 16.7% of the respondents who found their new job through a contact indicated that they frequently saw their contact. Thus job seekers mainly receive information on job openings through weak ties. (Granovetter, 1973, 1995). In later surveys (Granovetter, 1983, 1995) Granovetter provides more empirical evidence. However, empirical research that directly tests the structural content of the strength of weak ties hypotheses using a large data set appears to be lacking.[1] There are two main reason for why previous researchers have focused on small network samples or indirect tests. First, data on large networks was not available, and, second, until recently computers lacked the computational power to tackle data sets of tens of thousands of actors. However, with the availability of large data sets and the advances in computing power, these problems can now be overcome.

In this paper we use one such data set, a data set of co-authorship relations of economists publishing in scientific journals, to explore the validity of Granovetter's hypothesis. This

[1] After this paper was first written in May 2005, we became aware of Onnela et al. (2007), which is discussed below.

data set contains about 150,000 articles from 120,000 economists collected over a 30 year period. This data set is, in particular, convenient due to its large coverage and the fact that it allows us to define the existence and the strength of a tie objectively and unambiguously; a tie between authors A and B exists when they have published an article together, and the strength of their tie is measured by the number of articles A and B have jointly published.

We formalize the structural part of Granovetter's theory in terms of two testable statements: first, that strong ties are transitive; and, second, that weak ties are more important in reducing shortest path lengths between actors. Our principal findings with regard to these hypotheses are as follows: we find strong support for the first hypothesis that strong ties are transitive. However, we reject the hypothesis that weak ties are more important in reducing shortest path lengths.

These findings are surprising and lead us to explore more closely the distribution of strong and weak links in the network. The key finding here is that 1) the network is connected by a set of hubs (nodes with a large number of ties), and 2) ties between actors with many ties are relatively stronger. These two properties help explain the relatively greater criticality of strong ties. To illustrate this we present examples of networks around specific individuals in our data set in Figure 1-2. These local networks suggest that strong ties often lie in the center of the co-author network.

The classical view has been that society consists of different communities with strong ties within and weak ties across communities.[2] Our paper provides evidence for the existence of societies in which the strong ties are located in the core. In such societies, strong ties are more important for bridging the network than weak ties. We believe that our findings describe a *general* aspect of social structure in professions and in organizations. This social structure arises out of the fundamental time dimension in the creation of links. Individuals create links and strengthen them over time. At any point in time, there will exist a demographic profile of young and old individuals. The young are likely to have relatively fewer ties and these ties are likely to have lower strength, compared to the old. In particular, older individuals will on average be better connected, their links will be with other well connected older individuals, and these links will be (relatively) stronger. Given the key role of highly linked individuals in the social structure, it follows that the strong ties will be more important for connecting up the network than weak ties. Hence, the strength of strong ties in a small world.

Our paper is a contribution to the empirical study of networks.[3] Theoretical work has highlighted the impact of networks on unemployment and inequality (Calvo-Armengol

[2]These ideas also find expression in a variety of spheres ranging from the study of personal identity in social philosophy (see e.g., Avineri and de Shalit, 1992; Taylor, 1969) to the diffusion of innovations via social communication (see e.g., Rogers 1995).

[3]For work on networks in sociology see Granovetter (1995), Moody (2004), and Watts (1989). The research in physics is surveyed by Albert and Barabási (2002), Dorogovtsev and Mendes (2002), Guimera et al. (2005), and Newman (2003); for work in economics, see Goyal (2007) and Jackson (2009).

and Jackon, 2004), diffusion of information and innovation (Bala and Goyal, 1998; Golub and Jackson, 2008), the competitiveness of firms (Goyal and Moraga, 2001), criminal activity (Bellester, Calvo-Armengol and Zenou, 2006), and social cooperation (Bandiera, Barankay, and Rasul, 2006). For general surveys of this work see Goyal (2007) and Jackson (2008). These findings motivate a closer examination of the empirical structure of social and economic networks. Granovetter's (1973) strength of weak ties is perhaps the best known single hypothesis on social networks. However, in spite of its great fame, we could locate only three empirical papers which directly assess this hypothesis, Friedkin's (1980), Borgatti and Feld (1994), and Onnela et al. (2007). We now relate our work to them.[4]

We first discuss Friedkin's (1980) work. His research is based on a survey of 136 faculty members in seven biological science departments of a single university. He defines a tie between A and B to be strong if A and B have discussed both their current research together, while a tie is weak if only either A or B's research has been discussed by the two. In his analysis Friedkin confirms the strength of weak ties theory on five hypotheses. However, the hypothesis that 'weak ties create more and shorter paths' is based on only a small simulation with 4 replications (Friedkin 1980, p.417, Hypothesis 4 and Footnote 6), and hence this test is very limited. In our analysis, this hypothesis is crucially rejected. This difference in results leads us to an investigation of general properties of networks under which the strength of weak ties may not hold.

Borgatti and Feld (1994) propose another test for the strength of weak ties theory based on the overlap and non-overlap of the neighborhoods of dyad members[5]. They apply their procedure to Zachary's (1977) Karate club data. For this particular data set they reject the hypothesis that the dyad members of weak ties have larger non-overlapping neighborhoods. This finding is echoed in our work.

Onnela et al. (2007) study a large scale network of mobile phone conversations. They find support for the strength of weak ties theory. In particular, when they remove links from the mobile communication network in order starting with the weakest links first, they observe a sudden collapse in the cohesion of the network, something that does not appear when the strongest links are deleted first. This is in contrast to our results. How do we explain the difference? The key difference is that academics reflects a strong hierarchy which is related to time: senior academics have more links than younger ones and typically their ties with other senior colleagues are stronger than ties among young academics or between young and old academics. This creates the key positive correlation between degree and strength of ties. This positive correlation is unlikely to obtain in the network of telephone conversations.[6]

[4]In an illuminating recent paper, Yakubovich (2005) examines the functional interpretation of the strength of weak ties hypothesis. He finds, controlling for person specific fixed effects, that on average, a weak tie of a person is more likely to yield a job as compared to a strong tie.

[5]Their test procedure requires the full adjacency matrix of actors. This makes the procedure for our networks of more than 80,000 actors in the 1990's infeasible, and we therefore use a different approach.

[6]While we have been unable to directly test this hypothesis on the mobile phone network due to data

The rest of the paper is outlined as follows. In Section 2 we lay out the main hypotheses of Granovetter's paper. In Section 3 we present our main empirical results. Then in Section 4 we give an explanation for the conflicting results, focusing on two properties of networks: an unequal distribution of links and the relation of degree to the strength of the links. Section 5 concludes.

2 The strength of weak ties: main hypotheses

We first recapitulate the arguments of Granovetter's 'strength of weak ties' theory. The theory consists of a logical sequence of hypotheses on structural network features, starting with hypotheses on the microstructure of networks and leading to hypotheses on the macrostructure of networks. In our test we focus on two hypotheses, one on the micro-level and one on the macro-level, which capture the essence of the 'strength of weak ties' theory.

Intuitively the strength of a social tie is a "(...) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie." (Granovetter, 1973, p. 1361). Granovetter argues that the strength of a tie is directly related to the network structure. In particular, consider a triad of three individuals A, B and C in a social network in which AB and AC are tied. We call such a triad a connected triple. Now, consider the likelihood that there is also a tie between B and C. In that case, the triad is called completed, and the connected triple is called transitive. Granovetter argues that triad completion is more likely if AB and AC are *strong*, because "(...) if C and B have no relationship, common strong ties to A will probably bring them into interaction and generate one" (Granovetter, 1973, p. 1362). Further, since AB and AC have a strong tie, B and C are likely to be similar to A and therefore similar to eachother, and this facilitates the formation of a tie between B and C.

According to Granovetter: "(...) the triad which is most *unlikely* to occur, (...) is that in which A and B are strongly linked, A has a strong tie to some friend C, but the tie between C and B is absent."(Granovetter, 1973, p. 1363). This leads to hypothesis 1.

Hypothesis 1: *For a set of three players with two links present, the probability of triad completion is much higher if the links are strong as compared to the case where the links are weak.*

The second step in the theory is then to relate Hypothesis 1 to the presence of shortest paths and bridges. These concepts are defined as follows; a *path* between i and j is a sequence of distinct actors $i = i_0, i_1, \ldots, i_{z-1}, i_z = j$ for which i_{k-1} and i_k, $k = 1, \ldots, z$ are tied. Here z is the path length; recall that the distance between two nodes i and j in

restrictions, personal communication with Dr. J-P Onnela suggests that there is NO positive correlation between degree and strength of ties in the mobile telephone data set.

network g refers to the shortest path length between these nodes in network g. A *bridge* is then a tie in a network which provides the only path between some actors i and j. A *local bridge* is a tie between i and j in which the length of the shortest path between i and j other than the tie itself is larger than 2. Figure 3a and 3b illustrates a bridge and local bridge.

Granovetter then shows that if Hypothesis 1 is true and if every person has several strong ties, then *a strong tie is unlikely to be a (local) bridge*, or equivalently, most (local) bridges are weak ties. Since strong links are transitive, they are unlikely to be critical. This does not apply to weak links, since many of the triads with weak links are likely not to be completed.

This statement leads to our second hypothesis. Granovetter says, "The significance of weak ties, then, would be that those which are local bridges create more, and shorter, paths. Any given tie may, hypothetically, be removed from a network; the number of path broken and the changes in average path length resulting between arbitrary pairs of points (with some limitation on length of path considered) can then be computed. " (Granovetter, 1973, p. 1366). Then the contention is that the removal of a weak tie would, on average, break more paths and increase average path length more than the removal of a strong tie. For computational reasons we concentrate on shortest paths, leading to the following hypothesis, separated into two parts.

Hypothesis 2A: *the removal of a random weak tie from the network would break more shortest paths between actors than the removal of an arbitrary strong tie.*

Hypothesis 2B: *the removal of a random weak tie from the network would increase average distance in the network more than the removal of an arbitrary strong tie.*

Hypothesis 2 closes Granovetter's theory as far as it concerns the structural network features.

In his original paper, Granovetter (1973) is also concerned with the implications of Hypothesis 2 for diffusion, social mobility, and political organization. These potential effects of networks are clearly important and have been the focus of extensive research in different disciplines. The present paper will however be concerned with the *structural* aspects of the strength of weak ties.[7]

3 Testing the hypotheses

We test the hypotheses using data on co-authorship networks of economists publishing in scientific journals. The data is derived from EconLit, a bibliography covering economic

[7]For pioneering work on the effects of social networks on the diffusion of innovations, see Coleman (1966) and Hägerstrand (1969). For a survey of this work, see Rogers (1995).

journals, and it is split up in three decades, 1970-1979, 1980-1989 and 1990-1999. Each node is an economist[8]. Two economists are linked whenever they wrote an article together either as the only two authors, or together with a third author. Note that an article with three co-authors automatically results in a closed triad. EconLit does not provide full information on author names of articles with 4 or more authors, hence these articles are excluded from the analysis.[9]

We measure the strength of a tie by the number of articles over a decade of which the two associated economists were co-authors. Our measure of strength has the merit that it does not rely on subjective interpretations of respondents. It is objective and easily and directly measurable. Furthermore, ties based on this measure are symmetric and positive. Granovetter's theory is expressed in terms of symmetric positive ties (Granovetter, 1973, footnote 2), and our measure is therefore directly applicable to the theory.

We now make some remarks on our interpretation of the coauthor network as a social network and the idea of using number of coauthored papers as a measure of strength of a tie between two authors. It seems to us that in the social sciences – where collaborative research typically involves one or two collaborators – it is very likely that the co-authors also know each other personally. Our second observation concerns the relation between number of co-authored papers and the strength of tie. Even though the mere existence of a coauthor tie suggest that the relation must be quite strong, it seems clear that coauthoring *more* papers is suggestive of *more* sustained interaction, and an even *stronger* tie. Hence, the number of co-authored papers is a good first measure of strength of tie. On the other hand, we are conscious that there is no direct causal relation between number of papers and strength of tie. This brings us to the more general point that is worth emphasizing. Our arguments identify structural features of a network under which the strength of weak ties hypothesis will not hold. The interpretation of the coauthor network as a social network is proposed by way of motivation. Thus, the formal validity of the general argument we are developing in the paper depends on characteristics of a network structure and it does not depend on the interpretation of the coauthor network as a social network.

Before proceeding with the tests we first provide some general properties of the co-author network over the period 1970-2000. These are summarized in Table 1.[10] The co-author network of economists has some characteristics which are common to many large networks and which have been explored extensively by physicists (Newman, 2003). That is,

[8]Author names appear in different forms, often without middle names or with only initials mentioned, and thus an extraction rule is necessary to match the different names to the correct author as best as possible. The rule we used can be obtained from the corresponding author.

[9]In the EconLit database 77% of all the co-authored articles had 2 authors, 19% had 3 authors, and 4% had 4 or more authors. Results presented in Van der Leij (2006, pp. 53-56) show that the structure of the co-author network is qualitatively unaffected when only 2-authored papers are considered or when (for a subset of the data) articles with 4 or more authors are included.

[10]The reported numbers differ slightly from the numbers reported in Goyal et al. (2006); this is due to differences in the rule that distinguishes author names.

many economists (up to 40% of the total population of economists) are part of one large cluster of connected nodes, called the 'giant component', while the second largest cluster is extremely small relative to the size of the whole network. Further, average distance is small, and the fraction of connected triples that are transitive, known as the clustering coefficient, is relatively high. Further, when we focus on the trends in the properties we observe that average distance has become much smaller, while the giant component has become much larger. For details of this study, see Goyal et al. (2006).

3.1 The transitivity of connected triples

We now turn to hypothesis 1. First, we define a tie between A and B as a *weak tie* if the strength of the tie is smaller than some threshold, $c_S = \{2, 3, 5\}$, and a *strong tie* otherwise. Next, for each of the three networks we gather the set of ordered triples with actors A, B and C for which there is a tie between A and B and between A and C. We partition this set of triples ABC into three subsets in which:

1. both A and B, and A and C have a *weak* tie (weak-weak);
2. if A and B have a weak tie, then A and C have a strong tie (weak-strong);
3. both A and B, and A and C have a *strong* tie (strong-strong).

For these three subsets we compute the fraction of triples that are completed, that is, for which there is also a tie between B and C. Table 2 shows the results for the three data sets of the 1970's, 1980's and 1990's and strength thresholds of 2, 3 and 5.

Moreover, the numbers in Table 2 clearly suggest that Hypothesis 1 is true in the co-author network. For instance take the definition that a link is strong if there are 5 or more papers involved. For the 1990's network, we find that triple completion occurs in (approximately) 16% of the cases where a weak-weak triple exists, while it occurs in almost 50% of the cases where a strong-strong triple exists. χ^2-tests suggest that these differences are statistically significant. Thus triples with two strong links are more likely to be transitive.

While the χ^2-test statistics are indicative, conclusions cannot be directly drawn from them. The problem is that these tests require independence of observations, while obervations in our datasets of connected triples are far from independent. In fact, the triples in the network overlap each other. Hence, the same author names appear over and over again in different triples. For example there are 243931 connected triples in the network of the 1990s, allowing for 3×243931 different author names. However, only 45110 different author names appear in these triples.[11]

[11] Our test also does not allow us to differentiate between the effect of a 2-authored paper and the effect of a 3-authored paper, which automatically creates a transitive triple. However, results in Van der Leij (2006, pp. 53-56) suggest that results are robust for the exclusion of papers of three authors.

In order to address the problem of 'overlapping triples' we perform a more rigorous analysis in the form of a logistic regression.[12] We test whether for an ordered connected triple ABC the probability of a link between B and C is increasing in the strength of the ties AB and AC. More precisely, we estimate $\theta = \{\alpha, \beta_1, \beta_2\}$ in the following logistic regression:

$$\Pr\left(s_{BC} \geq 1 | s_{AB} \geq 1 \wedge s_{AC} \geq 1\right) = \Lambda\left(\alpha + \beta_1 \frac{\ln s_{AB} + \ln s_{AC}}{2} + \beta_2 \left|\ln s_{AB} - \ln s_{AC}\right|\right), \tag{1}$$

for each ordered connected triple ABC. Here $\Lambda(\cdot)$ is the logistic function and s_{PQ} is the strength of a tie between P and Q.[13]

In order to obtain a data set with independent observations, we propose to thin the data set, that is, to perform a standard logistic regression on *random subsamples* from the set of all ordered connected triples ABC. The motivation for this method is that a fixed set of randomly chosen actors is unlikely to influence each other if the network becomes very large. For example, while in a class room of 30 students each actor might have some direct or indirect influence on their classmates, in a social network comprising the whole world one random actor in one part of the world is unlikely to have any dependence on a random actor in a different part of the world. Thus, if the size of a random subsample of ordered triples is large, but relatively very small compared to the size of the set of all ordered connected triples, then the observations in the subsample are close to independent while large-sample asymptotic results still hold. That is, estimators are consistent and the standard errors are correctly specified.

The proposed method has a drawback, however, since by taking a random subsample we do not make use of all the information in the data, and hence, we dramatically reduce the power of the test. This implies that there is a trade off between efficiency and accuracy in the choice of the size of the subset. If the sample is too small, then the test has not enough power to give significant results, while if the sample is too large overlapping triples would distort the statistical results.

We apply this subsampling method on the combined data set of ordered connected triples ABC in the 1970's, 1980's and 1990's. This combined data set contains 357198 observations. We randomly take 1000 observations from this combined data set and we perform

[12] An alternative approach would be to approximate the distribution of a triad census in which the census differentiates between different strengths of ties, and to base a statistical test on this approximate distribution. Approximate distributions under the assumption of random network formation have been derived for directed networks, and for networks allowing for different actor types, see Wasserman and Faust (1994, Ch. 14.3) for an overview. As far as we know the distribution of a triad census has not been derived in the case of undirected networks with different types of links.

[13] This method is closely related to a pseudolikelihood estimation of a variant of the ERGM/p^*-model (Frank and Strauss, 1986), although in our case the dependent variable is a triple instead of a dyad. As is well known, the pseudolikelihood may lead to incorrect inference, due to the dependence between the links. We approach this problem by taking random subsamples, see the discussion in this Section.

the logistic regression as in (1) in which we also include dummies for the three decades.[14] With this subsample size the power of the test is still quite high, while the overlap of triples is starkly reduced.[15] We repeat this procedure a 10000 times, and we report estimation results in Table 3.

Clearly, the variable $AVGSTRENGTH$ has a significantly positive coefficient ($p = .027$). Thus the stronger the ties AB and AC, the more likely it is that B and C are tied. To give some intuition of the numbers, if AB and AC are both based on one coauthored paper, then the estimated probability that there is a link BC is .189 in the 1970's, .177 in the 1980's, and .166 in the 1990's. If on the other hand AB and AC are both based on 10 papers, then the estimated probabilities increase to .572 in the 1970's, .553 in the 1980's, and .534 in the 1990's. Thus there is strong support for hypothesis 1.

It is interesting to note that the variable $DIFFSTRENGTH$ is significantly negative. This means that a triple with one weak and one strong link is less likely to be transitive than a triple with two intermediate strength links, let alone with two strong links. To give a numerical example, if AB is based on 10 articles, but AC on only one article, then the estimated probability of a tie BC is .141 in the 1970's, .132 in the 1980's and .123 in the 1990's. Thus the hypothesis holds only for triads for which AB and AC are *both* strong. Note that this is in agreement with hypothesis 1.

3.2 Testing the strength of weak ties

We next look at testing hypothesis 2A. We perform a regression analysis on *link betweenness* of the links in the giant component. Link betweenness was introduced by Girvan and Newman (2002)[16]. The formal definition is as follows. Let n be the number of nodes in the network, g the set of ties in the network and let $\mathcal{L}_{ij}$ be the set of shortest paths between i and j. Then the link betweenness of a link AB is

$$B_{AB} = \frac{2}{n(n-1)} \sum_{ij} \frac{1}{|\mathcal{L}_{ij}|} \sum_{L \in \mathcal{L}_{ij}} I_{AB \in L},$$

where $I_{AB \in L}$ is an indicator variable, which is 1 if $AB \in L$ and 0 otherwise. In short, link betweenness of a link AB measures the fraction of all pairs of actors i and j for which the link AB lies on a shortest path between i and j. If there are multiple shortest paths between i and j, then each shortest path contributes equally to the link betweenness of the links on these paths.

[14] A Lagrange-Multiplier test on the alternative of separate logistic regressions for each decade is not rejected. The LM statistic is on average 4.37, while the .05-critical value is 9.49. Thus the use of a pooled regression with decade dummies is appropriate.

[15] In the 1000 connected triples typically more than 2700 different author names appear.

[16] The concept of link betweenness is very similar to the concept of betweenness centrality (Freeman, 1977). While betweenness centrality measures the centrality of actors, link betweenness measures the centrality of the actor's ties.

We can also rephrase the definition by saying that the link betweenness of a link AB measures the number of shortest paths in the network that would be broken if the link AB would be removed from that network. This formulation reveals that it is appropriate to use this measure in a test on hypothesis 2A.

Our test proceeds as follows. First, we extract the giant components from the networks of the 1970's, 80's and 90's, and for each tie in a giant component we compute link betweenness using the algorithm of Newman (2001). We then stack the observations on betweenness of links in the giant components of the three decades (72640 observations), and we randomly select a subsample of 1000 observations. For this subsample of ties we estimate the effect of the logarithm of strength on the logarithm of betweenness by means of an OLS regression in which we include three decade dummies.[17] We repeat this procedure 1000 times. In Table 4 we report average estimation results.

The results in Table 4 are very surprising. The strength of a tie has significant positive relation with the tie's link betweenness. Thus *strong ties* have a higher link betweenness, and these strong ties are crucial in connecting different actors in the network. In other words, hypothesis 2A is clearly rejected.

A disadvantage of using link betweenness is that it only reveals the number of shortest paths that would be broken if a tie would be removed from the network; it does not tell us how much the distance between a random pair of actors increases. This motivates us to turn to hypothesis 2B; the removal of an arbitrary weak tie from the network would increase average distance in the network more than the removal of an arbitrary strong tie. Before we turn to our test procedure, it should be mentioned that the distance between two unconnected actors is undefined (or infinite), and therefore the average distance can only be computed for network *components*. Recall that a component is a group of actors in which all actors are directly or indirectly connected to each other.

Unfortunately, this raises the problem that the size of the component might change when a link is removed from the network, which affects average distance as well. We therefore look at the following aspects of hypothesis 2B. We only focus on the largest (giant) component in the network, and we say that hypothesis 2B is supported if the removal of an arbitrary weak link would both decrease the size of the giant component as well as increase the average distance within the component more than the removal of a random strong link.

To test this hypothesis we perform the following simulation on the co-authorship network of the 1970's, 1980's and 1990's. Let a tie based on two or more papers be a strong tie, and let a tie based on one paper be a weak tie. We randomly delete 50 *strong* ties from the giant component of the coauthor network, and we recompute the size of the giant component and the average distance within the giant component. Next, starting from the

[17] An LM test on the alternative of separate regressions for each decade is not rejected with an average LM statistic of 4.46, while the .05-critical value is 9.49.

original network we delete 50 *weak* ties from the giant component, and we again compute the size of the giant component and the average distance within the giant component. We repeat this procedure m times, thus we create for each network two samples of m observations[18]. One sample measures the effect of deleting strong ties on the size of the giant component and the average distance, while the other sample measures the effect of deleting weak ties.

The test then boils down to comparing the mean (both for size of giant component and average distance) of the two samples in a one-sided heteroskedastic two-sample t-test. First, we test the null hypothesis that the sample mean for the size of the giant component in the two samples is identical against the alternative hypothesis that the sample mean in the first sample, in which we delete strong links, is larger than in the second. Next, we test the null hypothesis that the sample mean for average distance in the two samples is identical against the alternative hypothesis that the sample mean for average distance in the first sample is smaller.

Table 5 shows the results of these t-tests. We find that the removal of weak ties indeed decreases the size of the giant component more than the removal of strong ties. This is in support of Granovetter's strength of weak ties hypothesis. However, with respect to average distance we draw a different conclusion. Not *weak* links, but *strong* links have a bigger impact on average distances in the giant components.

The above simulation results suggest that hypothesis 2B does not hold; although the removal of a weak tie has a bigger impact on the size of the giant component than the removal of a strong tie, the impact on average distance is smaller for the removal of a weak tie than for the removal of a strong tie. However, since the two effects are conflicting the simulation results are a bit unsatisfying, and we therefore perform more simulations on the 1970's network as a robustness check. In these simulations we increase the number of links that are removed from the network. While in the above simulations we measured the effect of randomly and simultaneously removing 50 links, in the following simulations we remove 100 up to 500 links at once. Since there are 542 strong links in the giant component of the 1970's, in these simulations we remove a considerable fraction of all strong ties in the 1970's network.

As before we delete k random weak links ($k = \{100, 200, 500\}$) from the giant component in the 1970's and we measure the size of the giant component and the average distance within the giant component. We then delete k random strong links from the giant component, and we again measure the size of the giant component and the average distance. We repeat this procedure m times, such that we have two samples of m observations; one in which we measure the effect of deleting weak ties on the size of the giant component and

[18]We repeat the simulation 200 times for the 1970's and 1980's and network, and 100 times for the 1990's network. These simulations are very time consuming, and this restricted the number of simulations we could perform.

average distance, and one in which we measure the effect of deleting strong ties. We then perform one-sided heteroskedastic two-sample t-tests on the size of the giant component and average distance.

Table 6 shows the results from these simulations. We observe that the effect on the size of the giant component becomes insignificant when comparing the removal of 500 weak ties to the removal of 500 strong ties. On the other hand, the effect on average distance becomes more strongly significant when removing more ties at once. That is, the removal of 500 strong ties significantly increases the average distance more than the removal of 500 weak ties. Thus these supplementary simulations show that support for hypothesis 2B becomes weaker when considering the removal of 500 strong ties compared to the removal of 500 weak ties.

3.3 The core network of active economists

The results reported above show that, as predicted, connected triples are transitive, but this support for Hypothesis 1 does not imply that weak ties are better in connecting the world of economists. These results are puzzling, and lead us to think more carefully about the causes of the rejection of Hypothesis 2. One potential factor may be the fact that a large part of the actors in the economics network publish very few papers: around 60% of the actors have only published 1 paper in a decade, and around 15% only 2 papers. These are mostly PhD students or researchers at the beginning of their career, and non-academic researchers. Including these "one-time" economists in the data set may bias the results, as actors with only one publication and one or two co-authors by definition only have a weak link that do not connect other actors, except for connecting themselves. Those weak ties may lead to a rejection of Hypothesis 2.

To check whether the inclusion of these inactive economists matters, we repeat the analysis on Granovetter's strength of weak ties theory, but now we remove all actors from the network that have fewer than 5 papers written in a decade's time. This leaves us with a core network of active economists only.

Table 7 gives summary statistics for this network. We observe the same patterns as in the whole network, that is, a giant component, small average distances, small degree, and high clustering. Thus according to this table the network of active economists is not very different from the network of all economists.

We next consider Hypothesis 1. For the core network of active economists, we again gather all ordered triples ABC with a link AB and AC, and we perform a logistic regression on the existence of a link BC with the strength of the links AB and AC as regressors. To dilute the dependence of observations, we again perform the logistic regression on random subsamples of 500 observations. Table 8 shows the regression results of a logistic regression for transitivity of triples. We observe that the co-efficient on strength of the links AB and AC is significantly positive. To put some numbers on it, if AB and AC both have a link based on 1 paper, then the probability that B and C have a tie is about 22%. On the other hand, if the links AB and AC are each based on 5 papers, then the probability of a link BC is about 97%. Hence, as in the network of all (active and inactive) economists, Hypothesis 1 is supported for the core network of active economists.

We then look at Hypothesis 2A, that is, the hypothesis that a weak tie breaks more shortest paths than a strong tie. For that purpose, we again regress link betweenness on the (logarithm of) strength of a tie. The results reported in Table 9 show that when we consider the core network of active economists, the strength of a tie is negatively related to link betweenness. This is in agreement with Hypothesis 2A, but it contrasts the results in Subsection 3.2 in which it was shown that, when considering the whole network, Hypothesis 2A was rejected.

To summarize, focusing on the set of active economists only does not affect the results on Hypothesis 1, but it does with respect to Hypothesis 2. Whereas we reject the 'Strength of Weak Ties' theory when we consider all the actors in the co-author network of economists, for the restricted network of active economists (who have published 5 or more papers in a decade) we find support of Granovetter's theory.

4 Explaining the rejection of Granovetter's theory

Our findings show that Hypothesis 1 is univocally supported; connected triples with strong ties are more likely to be transitive. However, support for the second hypothesis depends on the particular network that one considers; in the network of active economists the hypothesis is supported, whereas it is rejected for the network that includes all economists.

We now would like to put these results in perspective. In particular, for a given network we would like to explain under which measurable conditions on the network structure Hypothesis 2A is likely to be rejected. We argue that there are two network properties that are crucial for the rejection of hypothesis 2A.

The first property is the existence of hubs, that is, actors with a high number of links. As shown by Albert, Jeong and Barabási (2000) these hubs play the role of *connectors* in the network, that is, without these hubs the network would be fragmented in isolated subgroups. Hence, hubs are nodes in the network that bridge different subgroups together. Given their crucial role in the network, the hubs have a high betweenness centrality, and the links between hubs have high link betweenness.

The second property is that the links between hubs are stronger than other links. Because of the central position of the hubs in the network, this implies that the links with high link betweenness are typically *strong*. These two properties put together suggest that strong links connect individuals who have more links on average and this means that they lie on more shortest paths. This fact more than compensates for the clustering in strong links noted above and results in strong links having a greater criticality.

The working of the above two properties are best explained by considering two example networks, an island network structure (Figure 4) and the core-periphery network structure (Figure 5). In the island network both hypothesis 1 and 2 are true, while in the core-periphery network hypothesis 1 is supported, while hypothesis 2 is rejected.

Consider first the network structure of Figure 4. In a stylized way this network represents a view of the world which often has been put forward when explaining the 'strength of weak ties' theory[19]. That is, the world consists of families or communities with very strong ties between family members. These families are connected through trade relations or occupational colleagueships. However, these interfamily ties are typically weaker than intrafamily ties.

From Figure 4 it is easily seen that with such a view of the world, hypothesis 1 and 2 hold. First, hypothesis 1 is obviously true as the only connected triples with two strong ties are within an island, while triples with one or two weak ties involve nodes from different islands. As everyone within an island is directly connected, it must be that all connected triples with two strong ties are transitive. Second, the link betweenness of weak ties is higher then the link betweenness of strong ties. Weak ties directly connect two separate islands; $4 \times 4 = 16$ shortest paths depend on a weak tie. The strong tie, on the other hand, is only crucial for the connectedness of the actors involved in a strong tie. In Figure 4, the strong tie between A and B lies on the shortest path between A and B, between A and the actors of island I, and between B and the actors of island J; a total of $5 + 5 = 10$ shortest paths. Other strong ties have the same or lower link betweenness. Hence, in the island network weak ties have higher link betweenness than strong ties.

These computations suggest why Granovetter's 'strength of weak ties' theory typically holds in an island network structure. On the other hand, the theory typically fails to hold for a core-periphery structure, as given in Figure 5. This network consists of a core of four actors, who are strongly connected in a clique. These actors have a number of ties with peripheral players. In Figure 5 each core actor is connected to five peripheral actors. The peripheral actors themselves have only a weak link to one of the core actors, and no link to other peripheral actors. In contrast to the island network structure, the core-periphery network has a distinct hierarchy.

When we examine the two hypotheses of Section 2 in the context of the core-periphery structure, we observe the following. First, triples with two strong ties necessarily involve core actors only. Since the core is completely internally connected these triples are transitive. On the other hand, triples with a weak tie involve peripheral players, and these triples are typically not transitive. Hence, the first hypothesis holds. Second, in contrast

[19] See, for example, Figure 2 in Granovetter (1973, p. 1365) or Figure 1 in Friedkin (1980, p. 412).

to hypothesis 2A, *strong* ties have higher betweenness. In a core-periphery network a strong tie belongs to the shortest path of the two core actors *and* the peripheral 'clients' attached to the core actors. On the other hand, a weak tie only connects the peripheral player involved. So, in Figure 5, the link between A and B belongs to the shortest paths of all nodes between I and J, a total of $6 \times 6 = 36$ paths. A weak tie only belongs to the shortest paths that connect a peripheral player to the rest of the network; in Figure 5 23 paths. Hence, strong ties have a higher betweenness than weak ties in a core-periphery network. Thus hypothesis 2A is rejected.

To understand the failure of the 'strength of weak ties' theory in the core-periphery structure, we first note that in both Figure 4 and Figure 5 weak ties form (local) bridges. However, the importance of these bridges is very different in the two network structures. In the island network structure the bridges connect different *communities* to each other, while in the core-periphery network structure the bridges only connect a single peripheral player. Hence, the bridges are more 'crucial' in the island structure than in the core-periphery network structure.

We now examine these two properties in the co-author network of economists, both including and excluding economists who have published less than 5 articles in a decade. First, we note that there is indeed a considerable inequality in the link degree distribution of the networks. Table 1 shows that the maximum link degree exceeds the average link degree by around 20 standard deviations. A similar pattern is observed in Table 7 where we concentrate on the core network of active economists. Furthermore, in Goyal et al. (2006) we showed that the removal of these high degree 'hubs' from the network is catastrophic for the cohesion of the co-author network, while the removal of random nodes has only a small, gradual affect on the network cohesion (see also Albert et al., 2000). For example, the removal of 5% of the nodes with the highest degree from the co-author network would result in a complete breakdown of the giant component. Hence, this work shows that high degree nodes are very important in connecting different parts of the network.

Next, we examine the correlation between strength and average degree. We first focus on the network that includes all economists. We perform the following regression on the links in the giant component of the co-author networks of the 1970's, 1980's and 1990's. The dependent variable is the logarithm of strength of links, and the regressors are: 1) the average log degree of the two actors attached to a link and; 2) the difference between the log degrees of the two actors; 3) three decade dummies. We again take 10000 random subsamples of size 1000 and perform regressions on these 10000 subsamples. The estimation results of these regressions are shown in Table 10.

The results show that there is a significant positive relation between strength of links and average degree of the two co-authors, and a negative relation between strength of links and difference of degree between co-authors. Hence, if two actors A and B both have many links, then the link between them is expected to be strong.

We explore the role of this correlation more closely next via an experiment in which we consider only the network of active economists: these are economists who have published 5 or more papers in a decade. In Table 11 we show the results of a regression of strength on average and difference degree. The results show that average and difference degree have *no* effect on the strength of a tie.

The results above show that we can trace back the difference in the test results of the 'strength of weak ties' theory to a difference in the relation between link degree and strength of a tie. When there is a positive relation, such as in the complete network of economists, and when there are hubs in the network, then we should expect strong ties to be more important for connecting the network. On the other hand, when there is no relation between link degree and strength, as in the core network of active economists, then we should expect the 'strength of weak ties' theory to hold.

We conclude by emphasizing that *the entire network is important and the natural one to focus on.* This is because it captures the demographics of network evolution, which is a key feature of many social networks. People enter the network with few links and modest productivity; as they mature they accumulate links and their productivity improves and eventually as they age, their links dissolve and diminish and at the same time their productivity declines as well. Our results on the entire network bring out an important structural implication of this life cycle aspect of social networks: that mature individuals will on average have more links and the links they have among themselves will be stronger. Since highly linked individuals are more central to the network, the strong links will be more central to the network than links which are weak.

5 Concluding remarks

This paper examines the celebrated 'strength of weak ties' hypothesis from a structural point of view: we ask if weak ties are more critical for integrating the network as compared to strong links. The first part of our paper shows this hypothesis is not valid in the co-author network of economists, although it is supported for the network of core economists. The second part of the paper argues that two features of the network together help account for this finding: one, significant inequality in number of co-authors across individuals, and two, a positive relationship between the strength of a tie and the number of co-authors of the involved authors.

The arguments in the paper taken together make the following general point: the classical view has been that society consists of different communities with strong ties within and weak ties across communities. This view emphasizes the role of weak ties in connecting a society. In contrast, our work highlight the strength of strong ties: we believe that this finding reflects a fundamental time dimension in the creation of links in a profession. Individuals create links and strengthen them over time. At any point in time, there will exist a demographic profile of young and old individuals. The young are likely to have relatively fewer ties and these ties are likely to have lower strength, as compared to the old, simply because they have had less time. The older individuals will on average be better connected and their links will be with other well connected older individuals, and our work suggests that they will be stronger. It then follows as a corollary that strong ties will be more important for connecting up the network than weak ties.

Our findings raise the question of whether similar patterns also obtain in non-academic social networks. One context where this would be specially interesting to study is labor market contacts. There is some evidence on labor market contacts in East Asia which shows that strong ties are typically more important for job seekers as compared to weak ties (Bian and Ang, 1997). It would be very interesting to relate the network structure properties to these informativeness properties of strong ties.

A second line of enquiry concerns the formation of such networks, in other words, what are the circumstances – economic, cultural and technological – under which these two properties are likely to emerge. The economic co-authors networks is a specific context. Many factors play a role in the decision to co-author or not, such as competition for priority, or complementarities of research skills, factors that do not play a role in other social settings. Moreover, the institution of academic research is such that there is a clear hierarchy in which professors advise PhD students. Many PhD students have only one or a few papers with their advising professor, in particular those who do not continue their career in academics. We would like to develop a simple theory of how such factors may help lead to the observed coauthor network.[20]

A third line of enquiry concerns the implications of the network patterns.[21] In this paper the focus has been on the structural aspects of the network structure, and it remains an open question whether strong ties are indeed more important for information dissemination. In the context of scientific collaboration, a combined analysis of the co-author network and the pattern of citations would be one way to move forward on this interesting question.

[20]For a brief survey of models of strong and weak links, see Goyal (2005).

[21]For a recent study of the relation between the strength of ties and social contagion, see Centola and Macy (2007).

Appendix

Table 1: Network statistics for coauthorship networks in economics, 1970's 1980's and 1990's.

period	1970's	1980's	1990's
number of nodes	33768	48441	83209
number of links	15020	29952	67018
number of papers	62227	93976	152868
average link degree	.890	1.237	1.611
std.dev. link degree	(1.359)	(1.779)	(2.209)
max link degree	24	36	51
average strength	1.265	1.336	1.354
std.dev. strength	(.877)	(.952)	(1.043)
max strength	26	25	39
size of giant component	5164	13358	31074
as percentage	.153	.276	.373
second largest component	120	29	40
isolated	16846	19503	25881
as percentage	.499	.403	.311
clustering coefficient	.193	.181	.169
degree correlation	.124	.151	.137
average distance	12.41	10.82	9.95
std.dev. distance	(3.81)	(2.98)	(2.46)
maximum distance (diameter)	35	37	30

Link degree of a node: number of links attached to the node. Strength of a link: number of papers coauthored by the two authors of a link. Size of giant component: size of the largest component, a subset of nodes for which there is a path between each pair of node in the subset. Second largest component: size of the largest component except the giant component. Isolated: number of nodes without any links. Clustering coefficient: fraction of connected triples that are transitive. Degree correlation: correlation coefficient between the link degrees of two neighbouring nodes. Distance of a pair of nodes: shortest path length between the pair of nodes.

Table 2: Fraction of subsets of connected triples that are transitive in the coauthorship network for economists, 1970's, 1980's and 1990's.

period	1970's	1980's	1990's
observations	29515	83752	243931
Strong tie: ≥ 2 papers			
weak-weak	.209	.193	.177
weak-strong	.140	.138	.135
strong-strong	.300	.279	.257
χ^2-test	340.6	920.6	2046.8
p-value	.000	.000	.000
Strong tie: ≥ 3 papers			
weak-weak	.196	.176	.164
weak-strong	.160	.184	.171
strong-strong	.396	.355	.337
χ^2-test	152.8	322.6	974.5
p-value	.000	.000	.000
Strong tie: ≥ 5 papers			
weak-weak	.192	.176	.163
weak-strong	.211	.244	.232
strong-strong	.444	.425	.487
χ^2-test	21.3	224.6	890.4
p-value	.000	.000	.000
all	.193	.181	.169

Observations are connected triples. The set of connected triples is partitioned into three subsets. Weak-weak: connected triples consisting of two weak ties. Weak-strong: connected triples consisting of one weak and one strong tie. Strong-strong: connected triples consisting of two strong ties. All: all connected triples. χ^2-test: test statistic for χ^2-independence test (2 degrees of freedom).

Table 3: Estimation results of a logistic regression on the transitivity of connected triples.

variable	coefficient	std. error	t-stat
DUMMY70S	-1.455	.291	-4.965
DUMMY80S	-1.535	.185	-8.294
DUMMY90S	-1.612	.124	-13.031
AVGSTRENGTH	.759	.304	2.561
DIFFSTRENGTH	-.531	.225	-2.356
loglikelihood	-455.03		

Average results of regressions on 1000 random subsamples with each subsample 1000 observations. Each observation is a triple ABC for which A and B are tied and A and C are tied. The dependent variable, $TRANSITIVE$, is 1 if BC is also tied, and 0 otherwise. $AVGSTRENGTH = (X_{AB} + X_{AC})/2$ where X_{AB} is the natural logarithm of the number of papers A and B have written together. $DIFFSTRENGTH = |X_{AB} - X_{AC}|$. $DUMMY70S$, $DUMMY80S$, and $DUMMY90S$ are dummy variables to indicate whether the observations were drawn from the 1970's, 1980's or 1990's.

Table 4: Estimation results of a regression on link betweenness

variable	coefficient	.95-confidence	p-value
DUMMY70S	-7.871	.276	-29.044
DUMMY80S	-8.959	.175	-51.677
DUMMY90S	-9.895	.120	-83.015
LNSTRENGTH	.477	.181	2.764
R^2	.060		

Average results of regressions on 1000 random subsamples with each subsample 1000 observations. Each observation is a link AB in the giant component of either the 1970's, 1980's or 1990's network. The dependent variable is the natural logarithm of link betweenness of link AB. $LNSTRENGTH$ is the natural logarithm of the number of papers A and B have written together. $DUMMY70S$, $DUMMY80S$, and $DUMMY90S$ are dummy variables to indicate whether the observations were drawn from the 1970's, 1980's or 1990's. Standard errors are heteroskedasticity-consistent.

Table 5: Simulation results for the coauthorship network for economists, 1970's, 1980's and 1990's.

period	1970's	1980's	1990's
replications	200	200	100
Size of giant component			
actual	5164	13358	31074
after deleting 50 strong ties			
mean	5113.4	13333.2	31056.7
std.dev.	22.8	11.3	9.6
after deleting 50 weak ties			
mean	5106.4	13325.4	31049.9
std.dev.	22.8	16.3	9.9
one-sided t-test			
t-stat.	3.08	5.57	4.97
p-value	.001	.000	.000
Average distance			
actual	12.407	10.818	9.949
after deleting 50 strong ties			
mean	12.485	10.838	9.955
std.dev.	.062	.009	.003
after deleting 50 weak ties			
mean	12.444	10.828	9.952
std.dev.	.056	.016	.003
one-sided t-test			
t-stat.	6.88	7.74	6.61
p-value	1.000	1.000	1.000

Simulation compares the effect of randomly deleting 50 strong ties to randomly deleting 50 weak ties from the giant component. A strong tie is a tie based on 2 or more papers, and a weak tie is a tie based on 1 paper. One-sided t-test is a one-sided heteroskedastic two-sample t-test; the null hypothesis is equal mean for the sample after deleting weak ties and the sample after deleting strong ties. The alternative hypothesis for the size of the giant component is that the mean size of the giant component after randomly deleting 50 strong ties is larger than the mean size after randomly deleting 50 weak ties. The alternative hypothesis for average distance is that the mean average distance after deleting 50 strong ties is smaller than the mean average distance after deleting 50 weak ties.

Table 6: Simulation results for the coauthorship network for economists in the 1970's.

removed links k	100	200	500
replications	200	200	200
Size of giant component			
actual	5164	5164	5164
after deleting strong ties			
mean	5058.9	4950.5	4562.2
std.dev.	31.3	42.3	66.8
after deleting weak ties			
mean	5048.9	4927.5	4554.4
std.dev.	32.2	43.5	61.5
one-sided t-test			
t-stat.	3.14	5.36	1.20
p-value	.001	.000	.115
Average distance			
actual	12.407	12.407	12.407
after deleting 50 strong ties			
mean	12.565	12.733	13.254
std.dev.	.086	.119	.215
after deleting 50 weak ties			
mean	12.484	12.565	12.816
std.dev.	.080	.117	.174
one-sided t-test			
t-stat.	9370	14.30	22.38
p-value	1.000	1.000	1.000

Simulation compares the effect of randomly deleting k strong ties to randomly deleting k weak ties from the giant component of the 1970's. A strong tie is a tie based on 2 or more papers, and a weak tie is a tie based on 1 paper. One-sided t-test is a one-sided heteroskedastic two-sample t-test; the null hypothesis is equal mean for the sample after deleting weak ties and the sample after deleting strong ties. The alternative hypothesis for the size of the giant component is that the mean size of the giant component after randomly deleting k strong ties is larger than the mean size after randomly deleting k weak ties. The alternative hypothesis for average distance is that the mean average distance within the giant component after deleting k strong ties is smaller than the mean average distance after deleting k weak ties.

Table 7: Network statistics for coauthorship networks of active economists only, 1970's 1980's and 1990's.

period	1970's	1980's	1990's
number of nodes	3890	6943	12326
number of links	2524	6711	16455
average link degree	1.298	1.933	2.670
std.dev. link degree	(1.612)	(2.091)	(2.601)
max link degree	15	22	41
average strength	1.929	1.967	1.982
std.dev. strength	(1.788)	(1.690)	(1.825)
max strength	26	25	39
size of giant component	1476	4001	9013
as percentage	.379	.576	.731
second largest component	23	9	13
isolated	1504	1893	2206
as percentage	.387	.273	.179
clustering coefficient	.142	.144	.137
degree correlation	.179	.197	.142
average distance	10.28	9.12	8.45
std.dev. distance	(3.46)	(2.49)	(2.18)
maximum distance (diameter)	29	26	27

Network only contains nodes of economists that have written 5 papers or more in one decade. See the notes of Table 1 for a description of the variables.

Table 8: Estimation results of a logistic regression for transitivity of connected triples in the network of active economists only.

variable	coefficient	std. error	t-stat
DUMMY70S	-2.413	.789	-3.837
DUMMY80S	-2.309	.335	-6.860
DUMMY90S	-2.336	.238	-9.815
AVGSTRENGTH	1.118	.317	3.572
DIFFSTRENGTH	-.176	.247	-.699
loglikelihood	-191.05		

Network only contains nodes of economists that have written 5 papers or more in one decade. Average results of regressions on 1000 random subsamples with each subsample 500 observations. See the notes of Table 3 for a further description.

Table 9: Estimation results of a regression on link betweenness in network of active economists only.

variable	coefficient	std.error	z-stat
DUMMY70S	-5.916	.277	-22.678
DUMMY80S	-6.959	.142	-50.448
DUMMY90S	-7.893	.084	-95.622
LNSTRENGTH	-.379	.139	-2.670
R^2	.173		

Network only contains nodes of economists that have written 5 papers or more in one decade. Average results of regressions on 1000 random subsamples with each subsample 500 observations. See the notes of Table 4 for a further description.

Table 10: Estimation results of a regression on the strength of ties

variable	coefficient	std.error	z-stat
DUMMY70S	.0286	.0563	.453
DUMMY80S	.0246	.0461	.526
DUMMY90S	.0040	.0436	.083
AVGDEGREE	.1827	.0293	6.242
DIFFDEGREE	-.0421	.0188	-2.253
R^2	.057		

Average results of regressions on 1000 random subsamples with each subsample 1000 observations. Each observation is a link AB in the giant component of either the 1970's, 1980's or 1990's network. The dependent variable is the natural logarithm of the number of papers A and B have written together. $AVGDEGREE = (X_A + X_B)/2$ where X_A is the natural logarithm of the number of links A has. $DIFFDEGREE = |X_A - X_B|$. $DUMMY70S$, $DUMMY80S$, and $DUMMY90S$ are dummy variables to indicate whether the observations were drawn from the 1970's, 1980's or 1990's. Standard errors are heteroskedasticity-consistent.

Table 11: Estimation results of a regression on the strength of ties in the networks of active economists only.

variable	coefficient	std.error	z-stat
DUMMY70S	.4100	.1217	3.357
DUMMY80S	.4469	.1001	4.481
DUMMY90S	.4477	.1001	4.484
AVGDEGREE	-.0099	.0587	-.183
DIFFDEGREE	.0135	.0500	-.250
R^2	.008		

Network only contains nodes of economists that have written 5 papers or more in one decade. Average results of regressions on 1000 random subsamples with each subsample 1000 observations. See the notes of Table 10 for a further description.

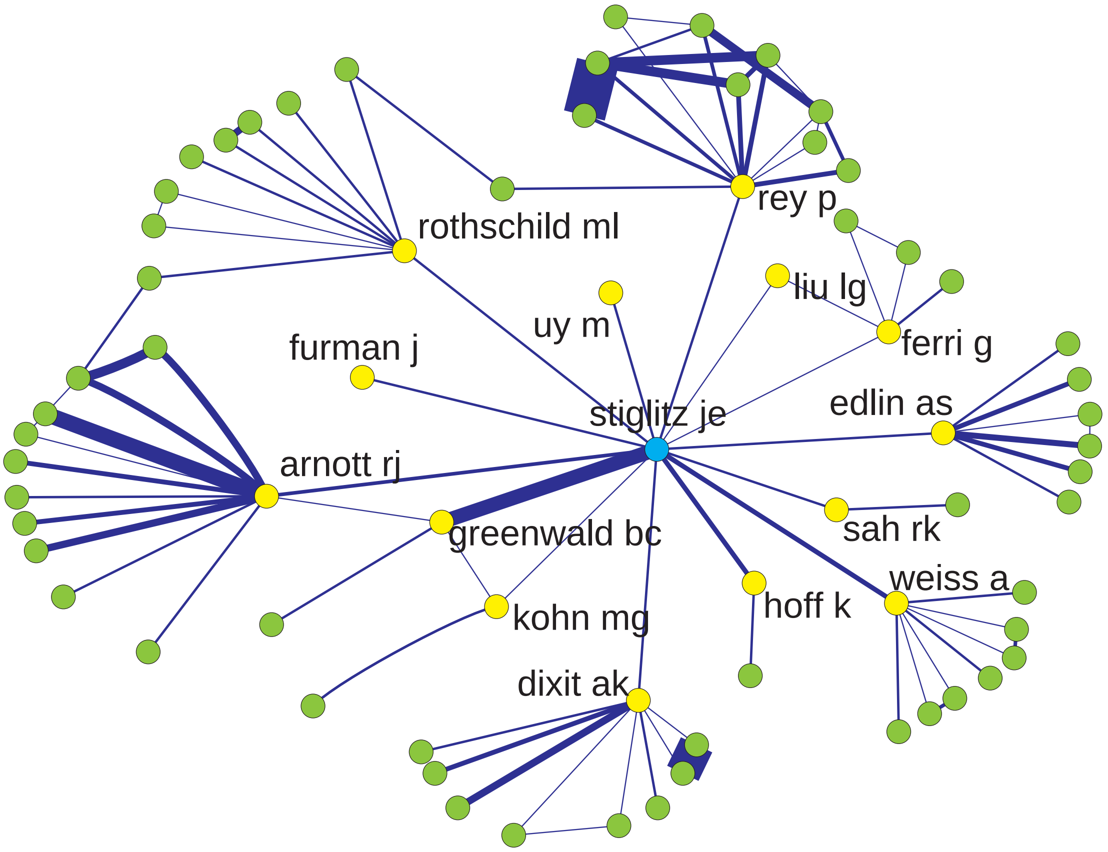


Figure 1: Local network of collaboration of Joseph E. Stiglitz in the 1990s.

Note: The figure shows all authors within distance 2 of J.E. Stiglitz as well as the links between them. The width denotes the strength of a tie. Some economists might appear twice or are missing due to the use of different initials or misspellings in EconLit. The figure was created by software program *Pajek*.

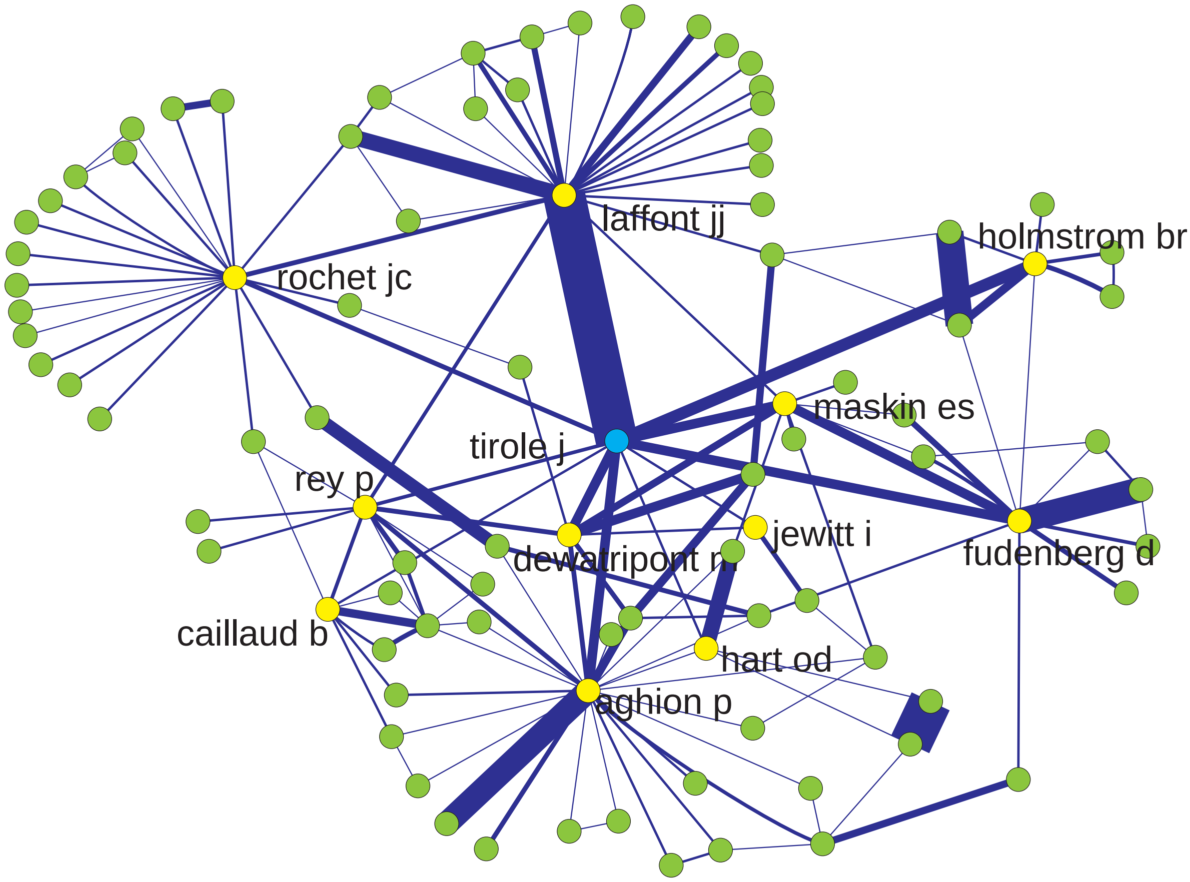
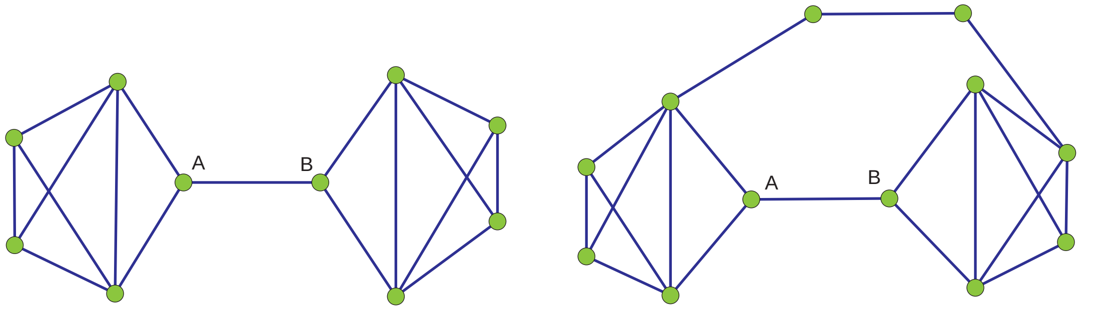


Figure 2: Local network of collaboration of Jean Tirole in the 1990s.

Note: The figure shows all authors within distance 2 of J. Tirole as well as the links between them. The width denotes the strength of a tie. Some economists might appear twice or are missing due to the use of different initials or misspellings in EconLit. The figure was created by software program *Pajek*.



(a) The tie between A and B is a bridge.

(b) The tie between A and B is a local bridge of degree 6.

Figure 3: Two networks with a bridge and a local bridge.

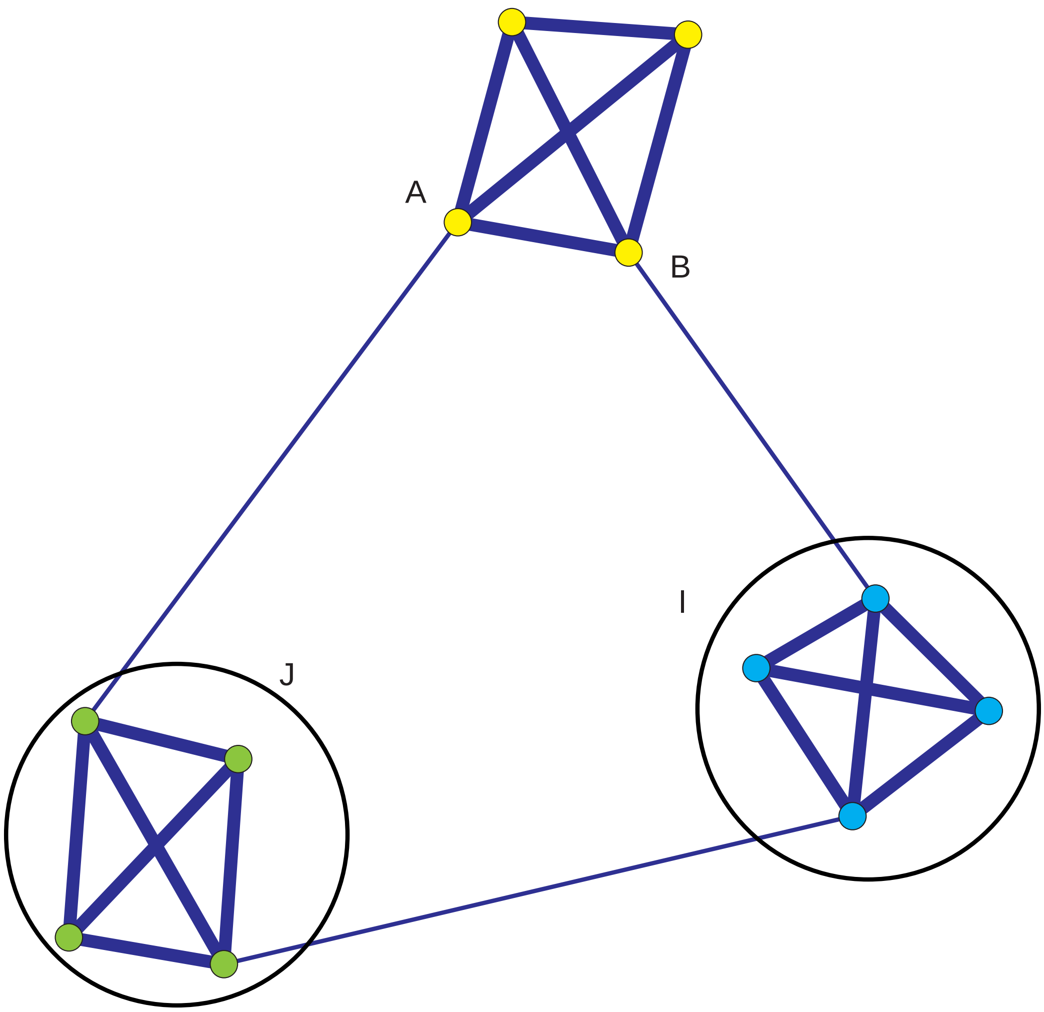


Figure 4: An island network.

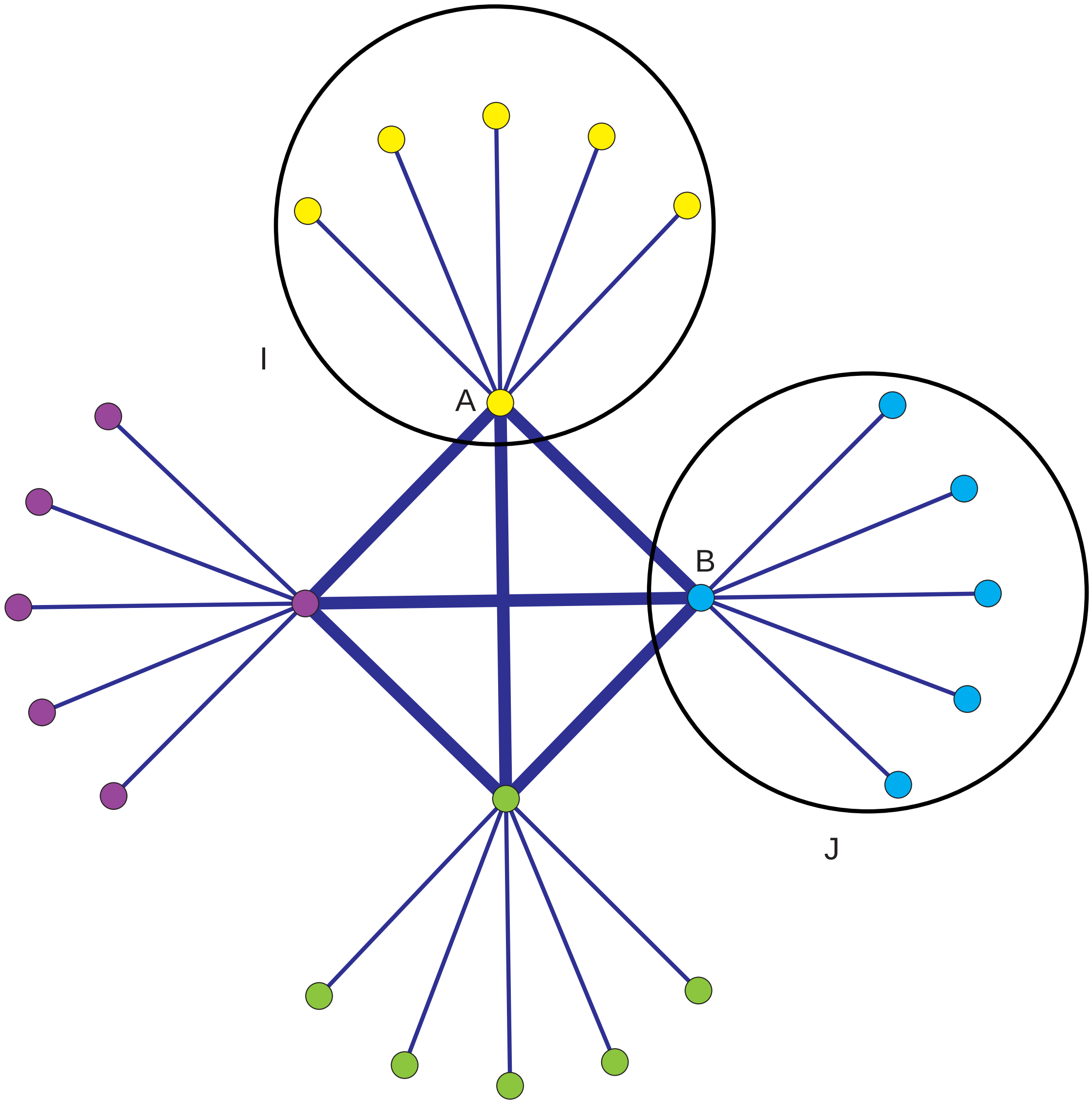


Figure 5: A core-periphery network.

References

[1] Albert, Réka and Albert-László Barabási (2002), 'Statistical mechanics of complex networks', *Review of Modern Physics*, 74, 47–97.

[2] Albert, Réka, Hawoong Jeong, and Albert-László Barabási (2000), 'Error and attack tolerance of complex networks', *Nature*, 406, 378–382.

[3] Aveneri, Shlomo and Avner de Shalit (1992), *Communitarianism and Individualism.* Oxford University Press. Oxford.

[4] Bala, V. and S. Goyal (1998), Learning from Neighbours, *Review of Economic Studies* 65, 595-621.

[5] Bala, V. and S. Goyal (2000), A Non-Cooperative Model of Network Formation, *Econometrica* , 68, 1181-1230.

[6] Ballester, C., A. Calvo-Armengol, and Y. Zenou (2006), Who's who in networks. Wanted: The Key Player, *Econometrica*, 74, 5, 1403-1417.

[7] Bandiera, O., I. Barankay, and I. Rasul (2006), Social connections and incentives in the work place: evidence from personnel data, *Review of Economic Studies*, to appear.

[8] Barabási Albert-László and Réka Albert (1999), 'Emergence of scaling in random network', *Science*, 99, 509–512.

[9] Barrat, Alain, Marc Barthélemy, and Alessandro Vespignani (2004a), 'Modeling the evolution of weighted networks', *Physical Review E*, 70, 066149.

[10] Barrat, Alain, Marc Barthélemy, Romualdo Pastor-Satorras, and Alessandro Vespignani (2004b), 'The architecture of complex weighted networks', *Proceedings of the National Academy of Sciences of the U.S.A.*, 101(11), 3747–3752.

[11] Bian, Yanjie and Soon Ang (1997), 'Guanxi networks and job mobility in China and Singapore', *Social Forces*, 75(3), 981–1005.

[12] Borgatti, Stephen P. and Scott L. Feld (1994), 'How to test the strength of weak ties theory', *Connections*. 17, 45–46.

[13] Burt, Ronald S. (1994), *Structural Holes: The social structure of competition.* Harvard University Press, Cambridge, MA.

[14] Calvó-Armengol, A. and M. O. Jackson (2004) "The Effects of Social Networks on Employment and Inequality, *American Economic Review*, 94, 3, 426-454.

[15] Centola, D. and M. Macy (2007), Contagions and the Weakness of Long Ties, *American Journal of Sociology*, 113, 3, 70234.

[16] Coleman, James S. (1966), *Medical Innovation: A Diffusion Study*, Second Edition, Bobbs-Merrill, New York.

[17] Dorogovtsev, Sergey and José Fernando Mendes (2002), 'Evolution of networks', *Advances in Physics*, 51, 1079-1187.

[18] Frank, Ove and David Strauss (1986), 'Markov Graphs', *Journal of the American Statistical Association*, 81(395), 832-842.

[19] Freeman, Linton C. (1977), 'A set of measures of centrality based on betweenness', *Sociometry*, 40(1), 35–41.

[20] Friedkin, Noah (1980), 'A test of the structural features of Granovetter's "strength of weak ties" theory'., *Social Network*, 2, 411–422.

[21] Girvan, Michelle and Mark E.J. Newman (2002), 'Community structure in social and biological networks', *Proceedings of the National Academy of Sciences of the U.S.A.*, 99, 8271–8276.

[22] Golub, B. and Jackson, M. (2008), Naive learning and wisdom of crowds, *American Economic Journal: microeconomics*, to appear.

[23] Goyal, Sanjeev (2005), 'Strong and weak links', *Journal of European Economic Association*, 3(2-3), 608-616.

[24] Goyal, Sanjeev (2007), *Connections: an introduction to the economics of networks.* Princeton University Press. New Jersey.

[25] Goyal, Sanjeev, Marco J. van der Leij, and José Luis Moraga-González (2006), 'Economics: an emerging small world', *Journal of Political Economy*, 114(2), 403–412.

[26] Goyal, S. and J. L. Moraga-Gonzalez (2001), R&D Networks, *Rand Journal of Economics,* 32, 4, 686-707.

[27] Granovetter, Mark (1973), 'The strength of weak ties', *American Journal of Sociology*, 78(6), 1360–1380.

[28] Granovetter, Mark (1983), 'The strength of weak ties: A network theory revisited', *Sociological Theory*, 1, 201–233.

[29] Granovetter, Mark (1995), *Getting a Job: A Study of Contacts and Careers*, 2nd edition, Harvard University Press, Cambridge MA.

[30] Guimera, R. B. Uzzi, J. Spiro, L. Amaral (2005), Team assembly mechanisms determine collaboration network structure and team performance, *Science*, 308, 697-702.

[31] Jackson, Matthew (2008), *Social and economic networks.* Princeton University Press. New Jersey.

[32] Hägerstrand, Torsten (1969), *Innovation diffusion as a spatial process.* University of Chicago Press. Chicago.

[33] Kumpula, Jussi M., Jukka-Pekka Onnela, Jari Saramäki, Kimmo Kaski, and János Kertész (2007), 'Emergence of communities in weighted networks', *Physical Review Letters*, 99, 228701.

[34] Milgram, Stanley (1967), 'The small world problem', *Psychology Today*, 2, 60-67.

[35] Moody, James (2004), 'The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999', *American Sociological Review*, 69, 213–238.

[36] Newman, Mark E.J. (2001), 'The structure of scientific collaboration networks', *Proceedings of the National Academy of Sciences*, 98, 404-409.

[37] Newman, Mark E.J. (2003), 'The structure and function of complex networks', *SIAM Review*, 45, 167–256.

[38] Newman, Mark E.J. (2004), 'Analysis of weighted networks', *Physical Review E*, 70, 056131.

[39] Onnela, J.P., J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási (2007), 'Structure and tie strengths in mobile communication networks', *Proceedings of the National Academy of Sciences*, 104, 7332-7336.

[40] Rogers, Everett M. (1995), *Diffusion of innovations.* Fourth Edition. Free Press, New York.

[41] Taylor, Charles (1999), *Sources of the self: the making of modern identity.* Harvard University Press, Cambridge, Ma.

[42] van der Leij, Marco J. (2006), *The Economics of Networks: Theory and Empirics*, Tinbergen Institute Research Series No. 384. Amsterdam: Thela Thesis

[43] Watts, Duncan (1999), *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton University Press, Princeton, New Jersey.

[44] Watts, Duncan J. and Steven H. Strogatz (1998), 'Collective dynamics of "small world" networks', *Nature*, 393, 440–442.

[45] Yook, Soon-Hyung, Hawoong Jeong, Albert-László Barabási, and Y. Tu (2001), 'Weighted evolving networks', *Physical Review Letters*, 86(25), 5835–5838.

[46] Yakubovich, V (2005), 'Weak ties, information and influence: how workers find jobs in a local Russian market', *American Sociological Review*, 70, 408-421.

[47] Zachary, Wayne (1977), 'An information flow model for conflict and fission in small groups', *Journal of Anthropological Research*, 33, 452–473.

PUBLISHED ISSUES *

WP-AD 2009-01	"Does sex education influence sexual and reproductive behaviour of women? Evidence from Mexico" P. Ortiz. February 2009.
WP-AD 2009-02	"Expectations and frward risk premium in the Spanish power market" M.D. Furió, V. Meneu. February 2009.
WP-AD 2009-03	"Solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm" L. Maliar, S. Maliar, F. Valli. February 2009.
WP-AD 2009-04	"Employee types and endogenous organizational design: an experiment" A. Cunyat, R. Sloof. February 2009.
WP-AD 2009-05	"Quality of life lost due to road crashes" P. Cubí. February 2009.
WP-AD 2009-06	"The role of search frictions for output and inflation dynamics: a Bayesian assessment" M. Menner. March 2009.
WP-AD 2009-07	"Factors affecting the schooling performance of secondary school pupils – the cost of high unemployment and imperfect financial markets" L. Farré, C. Trentini. March 2009.
WP-AD 2009-08	"Sexual orientation and household decision making. Same-sex couples' balance of power and labor supply choices" S. Oreffice. March 2009.
WP-AD 2009-09	"Advertising and business cycle fluctuations" B. Molinari, F. Turino. March 2009.
WP-AD 2009-10	"Education and selective vouchers" A. Piolatto. March 2009.
WP-AD 2009-11	"Does increasing parents' schooling raise the schooling of the next generation? Evidence based on conditional second moments" L. Farré, R. Klein, F. Vella. March 2009.
WP-AD 2009-12	"Equality of opportunity and optimal effort decision under uncertainty" A. Calo-Blanco. April 2009.
WP-AD 2009-13	"Policy announcements and welfare" V. Lepetyuk, C.A. Stoltenberg. May 2009.
WP-AD 2009-14	"Plurality versus proportional electoral rule: study of voters' representativeness" A. Piolatto. May 2009.
WP-AD 2009-15	"Matching and network effects" M. Fafchamps, S. Goyal, M.J. van der Leij. May 2009.

* Please contact Ivie's Publications Department to obtain a list of publications previous to 2009.

WP-AD 2009-16 “Generalizing the S-Gini family –some properties-”
F.J. Goerlich, M.C. Lasso de la Vega, A.M. Urrutia. May 2009.

WP-AD 2009-17 “Non-price competition, real rigidities and inflation dynamics”
F. Turino. June 2009.

WP-AD 2009-18 “Should we transfer resources from college to basic education?”
M. Hidalgo-Hidalgo, I. Iturbe-Ormaetxe. July 2009.

WP-AD 2009-19 “Immigration, family responsibilities and the labor supply of skilled native women”
L. Farré, L. González, F. Ortega. July 2009.

WP-AD 2009-20 “Collusion, competition and piracy”
F. Martínez-Sánchez. July 2009.

WP-AD 2009-21 “Information and discrimination in the rental housing market: evidence from a field experiment”
M. Bosch, M.A. Carnero, L. Farré. July 2009.

WP-AD 2009-22 “Pricing executive stock options under employment shocks”
J. Carmona, A. León, A. Vaello-Sebastiá. September 2009.

WP-AD 2009-23 “Who moves up the career ladder? A model of gender differences in job promotions”
L. Escriche, E. Pons. September 2009.

WP-AD 2009-24 “Strategic truth and deception”
P. Jindapon, C. Oyarzun. September 2009.

WP-AD 2009-25 “Do social networks prevent bank runs?
H.J. Kiss, I. Rodríguez-Lara, A. Rosa-García. October 2009.

WP-AD 2009-26 “Mergers of retailers with limited selling capacity”
R. Faulí-Oller. December 2009.

WP-AD 2010-01 “Scaling methods for categorical self-assessed health measures”
P. Cubí-Mollá. January 2010.

WP-AD 2010-02 “Strong ties in a small world”
M.J. van der Leij, S. Goyal. January 2010.

WP-AD 2010-03 “Timing of protectionism”
A. Gómez-Galvarriato, C.L. Guerrero-Luchtenberg. January 2010.

serie ad

Ivie

Guardia Civil, 22 - Esc. 2, 1º
46020 Valencia - Spain
Phone: +34 963 190 050
Fax: +34 963 190 055

Department of Economics
University of Alicante

Campus San Vicente del Raspeig
03071 Alicante - Spain
Phone: +34 965 903 563
Fax: +34 965 903 898

Website: http://www.ivie.es
E-mail: publicaciones@ivie.es