TI 2006-008/1
Tinbergen Institute Discussion Paper

Strong Ties in a Small World

Marco van der Leij[1]
Sanjeev Goyal[1,2]

[1] *Erasmus University Rotterdam, and Tinbergen Institute;*
[2] *University of Essex.*

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

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

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.
Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

Strong ties in a small world

Marco J. van der Leij* Sanjeev Goyal†

January 2006

Abstract

In this paper we test the celebrated 'Strength of weak ties' theory of Granovetter (1973). We test two hypotheses on the network structure in a data set of collaborating economists. While we find support for the hypothesis of transitivity of strong ties, we reject the hypothesis that weak ties reduce distance more than strong ties do. We relate this surprising result to two different views of society. Whereas the classical view has been that society consists of different communities with strong ties within communities and weak ties between, the community of economic researchers has an interlinked star structure with strong ties between the stars. In such a world, strong ties are more important than weak ties.

*Tinbergen Institute, Erasmus University Rotterdam. E-mail: mvanderleij@few.eur.nl.

†University of Essex and Tinbergen Institute. E-mail: sgoyal@essex.ac.uk. We thank Vincent Buskens, Antoni Calvo-Armengol, Dennis Fok, Matthew Jackson, Willemien Kets, Marcel Fafchamps, Michael Kosfeld, Richard Paap, Werner Raub and attendants at seminars at NAKE Research Day 2004, Erasmus University, Utrecht University, and workshop in Mannheim and Trieste 2005 for helpful discussions and comments.

1 Introduction

This paper examines the celebrated strength of weak ties hypothesis. In his seminal paper (Granovetter 1973), Granovetter 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 densily 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.

Granovetter 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). However, this result does not seem to be robust. For example, in many East Asian countries it is found that jobseekers depend heavily on *strong* ties in their job search (Bian and Ang, 1997).

In later surveys (Granovetter 1983, 1995) Granovetter provides more empirical evidence. However, a closer look reveals that this empirical work typically focus on the relation between strength of a tie and some social outcome of the actor involved, for example, employment, income, status. They usually do not test Granovetter's arguments relating to properties of the network structure. Thus, empirical research that directly tests the structural content of the strength of weak ties hypotheses using a large data set appears to be lacking. Exceptions are Friedkin (1980) and Borgatti and Feld (1994), but the size of their data set is quite small. The reason that previous researchers have focused on small network samples or indirect tests is probably twofold. 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 coauthorship relations of economists publishing in scientific journals, to explore the validity of Granovetter's hypothesis. This 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 specify the network 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 main findings are as follows: We find 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 leads us to ask: what is it about the arrangement of strong and weak links in the network that leads to them. The second part of our paper is then dedicated to exploring the structural features of Granovetter's theory more closely. We suggest that the key to understanding the relatively greater criticality of strong ties is the fact that 1) the network is connected by a set of interlinked stars, 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. We believe that these findings support an alternative view of the structure of the social world: There is significant inequality in the distribution of connections across individuals, with some authors having a very large number of connections relative to the average. Finally, the connections between highly connected individuals are on average stronger. These observations together suggest a very different social structure.

We place our findings in context now. To our knowledge only Friedkin (1980) has tested the structural network features of Granovetter's theory thoroughly. Friedkin's 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. It is therefore important to note thatwe have a more extensive analysis on this hypothesis, both based on the new concept of link betweenness and on a much more extensive simulation. Further, our data is much larger and covers scientists from many universities instead of one.

Borgatti and Feld (1994) propose another test for the strength of weak ties theory based on the overlap and non-overlap of the neighbourhoods of dyad members[1]. They apply their procedure to Zachary's (1977) Karate club data. For this particular data set they surprisingly reject the hypothesis that the dyad members of weak ties have larger non-overlapping neighbourhoods. Their results are very similar to ours.

Our paper is also related to research in physics on large networks. Physicists have tried to unravel the complex structure of large networks and the dynamics that govern them (Watts and Strogatz 1998, Barabási and Albert 1999, Albert and Barabási 2002, Dorogovtsev and Mendes 2002, Newman 2003). The two views of the world can be related to two main competing models in the complex networks literature. The view of highly clustered local neighbourhoods with few random weak link between them is echoed in the 'small world' model of Watts and Strogatz (1998). On the other hand, the view of a very unequal network in which stars dominate is related to the scale-free network of Barabasi and Albert (1999). In this literature little attention has been paid on the strength of ties, and complex dynamic models of weighted network are just recently emerging (Yook et al, 2001; Barrat et al, 2004a, 2004b; Newman, 2004).

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. In Section 5 we show that these properties indeed account for the rejection of the 'strength of weak ties' theory. Section 6 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

[1]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.

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. If that is the case, then 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.

For the development of his theory Granovetter takes this argument to the extreme, in the following way; consider three individuals A, B and C in a social network. Suppose A-B and A-C have a strong tie, while the B-C tie does not exist. About such a triad structure, Granovetter says the following: "I will exaggerate in what follows by saying that this triad never occurs – that is, that the B-C tie is always present"(Granovetter 1973, pp. 1363). This leads to our first hypothesis.

Hypothesis 1A: The forbidden triad: *in a social network a situation where A has strong ties with B and C and B and C do not have a tie between them does not occur.*

A more plausible form of this hypothesis would simply require that the likelihood of triad completion should be much higher in case A-B and A-C are strong ties, as compared to when they are both weak ties. 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, pp. 1363). This leads to the following version of hypothesis 1.

Hypothesis 1B: *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 1A to the presence of shortest paths and bridges. These concepts are defined as follows; a *path* between i and j is a set of 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. The shortest path is the path with the shortest path length, and the distance is shortest path length. A *bridge* or *critical link* 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 1A is true and if every person has several strong ties, then *no strong tie is a (local) bridge*, or equivalently, all (local) bridges are weak ties. Since strong links are transitive, none of these links are likely to be critical.

This does not apply to weak links, since many of the triads with weak links are likely not to be completed.

A weaker version of 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, pp. 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 an arbitrary 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 an arbitrary 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. However, in Granovetter (1973) he continues his theory to argue that Hypothesis 2 has implications on macro-phenomena as diffusion, social mobility, and political organization (Granovetter 1973). These aspects have been the focus of extensive research as discussed in the introduction section.

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 journals, and it is split up in three decades, 1970-1979, 1980-1989 and 1990-1999. Each node is an economist[2]. 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.

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 great 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.

[2]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 retrieved from the corresponding author.

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. It is also natural to assume that a tie between A and B is stronger when A and B collaborate repeatedly as compared to when they collaborate only once. In fact, a single joint project might be a signal of a difficult collaboration that both authors prefer to be 'once but never again'.

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[3]. 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, 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 remarkably small, and the fraction of connected triples that are transitive, known as the clustering coefficient, is much higher than one would expect in a random matching network. 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. Thus, according to Goyal, van der Leij and Moraga-González (2005), a *small world* (Watts 1999) is emerging in economics.

3.1 Testing hypothesis 1

We now turn to hypotheses 1A and 1B. 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.

From the table we directly observe that the strong hypothesis 1A is rejected as in all cases the fraction of transitive triples is below one half. Moreover, the numbers in Table 2 clearly

[3]It is worth noting that reported numbers slightly differ from the numbers reported in Goyal et al.(2005), since the rule that distinguishes different author names is different in Goyal et al.(2005).

suggest that the weak hypothesis 1B 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 can not be directly drawn from them. The problem is that these tests require independence of observations, while observations 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.

In order to address the problem of 'overlapping triples' we perform a more rigorous analysis in the form of a logistic regression. 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.

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 with 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 the logistic regression as in (1) in which we also include dummies for the three decades[4]. With this subsample size the power of the test is still quite high, while the overlap of triples is starkly reduced[5]. 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 1B.

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 1B.

3.2 Testing hypothesis 2

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)[6]. 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

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

[4]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.

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

[6]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.

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.

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[7]. We repeat this procedure a 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*. Remember that a component is a cluster 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 an arbitrary strong link.

[7] 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.

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 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[8]. 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 find an unexpected result. 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

[8]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.

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 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.

The above analysis shows that hypotheses 2A and 2B are not supported. Before we saw however that hypothesis 1B was supported. This seems a contradiction, and we therefore want to explain this apparent contradictory finding.

4 Explaining the rejection of Granovetter's theory

Under which measurable conditions on the network structure is Hypothesis 2A likely to be rejected? In this section we argue that there are two network properties that are crucial to understand the failure of hypothesis 2A. We show that the co-author network of economists has these two properties. Further, if we control for the effect of these properties, we show that there is a weak residual effect in support of Granovetters's theory. These findings provide evidence for our explanation of the rejection of hypothesis 2A in Section 3.

The two crucial network properties to understand the rejection of hypothesis 2A are as follows. The first property is the existence of stars, that is, actors have a high number of links. These stars play the role of connectors in the network (Albert et al. 2000). That is, they connect different subgroups in the network that would be isolated if the stars were not there. Given their crucial role in the network, the stars have a high betweenness centrality, and the links between stars have high link betweenness.

The second property is that the links between stars are stronger than other links. Because of the central position of the stars 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.

We explain the working of the above two properties by considering two types of network structures, the island network structure and the core-periphery network structure. 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. There are three islands. Each island consists of four nodes, and all four nodes on an island are directly connected to each other with a strong tie. Furthermore, different islands are connected with exactly one weak tie. In a stylized way this network represents a view of the world which has been often put forward when explaining the 'strength of weak ties' theory[9]. 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. In fact, Granovetter puts forward this view of the world when he discusses the effects of weak ties on the ability of communities to organize themselves (Granovetter 1973, pp. 1373–1376).

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; in Figure 4 $4 \times 4 = 16$ 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.

While Granovetter's 'strength of weak ties' theory typically holds in an island network structure, we now show that the theory fails to hold for another type of network. Consider the network 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 periheral actors themselves have only a weak link to one of the core actors, and no link to other peripheral actors. We refer to this network structure as the core-periphery network structure. In contrast to the island network structure, the core-periphery network has a distinct hierarchy.

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

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 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, we argue that the bridges are more 'crucial' in the island structure than in the core-periphery network structure.

5 The network structure of the co-author network

We now show that the co-author network has the two properties mentioned in Section 4, and that these two properties indeed explain the rejection of Hypothesis 2A in Section 3.

We first refer to our earlier work of Goyal, van der Leij and Moraga-González (2005). In this work we showed that the co-author network has a very unequal degree distribution. Most economists have only one or two links, but there is small fraction of *stars* with many, up to 50 links. Furthermore, we showed that the removal of these high degree 'stars' 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, Jeong and Barabási 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.

We now show that there exists a strong and positive correlation between strength and average degree. 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 7.

We observe that there is a 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.

To support our claim that it is indeed the two above observations that explain the rejection of hypothesis 2A, we repeat the regression of link betweenness on strength of a link; however, this time we add variables on average degree and difference in degree to the regression. Thus we control for the indirect effect that strong links are often between high degree nodes, and a link between these high degree nodes have a high link betweenness. If our explanation is valid, then we would expect the degree variables to be highly significant. Furthermore, our explanation for the rejection of Hypothesis 2A becomes irrelevant when we compare ties with the same degree of its actors. In that case, Granovetter's arguments should become more prominent again. Thus, after we control for the degree of the tie's actors, we should expect Hypothesis 2A to hold. That is, we should expect a negative coefficient for the strength variable.

We again stack observations on links in the giant components of the networks in the three decades and we again take 10000 random subsamples of 1000 observations. We then perform regressions on the subsamples with log betweenness as dependent variables and decade dummies, log strength, average log degree and difference log degree as explanatory variables. Table 8 shows the results of these regressions. We first observe that R^2 increases to .27 from .06 in the earlier regression reported in Table 4. Hence, the variation in link degree explains a large portion of the variation in link betweenness. Second, we observe that the coefficient of $AVGDEGREE$ is significantly positive, thus showing that a link between two high degree nodes has a higher betweenness. Third, we observe that the coefficient of $LNSTRENGTH$ is negative, although not significantly. These results clearly support our intuition that it is the indirect relation of between strong ties and high degree actors and between high degree actors and high betweenness that explains the results in Section 3, and that this indirect effect dominates the effect of high clustering of strong ties on betweenness.

6 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. 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.

It should be noted that we analyzed only one particular dataset, the co-author network of economists. This network is quite specific. 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. It is probable that this drives some of the results. In fact, this suspicion is supported in a supplementary analysis, in which we analyzed the network of economists with at least 5 articles. For this subnetwork of 'active' economists we found that both hypotheses 1 and 2 are supported.

The above result raises the question if we can induce our conclusions to other social settings outside the (economic) academic community. That is, do other social networks have the same properties, stars with strong ties, as the co-author network of economists? And, if so, are strong ties in these networks also more important than weak ties? Further research on other datasets should shed light on that. However, we have some clues that our conlusions indeed hold more generally.

First, we draw the attention to the results of Borgatti and Feld (1994) who performed a small test on the 'Strength of weak ties' theory on data of the friendship network of karate club members (Zachary, 1977). Their results are very similar to ours. In fact, they conclude that their "(...) results do not support (Granovetter's) theory, and instead suggest that strong ties tend to occur between actors who have many links in general (...)" (Borgatti and Feld, 1994, p.46). Thus, they make the same conclusions as we do, even though the data is very different.

Second, physicists have analyzed a number of large social networks, and they have found that inequality of links is a common property of these networks (Newman, 2003). This research typically considers a tie a dychotomous variable. Hence, results on the strength of ties and its relation to the degree of nodes is scarce (Barrat et al. 2004). Further research on other data sets should reveal if the relation between strength and degree holds more generally. It would be most interesting to do this for job contact networks in East Asia, as this region strong ties are typically found to be more important for job seekers (Bian and Ang, 1997).

A final issue for future research is the implication of the network patterns we found on information dissemination. As we only considered properties in the network structure, it remains an open question whether strong are indeed more important for information dissemination. For the co-author network of economists a combined analysis of co-author network and citation data could reveal if this is really true.

References

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

[2] Albert, R. Jeong, H. and A.-L. Barabási (2000), Error and attack tolerance of complex networks, *Nature*, 406, 378–382.

[3] Barabási A.-L. and R. Albert (1999), Emergence of scaling in random network, *Science*, 99, 509–512.

[4] Barrat, A., M. Barthélemy and A. Vespignani (2004a), Modeling the evolution of weighted networks, *Physical Review E*, 70.

[5] Barrat, A., M. Barthélemy, R. Pastor-Satorras and A. Vespignani (2004b), The architecture of complex weighted networks, *Proceedings of the National Academy of Sciences of the U.S.A.*, 101(11), 3747–3752.

[6] Bian, Y. and S. Ang (1997), Guanxi networks and job mobility in China and Singapore, *Social Forces*, 75(3), 981–1005.

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

[8] Dorogovtsev, S. and J. Mendes (2002), Evolution of networks, *Advances in Physics*, 51, 1079-1187.

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

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

[11] Girvan, M. and M. Newman (2002), Community structure in social and biological networks, *Proceedings of the National Academy of Sciences of the U.S.A.*, 99, 8271–8276.

[12] Goyal, S., M.J. van der Leij and J.L. Moraga-González (2005), Economics: an emerging small world, *Journal of Political Economy*, forthcoming.

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

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

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

[16] Newman, M. (2001),The structure of scientific collaboration networks, *Proceedings of the National Academy of Sciences*, 98, 404-409.

[17] Newman, M. (2003), The structure and function of complex networks, *SIAM Review*, 45, 167–256.

[18] Newman, M. (2004), Analysis of weighted networks, *Physical Review E*, 70, 056131.

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

[20] Watts, D. and Strogatz, S. (1998), Collective dynamics of 'small world' networks, *Nature*, 393, 440–442.

[21] Yook, S.H., H. Jeong, A.-L. Barabási and Y.Tu (2001), Weighted evolving networks, *Physical Review Letters*, 86(25), 5835–5838.

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

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 chi^2-independence test (2 degrees of freedom).

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

variable	coefficient	.95-confidence	p-value
DUMMY70S	-1.459	(-2.086, -0.933)	0
DUMMY80S	-1.534	(-1.917, -1.179)	0
DUMMY90S	-1.613	(-1.866, -1.371)	0
AVGSTRENGTH	.751	(.095, 1.343)	.027
DIFFSTRENGTH	-.522	(-.996, -.056)	.027
loglikelihood	-455.19		

Results of regressions on 10000 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. *coefficient* is the average of the estimated coefficients in the 10000 regressions. *.95-confidence* is the .025- and .975-quantile of the coefficients estimated from the 10000 regressions. *p-value* is $1 - |2s - 1|$ where s is the fraction of estimated positive coefficients.

Table 4: Estimation results of a regression on link betweenness

variable	coefficient	.95-confidence	p-value
DUMMY70S	-7.870	(-8.451, -7.345)	0
DUMMY80S	-8.954	(-9.313, -8.624)	0
DUMMY90S	-9.899	(-10.140, -9.666)	0
LNSTRENGTH	.480	(.108, .825)	.013
R^2	.060		

Results of regressions on 10000 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. *coefficient* is the average of the estimated coefficients in the 10000 regressions. *.95-confidence* is the .025- and .975-quantile of the coefficients estimated from the 10000 regressions. *p-value* is $1 - |2s - 1|$ where s is the fraction of estimated positive coefficients.

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: Estimation results of a regression on the strength of ties

variable	coefficient	.95-confidence	p-value
DUMMY70S	.0244	(-.0854, .1369)	.666
DUMMY80S	.0229	(-.0660, .1127)	.613
DUMMY90S	.0030	(-.0818, .0863)	.943
AVGDEGREE	.1827	(.1258, .2402)	0
DIFFDEGREE	-.0411	(-.0762, -.0048)	.027
R^2	.057		

Rresults of regressions on 10000 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. *coefficient* is the average of the estimated coefficients in the 10000 regressions. *.95-confidence* is the .025- and .975-quantile of the coefficients estimated from the 10000 regressions. *p-value* is $1 - |2s - 1|$ where s is the fraction of estimated positive coefficients.

Table 8: Estimation results of a regression on link betweenness controlling for degree

variable	coefficient	.95-confidence	p-valu
DUMMY70S	-11.160	(-11.985, -10.374)	0
DUMMY80S	-12.562	(-13.322, -11.810)	0
DUMMY90S	-13.807	(-14.588, -13.028)	0
LNSTRENGTH	-.084	(-.422, .227)	.631
AVGDEGREE	2.410	(2.096, 2.728)	0
DIFFDEGREE	.607	(.363, .857)	0
R^2	.270		

Results of regressions on 10000 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. $AVGDEGREE = (X_A + X_B)/2$ 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. *coefficient* is the average of the estimated coefficients in the 10000 regressions. *.95-confidence* is the .025- and .975-quantile of the coefficients estimated from the 10000 regressions. *p-value* is $1-|2s-1|$ where s is the fraction of estimated positive coefficients.

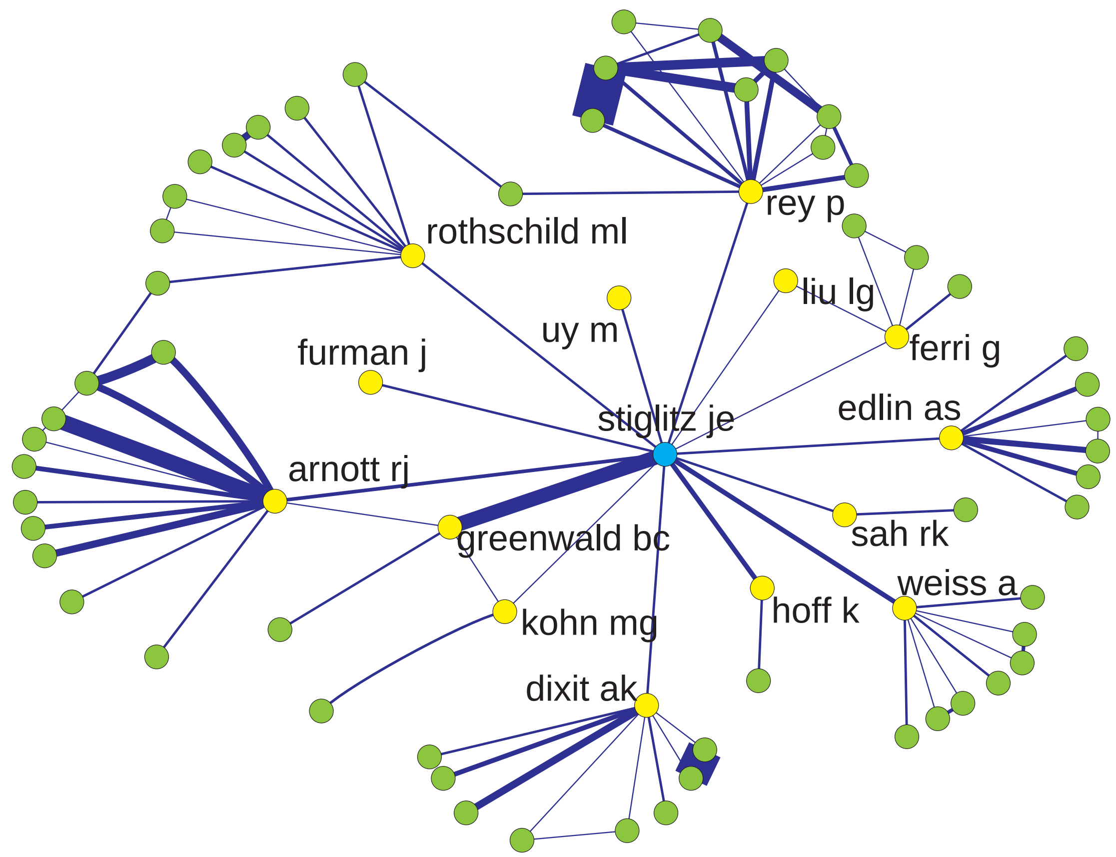


Figure 1: Local network of collaboration of Joseph E. Stiglitz in the 1990s.

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*.

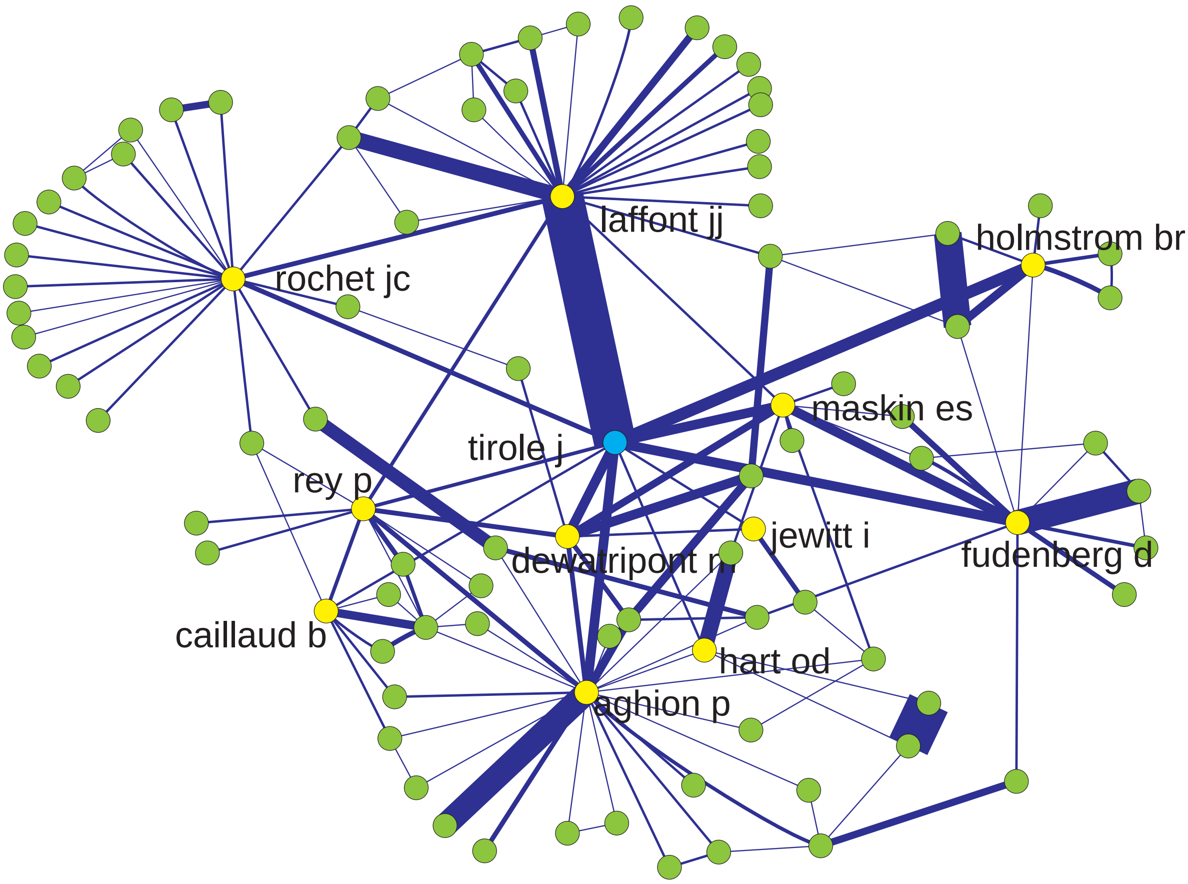
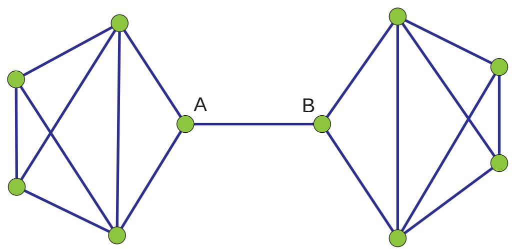
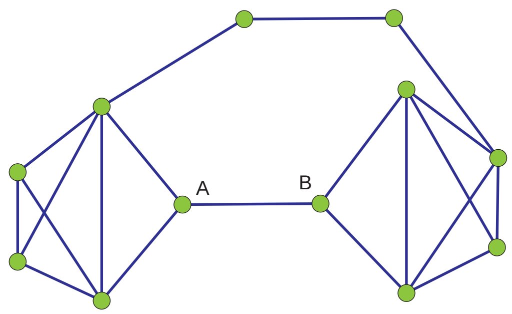


Figure 2: Local network of collaboration of Jean Tirole in the 1990s.

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.

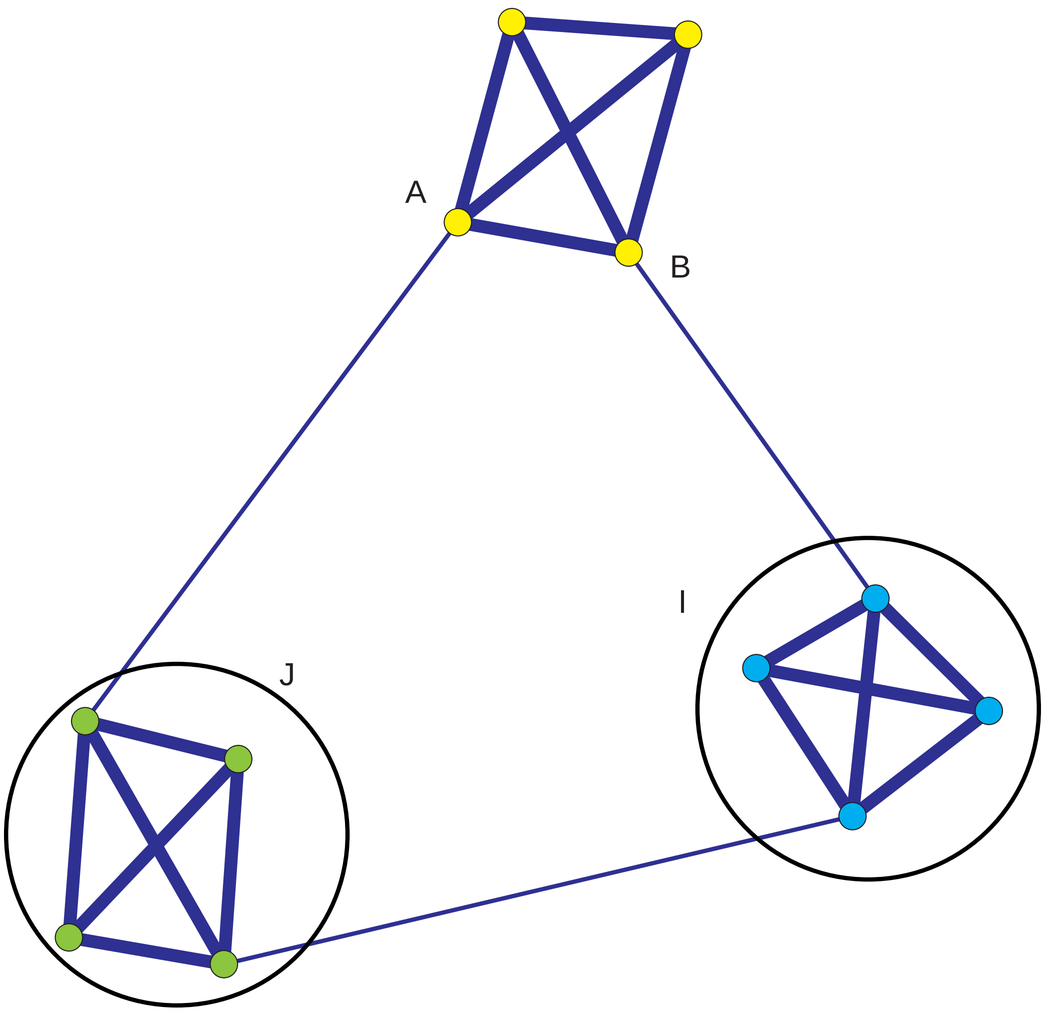


Figure 4: An island network.

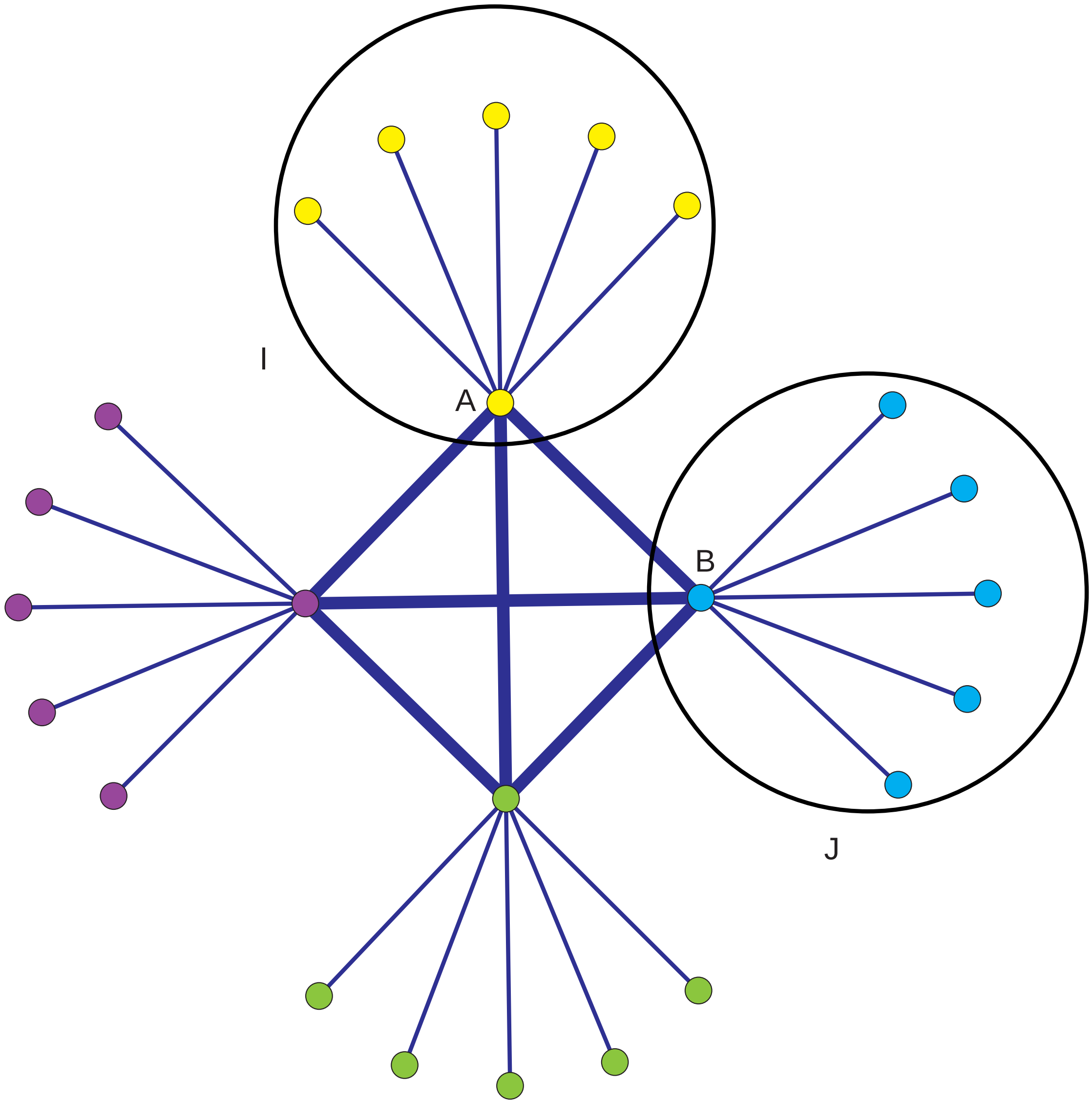


Figure 5: A core-periphery network.