Globalization, Education, and the Topology of Social Networks

Adrien Vigier*

September 17, 2008

Abstract

The present paper suggests a possible framework to analyze the impact of changes to the economic and social environment on the topology of networks formed. Economic (costs) and social (norms) constraints bind individuals in their ability to create ties with others. When global phenomena affect these constraints, the overall shapes of resulting networks naturally alter. I attempt to shed light on this relationship.

1 Introduction

The importance of social structure in determining economic outcomes is widely documented. Word of mouth communication plays an important role in disseminating information about products, prices, and quality (Katz and Lazarsfeld (1955)), network effects pervade the adoption and spread of new technologies (Conley and Udry (2001)), and individuals often rely on friends or acquaintances to obtain information on job opportunities (Granovetter (1973)) to name only a few examples.

The present paper aims to evaluate the way in which changes to the economic and social environment naturally translate into changes in the networks of interactions formed. The latest wave of globalization, for example, is redefining the face of communication on a world scale. Not only are people able to make long distance phone calls at very little cost, but the

*Faculty of Economics and Queens' College, Cambridge CB3 9ET. Email: av301@cam.ac.uk. I am greatly indebted to my supervisor, Sanjeev Goyal, for making himself available throughout the writing of the present paper. I am also grateful to Adam Clark-Joseph, and Willemien Kets for their careful reading of an early version of the present paper and wish to thank Matthew Jackson, Paolo Pin, and Marco Van der Leij, for giving me useful comments and suggestions. Lastly, the author gratefully acknowledges financial support from the Gates Cambridge Trust. All remaining errors are of course mine.

widespread learning of the English language across countries and cultures mean that it is now possible for many to maintain relationships, private or professional, with others on the other side of the globe. The impact of education on racism and tolerance is another important example. Psychologists hold that fear resulting from mis-information (or lack of) is one of the most important factor explaining discriminatory behaviour vis-à-vis other identity groups[1]. As a response, public funding of awareness schemes have developed on a large scale in an attempt to ease tensions between communities and promote the normalization of dialogue across them. In both cases, it is natural and compelling to enquire about the transformations implied at the global level.

To this end, we take a broad 'geographic' approach by assuming that nodes are located in some underlying s-dimensional Euclidian space and derive positive utility solely from forming ties with others sufficiently close to them within the metric. The usefulness of this approach stems from the fact that distances may be customized to account for economic constraints. For example, a rise in the costs of maintaining links may be modeled as an effective inflation of distances. As distances are inflated, the number of nodes lying within the radius of positive utility of any given node falls. Each node therefore finds it increasingly difficult to form links.

To model the network formation *per se*, we borrow from the recent work of Vazquez (2003) and Jackson and Rogers (2007). Nodes are introduced into the network sequentially and meetings proceed from a combination of random and network-based devices. Each new node entering the network is first introduced to a subset of nodes picked at random from the existing set. She then goes on to meet some of the neighbours of those randomly chosen nodes. The Jackson-Rogers-Vazquez (henceforth JRV) framework is both simple and intuitive, however its greatest credentials lie in its ability to generate networks which reproduce accurately their empirical counterparts[2]. In view of our objective, the JRV framework is therefore a natural one to base our analysis upon.

Our main results relate the topology of the networks formed and the underlying economic constraints. First, I show how the constraints affect the distribution of links arising. Second, and contrastingly, I show that clustering is essentially determined by the dimension s of the

[1] "Le Racisme", Que sais-je n° 1603.

[2] See, e.g., Vega-Redondo (2007). In short, a large number of empirical networks tend to exhibit the following regularities: (i) short average distance, (ii) degree distribution exhibiting a power law in the tail, (iii) high clustering, (iv) assortativity, and (v) negative clustering-degree relationship.

underlying space. We are thus able to make sharp predictions regarding some important aspects of network topology following changes to the underlying economic environment.

Next I show that in a geographic context and when nodes are searching potential partners locally in the network, the intensity of the constraints is unimportant in the sense that having more stringent constraints does not reduce the prospects of finding matching nodes. This adds importantly to previously recorded motivations for using local search in models of network formation, particularly in hostile environments for which constraints are tight.

Lastly, recent empirical evidence (Goyal, Moraga, and van der Leij (2007)) seems to indicate that, at least for some networks evolving over time, the number of links formed using the network tends to grow faster than links formed at random when the overall density of the network rises. I show how this phenomenon can be accounted for in a simple way within the model presented in this paper. Essentially, when constraints are relaxed nodes' access to the network is enhanced.

The present paper is related to the economics literature on network formation, introduced by the work of Jackson and Wolinsky (1996), and Bala and Goyal (2000)[3]. Models using the idea of an underlying metric have been developed by Gilles and Johnson (2000), and Galeotti, Goyal, and Kamphorst (2006). The focus of these papers however is largely distinct from ours since they are concerned with the formation of 'small networks' for which standard game-theoretic assumptions may reasonably be expected to hold. This paper, on the other hand, incorporates a large amount of bounded rationality on the part of agents involved in the process of network formation. In the spirit of Vega-Redondo (2007) we assume that in such a complex environment nodes follow a number of simple rules.

This paper is also related to the work of Strogatz and Watts (1998) in which nodes populating a ring lattice connect their closest neighbours. The issues these authors address are however wholly distinct from ours since they focus on the sharp impact that introducing some long range links may have on average distances in the network. The literature on random geometric graphs finally (Penrose (2003)), shares some common features with the present paper. In the former, nodes are placed randomly in Euclidian space and edges added to connect points that are close to each other. The absence of time, and of any kind of local search, mark however some important differences with our work.

The rest of the paper is organized as follows. Section 2 presents the model and gives some

[3] See Goyal (2007) for an overview of this literature.

preliminary results. The topological analysis is carried out in Section 3. Section 4 applies some of our results to the data analyzed by Goyal, Moraga, and van der Leij (2007) as an illustrative example of their empirical applicability. Section 5.1 discusses geodesic distances in the model. Section 5.2 discusses the model vis-à-vis its non-geographic counterpart. Important extensions, including higher dimensions, are discussed in Section 5.3. Section 6 concludes.

2 The model

Consider a countably infinite set of nodes labelled according to $\mathbb{N}$. For $i \in \mathbb{N}$, X_i[4] defines a random variable locating node i in some s-dimensional Euclidian space S. Furthermore, we assume that $(X_i)_{i\in\mathbb{N}}$ are uniformly and independently distributed in S. Nodes derive utility $u_i(g, (x_i)_{i\in\mathbb{N}})$ from forming ties with others according to network g. Let ε, fixed for society as a whole, determine the neighbourhood $\Omega_i(x_i)$ in S within which agent i derives positive marginal utility from being linked with others. That is, we assume

$$\begin{aligned} \frac{\partial u_i}{\partial n_{\Omega_i}} &> 0 \\ \frac{\partial u_i}{\partial n_{\Omega_i^c}} &< 0 \end{aligned} \tag{1}$$

where $n_{\Omega_i} = \#\{j|i \text{ } and \text{ } j \text{ } linked, \ x_j \in \Omega_i(x_i)\}$, and $n_{\Omega_i^c} = \#\{j|i \text{ } and \text{ } j \text{ } linked, \ x_j \in \Omega_i^c(x_i)\}$. Unless stated otherwise we assume for simplicity that S is identified with the torus of unit length, and that $\Omega(x)$ is given by the interval of length ε centred at x. Higher dimensional S are discussed in Section 5.3.1. Notice that $x_j \in \Omega_i(x_i)$ if and only if $x_i \in \Omega_j(x_j)$. We shall say that i and j match whenever the previous conditions hold. Also, since nodes are symmetric under our assumptions we may write $\Omega(x_i)$ instead of $\Omega_i(x_i)$. Finally, we require $\varepsilon << 1$.

Nodes enter the world sequentially, one at a time, so that node t also enters the world at time t. Upon entrance, each new node randomly meets m_r existing nodes chosen uniformly at random from the current set of nodes $\{1, ..., t-1\}$. Nodes also meet some of the neighbours of their random meetings by following each of their (outgoing) links independently and with probability λ. We emphasize here the distinction between neighbourhoods in S and

[4]Upper case symbols are used to indicate random variables, while corresponding lower case symbols indicate realizations of the random variables.

neighbourhood of a node i in network g. The first concept is topological, while the second refers to the subset of nodes which are linked to i. It follows from (1) that whenever node i meets another node j such that $x_j \in \Omega(x_i)$ a directed link is formed from i to j, which we denote by ij. Throughout, we refer to the links initiated (received) by node i as the outgoing (incoming) links of i. Note finally the clear distinction made between the 'meeting' and 'matching' processes which jointly constitute the formation process. On the one hand the meeting process determines which pairs of agents are introduced to one another. The matching process on the other hand determines which of the former pairs give rise to a link being formed.

Since two nodes can only form a link if they find themselves within distance ε in S, proximity in the network also conveys valuable information regarding nodes' relative positions in S. In particular, nodes are much more likely to form links in the neighbourhood of those that they matched with rather than those with whom they did not. Our first result makes this statement precise. Proposition 1.(ii) gives the probability that a node matches with some other node's neighbour, given that she has matched with the former. Proposition 1.(iii) gives the same probability given that she has not matched with the parent node[5]. Proposition 1.(i) simply states the obvious probability that a node matches with some random other. All proofs are relegated to the appendix.

Proposition 1 *Let i, j , and $k \in \aleph$. Matching probabilities are given by*

$$\begin{aligned}
(i)\ \ \Pr(X_j &\in \Omega_i) = \varepsilon \\
(ii)\ \ \Pr(X_k &\in \Omega_i \mid (X_j \in \Omega_i) \wedge (X_k \in \Omega_j)) = 3/4 \\
(iii)\ \ \Pr(X_k &\in \Omega_i \mid (X_j \in \Omega_i^c) \wedge (X_k \in \Omega_j)) = \frac{\varepsilon}{4(1-\varepsilon)}
\end{aligned}$$

Note that Proposition 1 shows how the geographic model analyzed here provides additional motivation for using local search in models of network formation. By Proposition 1.(ii), the probability of matching with a neighbour's neighbour is independent of ε in the model. In hostile environments, as ε tends to zero, this means that friends of my friends provide a particularly favourable pool of individuals with whom to form new links[6].

[5] 'Parent' node in this context is used to indicate that node which outgoing link was used to generate the network-based meeting under consideration.

[6] Consider the connections model of Jackson and Wolinsky (1996) in which the benefit to node i from being connected to node j is given by $\delta^{d(i,j;g)}$, where $d(i,j;g)$ indicates geodesic distance between nodes i and j in network g. With these payoffs

Next, the number of nodes met during the random stage is fixed and equal to m_r. However, since the number of nodes met through the network is unbound, conditions should be imposed that guarantee convergence of our process for t large. Following the literature on stochastic network formation we henceforth make extensive use of mean-field approximations on the premise that doing so greatly simplifies the analysis[7]. Our next result states the conditions under which the average number of links formed by entering nodes approaches a steady state as $t \to \infty$.

Proposition 2 *Let m_t denote the number of (outgoing) links formed by node t. In the mean-field approximation, $E[m_t]$ approaches a steady state as $t \to \infty$ provided $\lambda m_r \varepsilon < 1$.*

In what follows we assume that the condition $\lambda m_r \varepsilon < 1$ is always satisfied, and let m denote the steady state average number of outgoing links. Upon entering the network each new node forms an average $m_r \varepsilon$ links with random nodes. By Proposition 1.(ii), she also forms an average $\frac{3}{4}\lambda m_r \varepsilon m$ links in the neighbourhoods of these nodes, along with a further $\frac{\varepsilon}{4(1-\varepsilon)}\lambda\ m_r(1-\varepsilon)\,m$ links in the neighbourhoods of the nodes she met randomly but failed to form a link with. Notice that the ratio of the average number of links formed respectively in the neighbourhoods of matching and non-matching parent nodes is $3:1$. Finally, adding contributions from the random and network-based processes gives

$$m = m_r\varepsilon(1+\frac{3}{4}\lambda m) + m_r(1-\varepsilon)\left[\lambda m \frac{\varepsilon}{4(1-\varepsilon)}\right] = m_r\varepsilon(1+\lambda m)$$

from which

$$m = \frac{m_r\varepsilon}{1-\lambda m_r \varepsilon} \tag{2}$$

Naturally, increases in any of m_r, ε, or λ raises the expected number of outgoing links as indicated by (2). Notice also that $m \to \infty$ as $\lambda m_r \varepsilon \to 1$.

A natural partition of links in our model results from the distinction made between random and network-based links. Following Jackson and Rogers (2007) we may define the ratio r of, respectively, the average number of random and network-based links in the network

in our model, the expected benefit from meeting a neighbour's neighbour is given by $\frac{4}{3}(1-\delta)$, while for t large the expected benefit from meeting a random node is ε. For equal meeting costs, local search may therefore be optimal even in situations of weak informational decay (provided $\delta < 1 - \frac{4}{3}\varepsilon$) for which the perfect monitoring model warrants link formation with distant nodes instead.

[7] For an account of the performance of mean-field approximation in statistical models of network formation the reader is referred to Vega-Redondo (2007).

formation process

$$r = \frac{average\ number\ of\ random\ links}{average\ number\ of\ network - based\ links} \tag{3}$$

In particular, the r statistic provides a useful measure of randomness in this kind of models. Using our previous results we obtain

$$r = \frac{m_r\varepsilon}{m_r\varepsilon\left(\frac{3}{4}\lambda m\right) + m_r\left(1-\varepsilon\right)\left[\lambda m\frac{\varepsilon}{4(1-\varepsilon)}\right]} = \frac{1}{\lambda m}$$

i.e.

$$rm = \frac{1}{\lambda} \tag{4}$$

Equation (4) is a distinguishing feature of the present model. It indicates that denser networks also tend to be less random. As later emphasized in Section 4, this result proves important in view of empirical applications. Combining (2) and (4) we have $r = \frac{1-\lambda m_r\varepsilon}{\lambda m_r\varepsilon}$, which shows that increases in any of m_r, ε, or λ reduces r. Intuitively, raising ε improves nodes' access to the network. It should therefore not be too surprising that raising ε also reduces the relative importance of randomness in the model.

The following proposition summarizes the above observations

Proposition 3 *A change in any one parameter keeping other parameters fixed induces opposite shifts on m and r respectively. In particular, a rise in any one of m_r, ε, or λ leads to greater network density and falling randomness.*

3 Topological Analysis

3.1 Degree Distribution

The sequential addition of nodes in the JRV framework naturally introduces heterogeneity among nodes in the network. Whereas all nodes have identical expected out-degree in the formation process, older nodes accumulate incoming links for a longer period of time and are therefore more likely to exhibit high in-degrees. In what follows we analyze the in-degree distribution of nodes resulting from our formation process.

Let $d_i(t)$ denote the in-degree of node i at time t. The probability of node i receiving a new link at time t is obtained by adding the probability of receiving a link through random selection with the probability of receiving a link through network-based meeting. The first term is (approximately) $(\frac{m_r}{t})\varepsilon$, the probability of her being randomly selected times the matching probability of two random nodes. In the same vein the second term is $m_r \lambda m \varepsilon (\frac{d_i(t)}{mt})$, each node receiving a share of the total expected $m_r \lambda m \varepsilon$ network-based links formed in proportion to her own in-degree. The total probability of node i receiving a new link at time t is therefore given by

$$\frac{m_r \varepsilon}{t} + (\frac{\lambda m_r \varepsilon}{t}) d_i(t) \tag{5}$$

Using (2) and (4), we may rewrite (5) as

$$\frac{1}{t}\left(\frac{r}{1+r} m + \frac{1}{1+r} d_i\right) \tag{6}$$

highlighting the respective shares of randomness and network-based linking in the process.

In the mean-field approach $d_i(t)$ is a 'probabilistic stock' variable, the evolution of which is described in continuous time by the 'probabilistic flow' given by (5). The random system is thereby transformed yielding a set of deterministic ordinary differential equations of the form

$$\frac{dd_i(t)}{dt} = \frac{m_r \varepsilon}{t} + (\frac{\lambda m_r \varepsilon}{t}) d_i(t) \tag{7}$$

Notice the proportionality in degree-growth in (7). Lastly, initial conditions for all i solve for $d_i(t)$, $\forall\, i, t$. The degree distribution follows immediately. Details of the proof are relegated to the appendix.

Theorem 1 *As $t \to \infty$, in the mean-field approximation, the cdf of the in-degree distribution tends to*

$$F(d) = 1 - (\frac{\lambda^{-1}}{\lambda^{-1} + d})^{1/\lambda m_r \varepsilon} \tag{8}$$

Notice that for d large the in-degree distributions resulting from our formation process approximate power laws. As shown by Albert and Barabasi (1999), power-law degree distributions follow from proportionality in degree-growth. In our model, as in the standard JRV framework, proportionality in degree growth results from network-based linking. Lastly, notice that for $\lambda \sim 0$ the network formation process approaches one of uniform random linking[8]. In that case it is easily shown that an exponential distribution obtains[9].

The following corollary provides important comparative statics results regarding the degree distribution obtained in (8)

[8] In our model, only by acting on λ can network-based links be affected independently of random links.

[9] See, e.g., Vega-Redondo (2007).

Corollary 1 *Let F and $F^{'}$ denote the cumulative distribution functions of the formation processes with parameters $(m_r,\ \varepsilon,\ \lambda)$, and $(m_r',\ \varepsilon',\ \lambda')$ respectively,*

(i) If $(m_r',\ \varepsilon',\ \lambda') > (m_r,\ \varepsilon,\ \lambda)$, then F' strictly first order stochastically dominates F.

(ii) Given $(m_r',\ \varepsilon',\ \lambda')$, and $(m_r,\ \varepsilon,\ \lambda)$ such that $m' > m$ and $r' = r$ then F' strictly first order stochastically dominates F.

(iii) Given $(m_r',\ \varepsilon',\ \lambda')$, and $(m_r,\ \varepsilon,\ \lambda)$ such that $r' > r$ and $m' = m$ then F' strictly second order stochastically dominates F.

Corollary 1 has important welfare implications. Such a systematic analysis however is besides the focus of the present paper. The reader is referred to Jackson and Rogers (2007) for a detailed discussion of these issues.

Before turning to the next section notice that, following Corollary 1, increases in r may be accompanied by second order stochastic dominance or its opposite. As noted above, fat-tails result from network-based links. When average degree is kept fixed, a rise in r signals a transfer from the contribution of network-based links to that of random links which therefore reduces the spread in the distribution of degree. On the other hand when the rise in r mirrors a fall in m, Corollary 1.(i) indicates a dominance of the mean effect.

3.2 Clustering

There exist a number of possible measures of clustering, each defining a variation on a theme. In the model we present, one measure naturally imposes itself. For a given network g, it indicates the fraction of times the dotted connection in Figure (A) of the appendix exists given the pair of bold links, i.e.

$$C(g) = \frac{\sum\limits_{i;j\neq i;k\neq j} g_{ij}g_{jk}g_{ik}}{\sum\limits_{i;j\neq i;k\neq j} g_{ij}g_{jk}} \tag{9}$$

where $g_{ij} = 1$ if link ij exists in g and $g_{ij} = 0$ otherwise.

More intuitively, $C(g)$ can be expressed in terms of the percentage of times $\Xi(g)$ that two of a node's neighbours are linked (see Figure (B)),

$$\Xi(g) = \frac{\sum\limits_{i;j\neq i;k\neq j} g_{ij}g_{jk}g_{ik}}{\sum\limits_{i;j\neq i;k\neq j} g_{ij}g_{ik}} = \frac{\sum\limits_{i;j\neq i;k\neq j} g_{ij}g_{jk}g_{ik}}{\sum\limits_{i;j\neq i;k\neq j} g_{ij}g_{jk}} \frac{\sum\limits_{i;j\neq i;k\neq j} g_{ij}g_{jk}}{\sum\limits_{i;j\neq i;k\neq j} g_{ij}g_{ik}} = C(g)\frac{\sum\limits_{i;j\neq i;k\neq j} g_{ij}g_{jk}}{\sum\limits_{i;j\neq i;k\neq j} g_{ij}g_{ik}} \tag{10}$$

We show in the appendix that for t large clustering may be approximated using the following expression

$$C = \frac{E(\text{triplets per node } i)}{m^2} \tag{11}$$

The average number of triplets per node can be calculated, in the mean-field approximation, by separating out the situations according to whether j, and k, were met randomly or through the network. The following Theorem is proven in the appendix.

Theorem 2 *As $t \to \infty$, in the mean-field approximation*

(i) $\sup_\varepsilon C = \lim_{\varepsilon \to 0} C = \frac{3\lambda}{4}$.

(ii) $\frac{\partial C}{\partial \varepsilon} < 0$ *where continuity holds.*

(iii) $\frac{\partial C}{\partial \lambda} \lesseqgtr 0$ *depending on parameter values.*

Interestingly, in a similar model Strogatz and Watts (1998) previously simulated a lowest upper bound for clustering equal to $\frac{3}{4}$. In view of Proposition 1.(ii), it is easy to see why such a result holds. In fact, referring back to (9) one can see that Proposition 1.(ii) delivers an immediate measure of clustering provided the formation process is such that a node meets all the neighhbours of those she links to. In the model presented here each new node only meets her random meetings' neighbours so that the former condition is only satisfied provided network-based links can safely be ignored, which occurs as $\varepsilon \to 0$ since in that case $r \to \infty$. This explains point (i) in Theorem 2.

As ε moves away from zero, network-based links start forming a non-negligible part of the total number of links existing. Since entering nodes do not meet the neighbours of the nodes they met through the network[10], the number of situations depicted in Figure (A) where the dotted link never materializes increases importantly too, thereby triggering downward pressure on overall clustering. While on the one hand this explains point (ii) of the Theorem, the same argument also gives an important sense in which the determinants of clustering are largely independent from the intensity of the constraints in the underlying space S. Indeed, without loss in the meeting process, changes to ε do not affect clustering. In Section 5.3 I show that the key element determining the level of clustering is the dimension of the underlying space S.

[10]Naturally, due to neighbourhood overlap in the process they may effectively meet some of them.

Our last comment concerns the finite discontinuity points observed regarding clustering. Clustering arises in the JRV framework as a result of local search. A triplet is formed as soon as a node randomly links another and goes on to form a link with some neighbour of that node. We term this process 'first order clustering'. However, when the average number of links formed in the neighbourhood of matching parent nodes is greater than one, more triplets are formed due to the fact that two neighbours of a given node have a non-trivial probability of being linked (so long as some amount of (first order) clustering already exists). We refer to this process as 'second order clustering' since it arises only when some clustering already prevails in the network. Clustering therefore exhibits discontinuous jumps whenever the average number of links formed, respectively, in the neighbourhoods of matching and non-matching parent nodes reach the critical value 1. Numerical estimates show that the effect of second order clustering can lead to discontinuous jumps which are large in magnitude. For $\lambda = 1$, and $m_r = 10$ we go from less than a third of triplets realized to well over a half at the threshold.

4 Empirical applications

Our model constitutes a powerful instrument for the empirical investigation of networks. First, by Theorem 1, $\lambda m_r \varepsilon$ can be obtained from the slope of the degree distribution. Second, using equation (1) along with the observed value of the average degree m gives $m_r \varepsilon$. Third, λ is retrieved by taking the ratio of $\lambda m_r \varepsilon$ and $m_r \varepsilon$. Notice that at this point equation (3) provides an estimate of r. Finally, the biggest difficulty lies in separating the effects of m_r and ε. Closer investigation of the model reveals that the only place in which the effect of ε is singled-out is in second-order clustering arising in the neighbourhoods of non-matching parent nodes. This poses two problems. Firstly, it means that for a large range of parameter values separate estimation of m_r and ε is precluded. Secondly, inspection of Theorem 2 reveals that the necessary condition for retrieving ε from $C(g)$ requires itself knowledge of ε to be evaluated. This means that even in the most favorable range of parameters we can hope for no more than a simple test of non-inconsistency. The impact of this limitation depends on the particular question one aims to address and the assumptions one is willing to make about the formation process. For example, in many situations $m_r = 1$ naturally suggests itself given the nature of the problem. Following the above procedure then delivers

ε.

As an illustrative example, we apply some of our results to the data analyzed by Goyal, Moraga, and van der Leij (2007) concerning the evolution of coauthorship in the economics literature through the 1970's, 80's, and 90's. The coauthor network provides a natural platform for empirical applications of our model. First, the time-sequencing and meeting process of JRV provide a very intuitive description of their real-world counterpart in the context of coauthorship. Second, the amount of interaction involved in such relationships is such that communication costs are likely to play a prominent role in the determination of matching outcomes. Third, the availability of data dating back to 1970 is particularly relevant in our context considering the way in which communication costs have evolved over that period. The question therefore is whether our model is able to shed light on the evolution of coauthorship over the last decades.

Table 1 below is adapted from Goyal, Moraga, and van der Leij (2007). First, notice the negative relationship exhibited between randomness and average degree. Second, while network density trends upward through time, λ repeatedly falls from one decade to the next. Provided nodes attempt to achieve a target number of links, it is easy to see that λ and m should indeed evolve in opposite directions in our model[11].

In view of our interpretation, these results therefore suggest that sharp falls in communication costs over the period under consideration largely contributed in lifting geographical barriers to coauthorship. While the resulting positive impact on coauthorship likely lead researchers to cooperate with a lesser proportion of their own coauthors' colleagues, increases in the average number of links indicate a dominance of the first effect. Falling randomness is then accounted for in our interpretation by agents' improved access to the network.

	70's	80's	90's
m	.445	.622	.836
r	2.94	2.70	2.49
λ	.76	.59	.48

Table 1

[11] Suppose agents tailor their behaviour so as to achieve an expected number of (outgoing) links equal to K (possibly due to resource constraints or other), taking other agents behaviour as given. Letting m^* denote average degree in the network, the expected number of links formed as a function of λ is given by $m_r\varepsilon(1+\frac{3}{4}\lambda m^*) + m_r(1-\varepsilon)\left[\lambda m^* \frac{\varepsilon}{4(1-\varepsilon)}\right] = m_r\varepsilon(1+\lambda m^*)$. Setting $m_r\varepsilon(1+\lambda m^*) = K$ then yields $\lambda = \frac{K - m_r\varepsilon}{m_r\varepsilon m^*}$, showing that $\frac{\partial\lambda}{\partial m^*} < 0$.

(adapted from Goyal, Moraga, and van der Leij (2007))

5 Discussions

5.1 Distances

Recently, a number of studies have provided analytical results in an attempt to shed light on the small world phenomenon. Influential work by Newman, Strogatz, and Watts (2001) has established standard methods to estimate diameters in large, complex networks. In this approach, the model is given by starting from a degree distribution P and choosing a graph uniformly at random from all graphs with this distribution of degrees. Within this framework, starting from a node chosen at random, the number of nodes one link away is given by the average degree $\overline{k}$ of P. Each such node in turn has degree distributed according to $P'(k) = \frac{kP(k)}{\overline{k}}$. The average number of nodes two links away is therefore $\overline{k}.\overline{k}'$, where $\overline{k}'$ indicates average degree under P'. After d such steps one covers $\sum_{l=1}^{d} \overline{k}\left(\overline{k}'\right)^{l}$ nodes, giving an approximate average diameter $\overline{d}$ solution to

$$\sum_{l=0}^{\overline{d}} \overline{k}\left(\overline{k}'\right)^{l} = n \tag{12}$$

The heuristic argument outlined above provides a useful benchmark in many cases, however ignoring the actual structure of the network may prompt largely misleading conclusions as shown by the work of Bollobas and Riordan (2004). Indeed, the authors are able to show that, in the preferential attachment model (PA), whereas heuristics correctly predict a diameter $O(\frac{\ln n}{\ln(\ln n)})$ when the number of links formed by entering nodes is greater than or equal to 2, predictions fail to be revised to the correct value of $O(\ln n)$ when a single link is formed on entry.

Our model exhibits some important similarities to PA. Bollobas and Riordan's (2004) contribution therefore rings a first alarm concerning the use of standard methods to estimate diameters in our case. However, simple inspection alone of the heuristic argument given above provides convincing case against its use for our purpose. Indeed, the neighbourhood expansion method underlying equation (12) implicitly assumes tree-like structure of the network. Clearly, such an approximation cannot be supported in our model considering the amount of neighbourhood overlap exhibited, even as $n \to \infty$. The caveat is that in our model many links are in fact redundant as they do not help to decrease the distance between nodes much. A detailed analysis of distances arising in our model therefore represents considerable challenge.

Short distances arise in PA (when the number of links formed by entering nodes is ≥ 2) from the combined effects of proportionality in degree-growth with randomness in linking. As indicated previously, proportionality in degree-growth leads to the existence of some very highly connected nodes. These nodes then act as hubs of information for newly entering nodes which (randomly) create bridges between them. Fat-tails do result in our model, however the number of distinct neighbourhoods connected by entering nodes is stochastic and has mean $m_r\varepsilon + m_r(1-\varepsilon)[1-(1-\lambda\frac{\varepsilon}{4(1-\varepsilon)})^m] \sim m$. Therefore, although the results from Bollobas and Riordan (2004) indicated in the previous paragraph do not transpose immediately here, it seems natural to conjecture that similar results hold in our model too provided $m \geq 2$.

Finally, notice that our model structurally imposes a lower bound on the distance separating two nodes due to the fact that on any path between them each 'step' size in S is bounded above by ε. Therefore, although the model guarantees existence of some very highly connected nodes, these hubs together form a chain in which each member connects only others close by in S. For some parameter values this effect may be large and affect distances importantly. At this stage, we simply point to the fact that minor amendments can be found that resolve this weakness. Introducing a few fully tolerant nodes would be one way of bringing together the different hubs. Alternatively, as discussed below, allowing for higher dimensional S can also reduce distances under appropriate assumptions.

5.2 Geographic vs non-geographic

In the most straightforward interpretation the space S may be identified with the familiar physical geographic space, in which communication costs effectively serve to inflate or deflate distances. A fall in ε for example corresponds to an increase in communication costs, disabling agents to maintain ties with others far away.

Our model however may also serve as a useful yardstick in view of empirical research concerning individuals' tendency to associate with others similar to them, a phenomenon usually coined as 'homophily'[12]. Our endeavour in this interpretation may be viewed as an

[12] See Cook, Smith-Lovin, and McPherson (2001) for a well documented survey on homophily. See also Currarini, Jackson,

attempt to extract some of the information conveyed by linking patterns regarding nodes' preferences. In this case S may be used to represent the underlying social characteristic space, with higher ε representing more 'open' societies.

Quite generally, it is insightful to view the model developed in the present paper as an embodiment of the JRV framework in a geographical context. Following the notation of Jackson and Rogers (2007), a non-geographic version of the model presented here is obtained by taking a probability p_r of link formation following random meetings and some independent probability p_n of link formation following network-based meetings[13]. A few remarks may be valuable. First, using (2) and (4) to substitute in (8) one obtains

$$F(d) = 1 - (\frac{rm}{rm+d})^{1+r} \tag{13}$$

It can be checked that, in this more general form, expression (13) for the cdf of the in-degree distribution applies to the non-geographic model too. This indicates that (13) essentially captures the dynamics of the JRV framework and is quite independent of the (non-)geographic aspect of the model.

Second, the implications of a geographic model are strongest regarding clustering. Loosely speaking, in the geographic model clustering is determined by the dimensionality of the underlying space[14]. The non-geographic model offers more flexibility. An interesting point of comparison consists in setting the network-based linking probability equal to the highest (conditional) value attained in the geographic model, i.e. to set $p_n = \max\{\frac{3}{4}, \frac{\varepsilon}{4(1-\varepsilon)}\} = \frac{3}{4}$. On the one hand it is quite compelling to choose p_r so as to set the average density of links m equal in both models[15]. There are two important drawbacks however to choosing p_r in this way. One, we indicated in Section 3.2 that clustering tends to naturally drop as network density rises due to the fact that network-based meetings occur in the first (random) stage only. And two, we wish to focus on first order clustering, which further requires binding m. For these reasons we contend that a more meaningful evaluation results from taking limits in which, respectively, $\varepsilon \to 0$ and $p_r \to 0$. Proposition 4 highlights the sense in which geographic settings tend to generate networks which are more clustered than their non-geographic counterpart. The intuition behind Proposition 4 is straightforward. For given numbers of links of each

and Pin (2007) for a recent theoretical investigation of the phenomenon.

[13] The non-geographic version of our model differs slightly from Jackson and Rogers (2007) since we do not restrict new entrants in the number of network-based meetings they make.

[14] See Section 5.3.1.

[15] Notice that this would imply choosing $p_r < \varepsilon$.

kind, network-based links tend to be concentrated in the neighbourhood of matching parent nodes in the geographic model whereas the same links are evenly spread between matching and non-matching parent nodes' neighbourhoods in the non-geographic model.

Proposition 4 *Let* $p_n = \frac{3}{4}$*. Then* $\lim_{p_r \to 0} C^{ng} < \lim_{\varepsilon \to 0} C^g$*.*

To complete, we should add that provided $m_r\lambda$ is small enough one may obtain $C^{ng} > C^g$ for some range of p_n above $\frac{3}{4}$. Such a result should not be too surprising since the non-geographic model is naturally less binding than its geographic counterpart.

5.3 Extensions

5.3.1 Higher dimensions

To keep things simple, all results in the present paper have been derived for the case in which S was identified with the Euclidian space[16] of dimension one. The model naturally extends into higher dimensions as well, subject to the following adjustments. In dimension $s > 1$, $\Omega_i(x_i)$ is chosen to be the s-dimensional cube of sidelength ε[17]. This is a technical requirement, and our results hold if we instead choose to work with balls of radius ε (though in that case they become approximations)[18]. The analysis carried out in Sections 2 and 3 is easily generalized to give Proposition 5. In particular, and in view of the discussion given in Section 5.2, Proposition 5.(ii) corroborates our early assertion that in the geographic model clustering is determined by the dimensionality of the underlying space.

Proposition 5 *In the general case, with* $\dim(S) = s < \infty$ *and with* $\Omega(x_i)$ *denoting the* s*-dimensional cube of sidelength* ε *centred at* x_i*, as* $t \to \infty$ *and in the mean-field approximation*

(i) $F_t(d) = 1 - (\frac{\lambda^{-1}}{\lambda^{-1}+d})^{1/\lambda m_r \varepsilon^s}$

(ii) $\sup_\varepsilon C = \lim_{\varepsilon \to 0} C = \left(\frac{3}{4}\right)^s \lambda$

A few remarks are useful to uncover the results of Proposition 5. First, notice that with $\Omega(x_i)$ so defined nodes are constrained to match along all s dimensions. Since the

[16] More precisely a close substitute to the Euclidian space of dimension one, namely the torus of unit length.

[17] Note that, just as in dimension one we chose to work with the torus of unit length, in dimension $s > 1$ we take S to be the $(s+1)$-dimensional sphere normalized with unit surface.

[18] In a nutshell, this condition allows us to generalize the point made by Proposition 1.(ii) whereby we showed that the extent of interval overlap was independent of ε (assuming $\varepsilon << 1$).

probability of matching along any one dimension is ε and we have assumed independence across dimensions, the overall matching probability is given by ε^s. As the dimensionality increases, this naturally generates networks which tend to exhibit fewer links. Second, in higher dimensions nodes typically find themselves close to a subset of neighbours along one dimension while they are close to other neighbours along a different dimension. The possibilities for any given node's neighbours to find themselves at odds therefore increase with the dimensionality. Figure (C) illustrates this effect. Whereas in the unidimensional case $x_i \simeq x_j$ implies $|\Omega_i \cap \Omega_j| \simeq |\Omega_i|$, in two dimensions we draw an example for which $|\Omega_i \cap \Omega_j| \simeq \frac{|\Omega_i|}{2}$.

To complete, let us note that alternative versions of Proposition 5 may just as easily be obtained for which $\Omega(x_i)$ is chosen differently. An interesting example can be given in two dimensions, where one only requires nodes to match along a single dimension. In that case short (geodesic) distances between widely dissimilar nodes nodes are rendered possible by the intervention of intermediate nodes matching both of them (along different dimensions).

5.3.2 Heterogeneity

Heterogeneity of agents naturally arises in the model as a consequence of timing. However, we maintained throughout the existence of a single ε for society as a whole. Although this is certainly a simplifying assumption, my view is that it is easily supported as a first approximation by the fact that technological constraints in communication are commonly shared, and that different societies do exhibit tendencies towards lesser or greater tolerance levels. Nevertheless, there are good reasons for which one may want to relax this assumption. First, in the context of communication costs we should expect heterogeneity in agents' budget constraints to be reflected in varying ε values. Second, in social dimensions contexts some agents do tend to be more tolerant than others and this may in turn have consequences for society as a whole. For example, highly tolerant agents can create bridges between parts of the network segregated by social characteristics. Studying the impact of having heterogeneity in ε may therefore be an interesting path for research.

6 Concluding Remarks

The present paper suggests a possible framework to analyze the impact of changes to the economic and social environment on the topology of networks formed. Economic (costs) and social (norms) constraints bind individuals in their ability to create ties with others. When global phenomena affect these constraints, the overall shapes of resulting networks naturally alter. I have tried to shed light on this relationship.

One weakness of the analysis in the present paper relates to the fact that the optimizing behaviour of nodes has by and large been ignored. Clearly, an interesting path for future research would be to investigate the consequences of giving nodes some freedom regarding their linking strategies. This is also a necessary step to truly understand the incentives mechanisms underlying the formation of social networks.

It is hoped finally that the results of the present paper will prove useful to those doing empirical research on networks.

7 Appendix

Proof of Proposition 1

(i) Obvious.

(ii) To begin, fix $X_i = x_i$ and $X_j = x_j$, with $x_j \in \Omega(x_i)$. Assume moreover that $X_k \in \Omega(x_j)$. Under such conditions, the probability that X_k lies within $\Omega(x_i)$ is given by

$$\frac{|\Omega(x_i) \cap \Omega(x_j)|}{|\Omega(x_j)|}$$

where $|.|$ denotes the length of an interval. This is given equivalently by

$$\frac{\varepsilon - |x_j - x_i|}{\varepsilon}$$

By varying x_j along $\Omega(x_i)$ and integrating out, we obtain

$$\Pr(X_k \in \Omega(x_i) \mid (X_j \in \Omega(x_i)) \wedge (X_k \in \Omega(X_j))) = \int \frac{\varepsilon - |x_j - x_i|}{\varepsilon} dF$$

where F denotes the distribution of $|X_j - x_i|$ conditional on $X_j \in \Omega(x_i)$. Under the hypothesis of the model F is the distribution of a uniform random variable with support on $[0, \frac{\varepsilon}{2}]$. Substituting in the above expression yields

$$\Pr(X_k \in \Omega(x_i) \mid (X_j \in \Omega(x_i)) \wedge (X_k \in \Omega(X_j))) = \frac{2}{\varepsilon} \int_0^{\frac{\varepsilon}{2}} \frac{\varepsilon - u}{\varepsilon} du = \frac{3}{4}$$

Since this result holds for arbitrary x_i, the proof is concluded.

(iii) Notice that the probability of matching with a node picked at random is the same as that of matching with any neighbour of a node picked at random. Therefore

$$\Pr(X_k \in \Omega_i | X_k \in \Omega_j) = \varepsilon$$

Conditioning on X_j and using (ii) we then get

$$\begin{aligned}
\varepsilon &= \Pr(X_k \in \Omega_i | X_k \in \Omega_j) \\
&= \Pr(X_k \in \Omega_i \mid (X_j \in \Omega_i) \wedge (X_k \in \Omega_j)) \Pr(X_j \in \Omega_i) + \Pr(X_k \in \Omega_i \mid (X_j \in \Omega_i^c) \wedge (X_k \in \Omega_j)) \Pr(X_j \in \Omega_i^c) \\
&= \frac{3}{4}\varepsilon + (1-\varepsilon) \Pr(X_k \in \Omega_i \mid (X_j \in \Omega_i^c) \wedge (X_k \in \Omega_j))
\end{aligned}$$

from which

$$\Pr(X_k \in \Omega_i \mid (X_j \in \Omega_i^c) \wedge (X_k \in \Omega_j)) = \frac{\varepsilon}{4\,(1-\varepsilon)}$$

as indicated in the statement of the Proposition.

Proof of Proposition 2[19]

Let m_t denote the number of (outgoing) links formed by node t. Adding contributions from the random and network-based processes gives

$$\begin{aligned}
E[m_t] &= m_r\varepsilon(1 + \frac{3}{4}\lambda E[m_s|s<t]) + m_r\,(1-\varepsilon)\,(\frac{\varepsilon}{4\,(1-\varepsilon)}\lambda E[m_s|s<t]) \\
&= m_r\varepsilon(1 + \lambda E[m_s|s<t])
\end{aligned}$$

It is easy to see, by induction, that $E[m_s|s<t] \le E[m_t]$, and so

$$E[m_t] \le m_r\varepsilon(1 + \lambda E[m_t])$$

Successive substitution of the previous expression into itself then shows that for $\lambda m_r\varepsilon < 1$, $E[m_t] \le \frac{m_r\varepsilon}{1-\lambda m_r\varepsilon}$. $E[m_s]$ is therefore increasing and bounded, and so converges as $s \to \infty$, as claimed in the statement of the proposition.

Proof of Theorem 1

The equation of motion for node i is given by

$$\frac{dd_i(t)}{dt} = \frac{m_r\varepsilon}{t} + (\frac{\lambda m_r\varepsilon}{t})d_i(t)\ ,\ \ t \ge i$$

[19] I am grateful to Marco Van der Leij for suggesting this simpler version of the proof.

with initial condition $d_i(i) = 0$. The solution to this standard ODE yields

$$d_i(t) = \lambda^{-1}\left(\frac{t}{i}\right)^{\lambda m_r \varepsilon} - \lambda^{-1}\ , \quad t \geq i$$

Let $i(d,t)$ denote the (unique) node with degree d at time t. Substituting in the previous expression this is given by

$$i(d,t) = t\left(\frac{\lambda^{-1}}{\lambda^{-1} + d}\right)^{\frac{1}{\lambda m_r \varepsilon}}$$

In our deterministic framework, $1 - F_t(d)$ corresponds to the fraction of nodes older than $i(d,t)$, i.e.

$$1 - F_t(d) = \frac{i(d,t)}{t} = \left(\frac{\lambda^{-1}}{\lambda^{-1} + d}\right)^{\frac{1}{\lambda m_r \varepsilon}}$$

as indicated in the Theorem.

Proof of Corollary 1

(i) It is easy to see that $1 - \left(\frac{\lambda^{-1}}{\lambda^{-1}+d}\right)^{\frac{1}{\lambda m_r \varepsilon}}$ is decreasing m_r and ε.

Showing that $1 - \left(\frac{\lambda^{-1}}{\lambda^{-1}+d}\right)^{\frac{1}{\lambda m_r \varepsilon}}$ is decreasing in λ, is the same as showing that $\left(\frac{x}{x+d}\right)^x$ decreases in x for $x > 0$. Since $\frac{d}{dx}(\left(\frac{x}{x+d}\right)^x) = \left(\frac{x}{x+d}\right)^x [\ln(\frac{x}{x+d}) + \frac{d}{x+d}]$, it is enough to show that $h(x) = \ln(\frac{x}{x+d}) + \frac{d}{x+d} < 0$, $\forall x > 0$. This in turn follows from the observation that $\frac{d^2 h(x)}{dx^2} = -\frac{(3x+d)d^2}{(x+d)^3 x^2} < 0$, $\forall x > 0$, while

$$\lim_{x \to 0} h(x) = -\infty$$

and

$$\lim_{x \to \infty} h(x) = \lim_{x \to \infty} [-\frac{d^2}{x(x+d)}] = 0^-$$

(ii) Substitute $\lambda^{-1} = mr$, and $\frac{1}{\lambda m_r \varepsilon} = 1 + r$. It is then easily verified that $1 - \left(\frac{mr}{mr+d}\right)^{1+r}$ is decreasing in m.

(iii) Substituting as in (ii), the result follows by Theorem 6 in Jackson and Rogers (2007).

Proof of Theorem 2

Let $\Lambda_i(g) = \sum\limits_{j \neq i; k \neq j} g_{ij} g_{ik}$ denote the number of pairs of outgoing links (bold links in Figure (B)) existing for node i in network g, $\Gamma_i(g) = \sum\limits_{j \neq i; k \neq j} g_{ij} g_{jk}$ denote the number of transitive pairs of outgoing links (bold links in Figure (A)) for node i in network g, and $\Delta_i(g) = \sum_{j \neq i; k \neq j} g_{ij} g_{jk} g_{ik}$ denote the number of triplets realized for node i in network g.

With this notation, (9) and (10) can be rewritten as

$$C(g) = \frac{\sum_i \Delta_i(g)}{\sum_i \Gamma_i(g)}$$

and

$$\Xi(g) = C(g) \frac{\sum_i \Gamma_i(g)}{\sum_i \Lambda_i(g)}$$

Note that, with respect to i, each of $\Lambda_i(g)$, $\Gamma_i(g)$, and $\Delta_i(g)$ involve outgoing links only. Since all nodes are treated symmetrically in the formation process regarding outgoing links, we can divide both nominator and denominator by t and approximate in the law of large numbers as $t \to \infty$ to get

$$C = \frac{E[\Delta]}{E[\Gamma]} = \frac{E[\Delta]}{m^2}$$

and

$$\Xi = C \frac{m^2}{m(m-1)/2}$$

The next step in the proof consists in calculating $E[\Delta]$, the expected number of triplets realized per node in the process. We consider in turn the contributions from the 3 cases highlighted in the text (see Figure(B))

1. Both j and k were met randomly. In this case, with t large, the probability of jk existing becomes arbitrarily small.

2. j was met randomly while k was met through the network. If k was met through $l \neq j$ then given the information set, the probability of jk is at most that of a highly connected node with some other random node. This is higher than in the previous case but still tending towards zero under weak conditions. However, when k was met through j then we have found such a triplet by definition. Each random link engenders

an average $3/4\lambda m$ matches in its neighbourhood. Therefore, the expected contribution of this scenario is $m_r\varepsilon(3/4.\lambda m)$.

3. Both j and k where met through the network. Once again, we need only consider the situation in which j and k belong to the same parent's neighbourhood. In the model, network-based links are primarily concentrated in the neighbourhoods of matching parent nodes. Each of these provides an expected $3/4\lambda m$ network-based links whereas non-matching parent nodes only contributes $\frac{\varepsilon}{4(1-\varepsilon)}\lambda m$. Three ranges of parameters must be considered. (i) When $3/4.m < 1$ second order clustering is altogether absent. (ii) When $3/4.\lambda m > 1 > \frac{\varepsilon}{4(1-\varepsilon)}\lambda m$ second order clustering obtains for matching parent nodes. The situation in Figure (B) takes place an average $\frac{3/4.\lambda m(3/4.\lambda m-1)}{2}$ times per parent node. In each of these events link jk exists with probability Ξ (see (10)). The total expected contribution to node i's triplets is therefore $m_r\varepsilon\Xi\left(\frac{3/4.\lambda m(3/4.\lambda m-1)}{2}\right)$.(iii) Finally when $\frac{\varepsilon}{4(1-\varepsilon)}\lambda m > 1$ we must add to the previous case the contribution from second order clustering arising in non-matching parent nodes' neighbourhoods. The total expected contribution to node i's triplets becomes $\Xi\left[m_r\varepsilon\frac{3/4.\lambda m(3/4.\lambda m-1)}{2} + m_r(1-\varepsilon)\frac{\frac{\varepsilon\lambda m}{4(1-\varepsilon)}\left(\frac{\varepsilon\lambda m}{4(1-\varepsilon)}-1\right)}{2}\right]$.

For $\frac{\varepsilon}{4(1-\varepsilon)}\lambda m > 1$, adding the contributions of cases 2 and 3 above, we find

$$C = \frac{m_r\varepsilon(3/4.\lambda m) + \Xi\left[m_r\varepsilon\frac{3/4.\lambda m(3/4.\lambda m-1)}{2} + m_r(1-\varepsilon)\frac{\frac{\varepsilon\lambda m}{4(1-\varepsilon)}\left(\frac{\varepsilon\lambda m}{4(1-\varepsilon)}-1\right)}{2}\right]}{m^2}$$

and, replacing $\Xi = C\frac{m^2}{m(m-1)/2}$

$$C = \frac{3\lambda m_r\varepsilon}{4m} + C\frac{m_r\varepsilon\left[3/4.\lambda m(3/4.\lambda m-1)\right] + m_r(1-\varepsilon)\left[\frac{\varepsilon\lambda m}{4(1-\varepsilon)}\left(\frac{\varepsilon\lambda m}{4(1-\varepsilon)}-1\right)\right]}{m(m-1)}$$

simple algebra then yields

$$C = \frac{12\lambda r\,(m-1)}{16\,(1+r)\,(m-1) - \lambda m\left(9 + \frac{\varepsilon}{1-\varepsilon} - 16r\right)}$$

Other cases are solved in the same way. For $\frac{\varepsilon}{4(1-\varepsilon)}\lambda m < 1$ and $\frac{3}{4}\lambda m > 1$ we obtain

$$C = \frac{12\lambda r\,(m-1)}{16\,(1+r)\,(m-1) - \lambda m\,(9-12r)}$$

while for $\frac{3}{4}\lambda m < 1$ we have

$$C = \frac{3}{4m(1+r)}$$

Part (i), (ii), and (iii) of Theorem 2 are easily verified by substituting out for m and r using (2) and (4) to express C in terms of the parameters of the model.

Proof of proposition 4

In the non-geographic model, we have $m = p_r m_r + m_r m \lambda p_n$, and $m = \frac{p_r m_r}{1 - m_r \lambda p_n}$. In particular $m \to_{p_r \to 0} 0$, and the average number of network-based links formed also tends to zero as $p_r \to 0$ (recall from Section 3.2 that no clustering ever arises in the neighbourhoods of these nodes according to the model). In the limit, we can therefore safely ignore second-order clustering. Using definitions from Section 3.2 it then follows that $\lim_{p_r \to 0} C^{ng} = \frac{m_r p_r \lambda p_n}{m} = \lambda p_n (1 - m_r \lambda p_n) < \frac{3}{4}\lambda = \lim_{\varepsilon \to 0} C^g$, from Theorem 2.

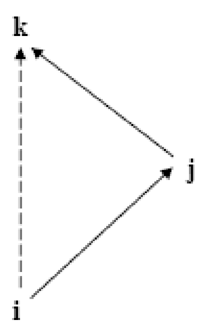

Figure A

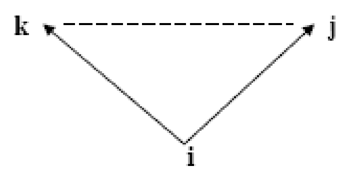

Figure B

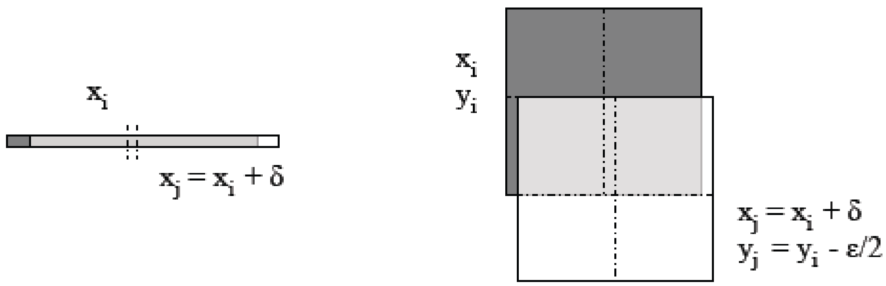

Figure C

References

[1] Bala, V. and Goyal, S. (2000). "A noncooperative model of network formation", Econometrica, vol. 68(5), pp. 1181–229.

[2] Barabasi, A., R. Albert, and H. Jeong (1999), "Mean-field theory for scale-free random networks," Physica A 272: 173-187.

[3] Bollobas, B. , and Riordan O. (2004), "The diameter of a scale-free random graph", Combinatorica, 24 (1), 5–34.

[4] Conley T. G., Udry C. R. (2001), "Social Learning Through Networks: The Adoption of New Agricultural Technologies in Ghana", Journal of Agricultural Economics, Vol 83 Issue 3, 668-673.

[5] Currarini S., Jackson M., and Pin P. (2007), "An Economic Model of Friendship: Homophily, Minorities, and Segregation", forthcoming, Econometrica.

[6] Galeotti A., Goyal S., and Kamphorst J. (2006), "Network formation with heterogeneous players", Games and Economic Behavior, 54, 353–372.

[7] Gilles R., Johnson C. (2000)."Spatial social networks," Review of Economic Design, Springer, vol. 5(3), pages 273-299.

[8] Goyal, S. (2007), "Connections: an introduction to the economics of networks". Princeton University Press.

[9] Goyal, S., Moraga, J-L, and van der Leij, M. (2007), "Why is the economics world becoming smaller?", unpublished notes.

[10] Granovetter, M. (1973) "The Strength of Weak Ties," American Journal of Sociology, 78, 1360-1380.

[11] Jackson, Matthew O. and Asher Wolinsky (1996) "A strategic model of social and economic networks," Journal of Economic Theory 71, 44-74.

[12] Jackson, M. O. and Rogers B. (2007), "Meeting strangers and friends of friends: how random are social networks", American Economic Review, Vol. 97, no. 3, 890-915.

[13] Katz E., and Lazarsfeld P.F. (1955), "Personal Influence: The part played by people in the flow of mass communication", Glencoe, IL: Free Press.

[14] McPherson, M., L. Smith-Lovin, and J.M. Cook (2001), "Birds of a Feather: Homophily in Social Networks" Annu. Rev. Sociol., 27, 415-44

[15] Newman M. E. J., Strogatz S. H., and Watts D. J. (2001), "Random graphs with arbitrary degree distribution and their applications", Physical Review E, 64, 026118.

[16] Penrose M. (2003), "Random Geometric Graphs", Oxford Studies in Probabilty, 5, Oxford University Press.

[17] Vazquez A. (2003), "Growing Network with Local Rules: Preferential Attachment, Clustering Hierarchy, and Degree Correlations." Physical Review E, 67(5), 056104.

[18] Vega-Redondo (2007), "Complex Social Networks", Econometric Society Monographs, Cambridge University Press.

[19] Watts, D. and S.H. Strogatz (1998), "Collective dynamics of 'small-world' networks", Nature, 393: 440-442.