Knowledge Portfolios and the Organization of Innovation Networks*

Robin Cowan[1,2], Nicolas Jonard[3]

[1]BETA, Université Louis Pasteur,
61 Avenue de la Forêt Noire, 67000 Strasbourg, France.
[2]UNU-MERIT, University of Maastricht,
P.O. Box 616, 6200 MD Maastricht, The Netherlands.
[3]Université du Luxembourg,
162a Avenue de la Faiencerie, L-1511 Luxembourg.

September 14, 2007

Abstract

Properties of strategic alliance networks such as small worlds, skewed link distributions and patterns of repeated tie occurrences are often explained in terms of social capital theories. A simple model shows that merely assuming that firms must have a certain degree of commonality in their knowledge to have a successful alliance, is enough to produce the above features, without recourse to social capital at all.

1 Introduction

Recent empirical work on knowledge networks has observed robust regularities. Whether R&D alliances, patent citations, patent co-invention or scientific co-authorship, we observe network architectures that tend to share three properties. First they are sparse: the number of observed partnerships is much smaller than the number that could form in principle. Second, innovation networks are asymmetric: distribution of the number of ties held by a firm (also known as the degree distribution) tends to be skewed, with most nodes having few ties, and a few nodes have very many. Finally, they tend to display local clusters, that is, dense local sub-groups of interconnected agents, connected by (few) clique-spanning ties.[1]

*The detailed suggestions and comments of three anonymous referees and Ron Adner have been extremely helpful in improving the paper. We also thank Moritz Mueller for his help with the biotech data.

[1]On the general properties of interfirm networks, see Garcia-Pont and Nohria (2002); Walker et al. (1997); Kogut et al. (1992) or Powell et al. (2005); on scientific co-authorship see Newman (2001); on patent citation in US biotech see Johnson and Mareva (2006); on the life sciences and ICT industry networks see Riccaboni and Pammolli (2002); on bank syndicates, see Baum et al. (2003).

All three characteristics are compatible with the notion of small world networks. Small worlds are essentially characterized by the conjunction of higher amounts of clustering (roughly the ratio of existing to possible triangles in the network) and lower pairwise distances than those a random graph of similar degree distribution would display.[2]

In this paper we present a simple model of alliance formation in which the properties of knowledge and the nature of joint innovation generate the network properties just described. The structural properties of the network are created without recourse to any social capital effects.

1.1 Innovation networks

In the literature, concern for detailed explanation has focussed on the small world properties of these networks — the combination of local clustering with sparse but efficient inter-cluster connections. Sparseness is understood to be associated with the cost of forming a link — typically links have real costs of creation and maintenance (though citation links have no maintenance costs they will still have a formation cost, as any paper or patent can cite only a very finite number of other papers or patents relative to the population), and this means that relatively few will be created. Skewness can easily arise from other properties that are known to have a skewed distribution: firm size; patent or paper quality; author quality; or preferential attachment arising from a self-enforcing mechanism (it is important to cite canonical papers in any field, canonical papers are those cited a lot, and so on, leading to a highly skewed distribution).

For the clustering aspect of small worlds, there are two types of explanation: social capital or control; and critical mass.

Forming an alliance with another firm is a source of risk, but finding partners that are "well-embedded" can reduce it. Relational embeddedness will introduce inertia into a network, but does not induce clustering. What does introduce clustering is the value of structural embeddedness. Firms who share a common partner have a source of information about each other's trustworthiness, capabilities or competences, objectives and so on (see Kogut et al., 1992, or Ahuja, 2000 for example), which is clearly valuable both in evaluating ex ante, and managing an alliance once it has formed. Structural embedding also enforces reputation effects and mitigates the effects of power asymmetries, both inducing "good behaviour" in partnerships.[3] All of these are in essence arguments that local clusters create valuable social capital (Coleman, 1988), which is beneficial for innovation. Indeed, Chung et al. (2000), Gulati (1995), Gulati and Gargiulo (1999), examining very different industries, find empirically that the existence of indirect ties between two firms either increases the probability that they will form an alliance, or increases the performance of that alliance. Social capital is valuable, and while taking advantage of that, firms (possibly

[2]The commonly-used Watts and Strogatz (1998) algorithm for generating small worlds does not typically create networks with skewed link distributions.

[3]For various discussions of these issues see, for example, Uzzi (1997), Walker et al. (1997) or Dyer and Nobeoka (2000). Baum et al. (2003) summarize this literature nicely, and this paragraph is drawn from their summary. Rowley et al. (2000) argue that relational and structural embeddedness have joint effects which are contingent on the nature of innovation in the industry (exploration versus exploitation). Cowan et al. (2007) present a model in which firms' relative valuation of relational and social embeddedness affects the types of innovation networks that form.

inadvertently) form networks that exhibit strong local clustering.

The second type of explanation for local clustering has to do with the importance of tacit knowledge and the benefits of face-to-face interaction. Many studies document the geographic agglomeration of innovation, and argue that knowledge spillovers explain clustering, geographical clustering directly implying network clustering.[4] A locally dense structure permits agents to have concentrated interactions, to create common languages, problem definitions and problem-solving heuristics. In addition, the high frequency of face-to-face interactions with many agents who share a common approach facilitates rapid diffusion of tacit knowledge, encouraging localized innovation. This is often given as an explanation for the success of Silicon Valley (Saxenian, 1991).[5] Von Hippel (1998) shows empirically that trading of tacit knowledge, which can be central to innovation success, depends on the existence both of long-term stable relationships, and of a shared language and tradition. The strength of the critical mass arguments depends on the importance of face-to-face interactions, and so its validity demands geographical proximity. Our mechanism makes no such demands. What the clustering and critical mass arguments cannot explain is dense networks among geographically disperse agents, which is not a problem for the model we develop below.

The second property of small worlds, namely short path lengths, arises from clique-spanning ties. Firms may explicitly consider the entire network as a strategic variable, or they may simply create links to optimize knowledge flows. In either case, too much clustering implies too much redundancy in a firm's ego-network. Some of a firm's links could be eliminated with a minimal impact on the knowledge that firm can access. Thus a firm successfully optimizing knowledge flows will have structural holes in its ego-network. This is Burt's (1992) structural holes argument. The firm will use some of its links to connect to distant parts of its network, and in particular parts of the network to which its neighbours are not well-connected. If this is the case, then a firm's network connections permit it to access rapidly knowledge in different parts of the network. This insures the firm against technological surprises that take place at a distance, which can arise if a firm is too deeply embedded locally, and gives it access to information that is not available in its local cluster. Whether this is a deliberate "network strategy" or not, if knowledge access is important, firms that do this well will have a competitive advantage over those that do not, and so a network should evolve to one that contains many structural holes. Baum et al. (2003), for example, find that spanning cliques can be a deliberate strategy aimed at information control, either by strong firms consolidating their positions, or by peripheral firms attempting to move to a more central position in a network. Here, firms are explicitly considering network structures and how particular positions in the structure allow them to control different parts of the network.

In this paper we develop a model in which the network properties observed above — sparseness, skewness, and small worlds — are the consequence not of social capital or network-based arguments, but simply of the nature of joint innovation. While we do not deny the appeal of social capital or network-based arguments to explain the observed patterns of network formation and dynamics, typically the nature of joint innovation has not

[4] See for example Audretsch and Feldman (1996), Baptista (2000) and Prevezer and Swann (1996). Feldman (1999) provides a survey.

[5] This is an example of the creation of club-goods more generally, and a tightly inter-connected clique creates a club-like atmosphere, conducive to generating these goods (Capello, 1999).

been controlled for adequately in empirical work. This could very well bias the empirical results, tending to over-ascribe influence to social and network explanations (such as embeddedness) to alliance formation and network structure. Therefore, one contribution of our formalization is to examine carefully an alternative to trust, reputation, shared expectations and so on, as explanation for these patterns in the network data.

1.2 Knowledge and innovation alliances

Joint innovation implicitly involves at least two firms combining their knowledge stocks to create new knowledge. These knowledge stocks will be located at different points in the underlying knowledge space, and so the issue of the distance between them can be important. If firms are too close together, their knowledge overlaps too much and there is little point in sharing; if they are too far apart they have difficulty understanding each other, and so sharing is too difficult (Nooteboom, 2000 or Grant, 1996). In addition to being intuitively appealing, this relationship is largely supported by empirical evidence reported in Mowery et al. (1998, 1996), Ahuja and Katila (2001) and Schoenmakers and Duysters (2006).

Mowery et al. (1998) use patent citation data to compare the intersection between two firms' technology portfolios before and after a partnership. They find that joint venture partners display a higher degree of technological overlap compared with a similar sample of non-collaborators, but the effect is non-monotonic. There is an optimal distance in knowledge space, and the likelihood of forming an alliance falls as a pair of firms moves away from it. In addition, Mowery et al. find that after two firms have forged an alliance, the overlap in their knowledge increases, and their distance in knowledge space falls. If two firms together create new knowledge, both can possess it after it is created (ignoring intellectual property rights issues). Adding the same piece of knowledge to their respective initial knowledge profiles will bring them closer together in knowledge space. This is true whether or not there are knowledge spillovers (inadvertent or intended) from one firm to the other.[6] This, of course, changes the probability that they will forge a new alliance in the future: increasing it if they were relatively far apart; decreasing it if they were relatively close together.[7]

Ahuja and Katila (2001) find a similar inverted-U relationship between knowledge-base overlap and innovation performance, not in strategic alliances, but in acquisitions. Schoenmakers and Duysters (2006) find that the increase in the knowledge-base overlap of two alliance partners also has an inverted-U relationship with initial overlap (and that the increase is positive).[8]

[6]Mowery et al. (1996) have a slightly more nuanced view, and conclude, following Nakamura et al. (1996), that some alliances produce "complementary specialization" (p. 87). They suggest that some alliances "are vehicles for accessing rather than acquiring capabilities." In the present paper, we restrict attention to alliances aimed at creating capabilities.

[7]Mowery et al. measure technological overlap (the extent to which two firms' knowledge is similar), by patent cross- or common-citations. If firm i heavily cites the patents of firm j, and vice versa, this suggests that their technological competences must be similar. Similarly, if the two firms draw from the same external pool of patents, they are using and absorbing the same knowledge, which provides another source of similarity or overlap.

[8]Both Ahuja and Katila, and Schoenmakers and Duysters define overlap in the following way. A firm's knowledge base is the set of patents it owns plus the set of patents its patents cite. The overlap of two

The above results are compelling, and in this paper we use them as the heart of a stylized model of alliance formation and joint innovation. The model's assumptions mimic the above results: in knowledge space, there is an optimal distance for partners to an alliance, and the effect of an alliance is to increase the partners' overlap. Given these assumptions on innovation, we find that: networks are often sparse; degree distributions are typically skewed; pairwise distances are consistently short; and excess clustering is always present. These properties combined point to the presence of small world networks. In addition, we observe that the incidence of repeated ties tends to weaken as time passes, typically after a phase of growth. These results are obtained without any assumption about the nature of knowledge (tacitness, imperfect absorption, etc), and without any assumption on the value of third parties in preventing opportunistic behaviour, sharing efficiently routines and favouring the emergence of trust, etc, in short abstracting entirely from any effects of social capital.

2 The model

Each firm in a finite population $\mathcal{N} = \{1, \dots, n\}$ is trying to acquire new knowledge. Firms pursue knowledge for its own sake, regardless of any other (e.g. market) motive. Knowledge is modelled as a finite set of discrete elements (which we refer to as ideas or facts), denoted $\mathcal{W} = \{1, \dots, w\}$. Firm i's knowledge portfolio is the list of facts it knows, $\theta_i \subseteq \mathcal{W}$ which evolves over time through innovation.

Firms can innovate either alone or as part of bilateral alliances. Our concern is with innovation arising from alliances, so we make no attempt to model explicitly the possibility of autarchic innovation: we simply adopt the very reduced form assumption that if a firm makes no alliances, it innovates on its own, with a fixed probability. But a firm can conduct projects that are joint with other firms. In an alliance, the probability of success is single-peaked in the overlap of the two firms, where by overlap we mean the size of the intersection of the partners' knowledge sets. Given the cost of forming an alliance, only alliances with overlap not "too far" from the optimal one will be profitable in expected terms. A successful firm is one that discovers, alone or in a partnership, an idea new to it, and this idea becomes part of the portfolio(s) of the firm(s). Each period firms form all profitable alliances and attempt to innovate. At the end of the period all alliances dissolve and the process repeats, with updated knowledge portfolios for the successful firms. Thus we have an evolving network of alliances, where the architecture of the network is determined entirely by the nature of innovation.

We have made the very simple assumption that pair formation is driven entirely by complementarities in partners' knowledge stocks. From this simple assumption on dyad creation we are able to derive the properties of the larger network of all alliances. Our interest is in how these properties respond to changes in the optimal overlap relative to the size of the underlying knowledge space, and how strictly the optimal overlap must be met.

firms' knowledge bases is the intersection of their knowledge bases, normalized by the union of them.

2.1 Innovation and equilibrium

In this section we show how the nature of joint innovation determines a unique equilibrium.

2.1.1 Innovation

Each firm initially holds each idea with probability $\Pr\{z \in \theta_i\} = 1/2$ independently for all $z \in \mathcal{W}$ and all $i \in \mathcal{N}$. An innovation, whether joint or individual, is the discovery of an idea of value equal to 1.[9]

Define $\theta_{ij} = \theta_i \cap \theta_j$ as the intersection of the knowledge portfolios of i and j, and define $y_{ij} = \#\theta_{ij}$ to be the size of the overlap. If the partnership ij forms, the two partners jointly innovate with probability $f(y_{ij})$, independently of the other alliances formed by i and j.[10] Specifically it is assumed that there is a cost of $c \geq 0$ incurred by each firm involved in a given link. To capture the inverted-U relationship between knowledge overlap and innovation success we assumedthat $f(y)$ is positive, symmetric, single-peaked at δ with $c < f(\delta) \ll 1$, increasing (decreasing) monotonically on the left (right) of δ. If i partners with j and they innovate successfully, they discover $z \notin \theta_i \cup \theta_j$, an idea new to both of them. They both receive the innovation, and thus their portfolios become $\theta_i + \{z\}$ and $\theta_j + \{z\}$.

A partnership at optimal distance has a probability of success strictly larger than the cost of the partnership. Consequently, partnerships "not too far" from the optimal distance will also have positive expected value, and thus will form.

2.1.2 Equilibrium

An industry network is a set g of links between pairs of firms in $\mathcal{N}$. The neighbourhood of firm i consists of the agents to whom i is directly connected, denoted $N_i^g = \{j \neq i : ij \in g\}$. The size of the neighbourhood is the number of ties held by a firm, or its degree, is denoted $n_i^g = \#N_i^g$.

What is the total profit of a firm? For each link ij it is involved in, i has a probability $f(y_{ij})$ of innovating. Unlike Jackson and Wolinsky (1996)'s communication model, in the

[9] Implicitly there is the assumption that a fact new to a firm and a fact new to the world have the same value. It could be argued that in a second stage in which firms use the knowledge they have discovered to create competitive advantage, facts "new to the world" are more valuable than facts "new to me", but here we only focus on the immediate production of knowledge. A second implicit assumption is that firms are indifferent about which facts they discover. To include the more realistic idea that firms get more value from one fact than another, would demand both a detailed specification of how firms turn facts into profits and how facts interact in that process, and a detailed specification of how firms turn existing knowledge into new knowledge. These are of course possible in principle, and material for an extension of the model, but the complication they would add at this point would distract significantly from the main message.

[10] There is a second order effect that we ignore. If i has a partnership with j, ij can innovate in locations $L = \mathcal{W} - \theta_i \cup \theta_j$. If i now considers k as a potential partner, any innovations that ik might make in L would be duplicates, and therefore of less value than innovations that take place outside L. Thus the evaluation of k as a partner should involve this second order effect, and in general, the evaluation of a portfolio of partners should include these interactions. We ignore this here and in what follows, since in the numerical experiment below success probabilities are small enough that these second order effects will have very little effect on decisions. In the experiment below, a firm innovates on average once per period, thus the risk of duplicate innovations is very remote. Including this effect would on average lower the expected value of an alliance, and so decrease the degree of the network.

present approach links do not transmit spillovers from innovations created elsewhere, nor do they perform any other task (control, information brokerage, etc.). Recalling that the value of an innovation is 1, firm i's expected one-period profit from networking activities is thus simply written

$$\pi_i^g = \sum_{j \in N_i^g} f(y_{ij}) - cn_i^g. \tag{1}$$

All firms in the industry face a similar problem when evaluating the partnerships they are engaged in, and for a partnership to form or be maintained, both parties must agree to it. A network is thus stable if any existing partnership is beneficial to both firms involved, and if the creation of any non-existing partnership would reduce the profit of at least one of the firms involved.[11] In the present model, the simple form of firms' profits implies that both existing and non existing links have their potential value determined in exactly the same way, the value of ij being $f(y_{ij}) - c$ to both i and j. As a consequence, the stable network g is simply $\{ij : f(y_{ij}) \geq c\}$. The effect of c is intuitive. For $c > f(\delta)$, no partnership is worth the cost, and the equilibrium network is the empty network (firms innovate only in isolation). When c is small enough, $c \leq \min\{f(0); f(w)\}$, all the links will form and the equilibrium network is the complete network. Finally when $\min\{f(0); f(w)\} < c < f(\delta)$, only those partnerships with firms having overlap y_{ij} such that $f(y_{ij}) \geq c$ will form, thus again there is a unique equilibrium network.

To re-frame things in terms of the inverted-U, define $\rho \geq 0$ by $f(\delta \pm \rho) = c$. Alliances at distance between $\delta - \rho$ and $\delta + \rho$ are profitable, since they have expected benefit greater than the cost c. Figure **??** illustrates this result and the proposition below states it.

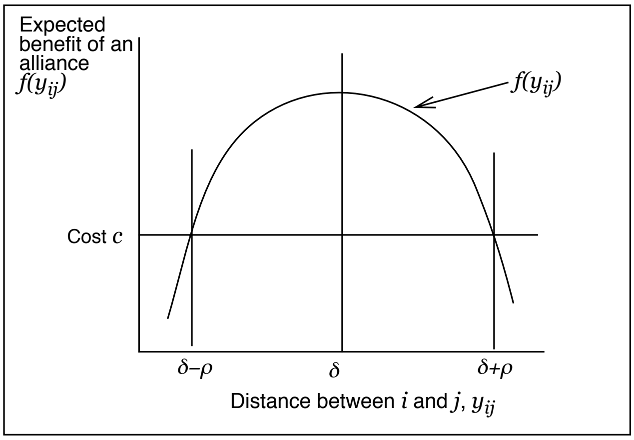

Figure 1: Costs and benefits of an alliance as a function of distance in knowledge space.

Proposition 1 *For any $c \geq 0$, there exists a unique equilibrium network g. When $c > f(\delta)$ the empty network is (uniquely) stable; when $c < \min\{f(0); f(w)\}$ the complete network is stable; when $\min\{f(0); f(w)\} < c < f(\delta)$ the stable network is $g = \{ij : |y_{ij} - \delta| \leq \rho\}$.*

[11]In the game theory literature, this equilibrium condition is called "pairwise stability" (Jackson and Wolinsky, 1996). Technically, in a pairwise stable network g, for any $ij \in g$ we must have $\min\{\pi_i^g - \pi_i^{g-ij}; \pi_j^g - \pi_j^{g-ij}\} \geq 0$ whereas any $ij \notin g$ must be such that $\min\{\pi_i^{g+ij} - \pi_i^g; \pi_j^{g+ij} - \pi_j^g\} < 0$.

Given the nature of joint innovation and the need for overlap in firms' knowledge stocks we are able to specify precisely which ties form in equilibrium, and that the equilibrium is unique.

2.2 The restricted model

Given the mechanics of innovation and firms' motivation to increase knowledge, a unique equilibrium exists. But its characterisation is very general and contains little information about the network's structural properties. To proceed further in describing them, we now adopt the assumption that $\rho = 0$, i.e. a partnership will form if and only if the two partnering firms' knowledge sets overlap in exactly δ places. As a consequence, the stable network, when neither empty nor complete, is $g = \{ij : y_{ij} = \delta\}$.

2.2.1 Degree

A common concern in research on alliance formation and networks in general has to do with the number of ties a firm forms. What determines degree, and why do agents differ in the numbers of partners they have? In this model, firms are differentiated only by their knowledge stocks. Thus the question we ask here is how the quantity of a firm's knowledge affects the probability that it can find a partner, and the expected number of partners it will have. Answers to these questions allow us to see how the distribution of knowledge drives the degree distribution of the network. Technically, this involves calculating probabilities conditional on the amount of knowledge a firm holds. The details are expositionally trying, so we consign them to an appendix.

We begin by distinguishing among firms according to how much knowledge they hold. That is, we condition on a_i, the number of ideas held by firm i. First note that a_i is binomially distributed with parameters w and $1/2$, yielding $\Pr\{a_i = a\} = \binom{w}{a}/2^w$. Given that firm i holds a_i pieces of knowledge, what is the probability that it forms a link with an arbitrary firm j? We can characterise the potential partnership, ij as a $2 \times w$ matrix, where each column represents a firm's knowledge endowment, and each row represents a place in the knowledge vector. Re-label the vectors so that i's knowledge is found in the first a_i positions. Up to this re-labeling, the condition under which i and j form a link can be represented as in Table **??**.

positions in knowledge vector	i	j	
1	1		
$\vdots$	$\vdots$		$\leftarrow \delta$ instances of 1
a	1		
$a+1$	0	$\star$	
$\vdots$	$\vdots$	$\vdots$	
w	0	$\star$	

Table 1: i and j's portfolios and the possibility of an alliance; $\star$ is the "don't care" symbol.

Conditional on $\{a_i = a\}$, the probability that i forms a partnership with $j \neq i$ is the probability that j has δ instances of 1 in the first a positions and anything in the remaining $w - a$. As a partnership requires that each firm holds at least δ ideas, one has

$$\Pr\{ij \in g | a_i = a\} = P_a = \begin{cases} \binom{a}{\delta} \left(\frac{1}{2}\right)^a, \ a \geq \delta, \\ 0, \ a < \delta. \end{cases} \tag{2}$$

This gives the probability that i forms a link to an arbitrarily chosen j.

More generally, the probability that i forms exactly l links, conditional on i having a_i ideas, is simply binomially distributed.[12] If $a > \delta$ then

$$\begin{aligned} \Pr\{n_i^g = l | a_i = a\} &= C_{n-1}^l P_a^l (1 - P_a)^{n-1-l} \\ &= \binom{n-1}{l} \left(\binom{a}{\delta} / 2^a \right)^l \left(1 - \binom{a}{\delta} / 2^a \right)^{n-1-l}, \end{aligned} \tag{3}$$

whereas if $a < \delta$

$$\Pr\{n_i^g = l | a_i = a\} = \begin{cases} 0, \ l > 0, \\ 1, \ l = 0. \end{cases} \tag{4}$$

The conditional probability of finding at least one partner is $1 - \Pr\{n_i^g = 0 | a_i = a\} = 1 - (1 - P_a)^{n-1}$. Finally the conditional degree distribution has mathematical expectation $E\left(n_i^g | a_i = a\right) = (n-1) P_a$. The following proposition, which follows straightforwardly from Equations **??** and **??**, summarizes the conditional results.

Proposition 2 *In the equilibrium network g, both a firm's probability of finding at least one partner and its expected degree increase and then decrease with a, the number of facts held by the firm, reaching their maximum for $a = 2\delta$.*

The appendix contains explicit functional forms of the degree distribution, the expected degree and the covariance of knowledge and degree. These functions are not particularly tractable, but Figure **??** shows exact curves for several parameter values. Three observations follow:

1. The mathematical degree distribution has a long tail, but is less than a power law in general.

2. For a given w, expected degree increases and then decreases with δ.

3. For a given w, the correlation between degree and knowledge is negative and falling (from 0) rises and becomes positive, and then falls to zero as the optimal overlap increases.

The upper right panel illustrates Proposition **??** and shows the relationship between success in finding a partner and knowledgeability, with the former being maximized when

[12] We should note that while the events ij and ik are not independent, the conditional events $\{ij | a_i\}$ and $\{ik | a_i\}$ are.

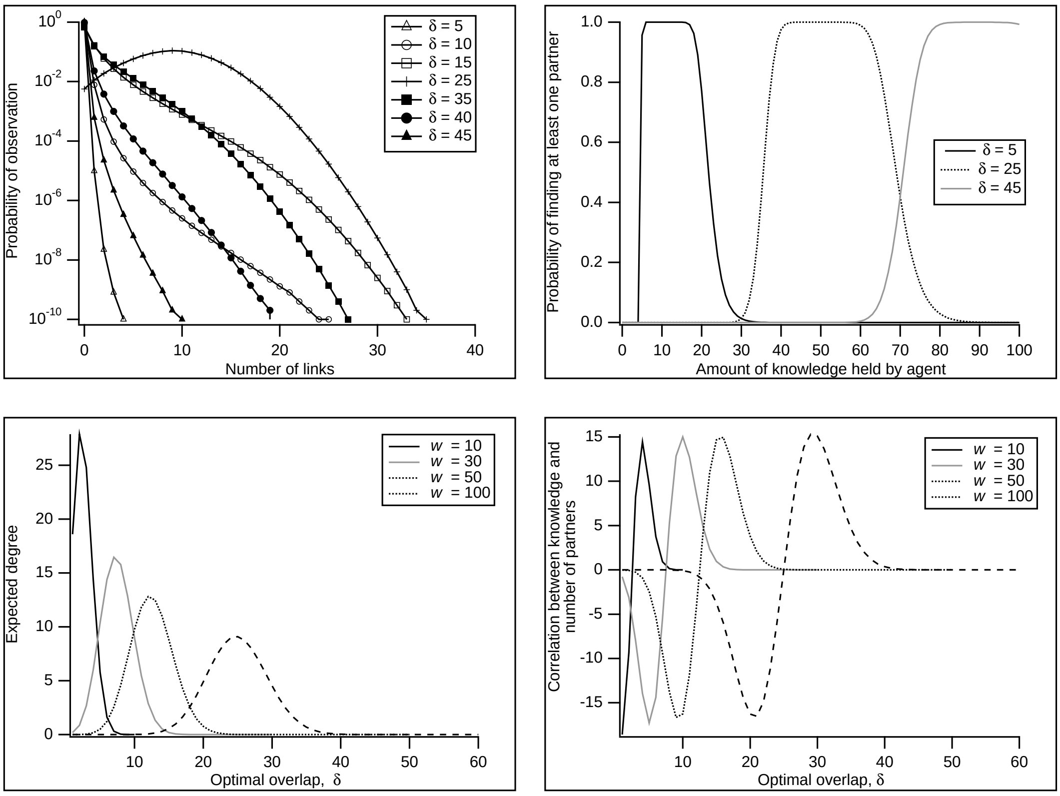

Figure 2: Upper left: Degree distribution as a function of δ with $w = 100$; Upper right: probability of finding at least one partner as a function of the amount of knowledge held by an agent for three values of δ with $w = 100$; lower left: expected degree and lower right: correlation between knowledge and degree as functions of δ for four values of w (note in the bottom panels that $0 \leq \delta \leq w$). In every case $n = 100$.

$a_i = 2\delta$. Each curve represents a different optimal overlap, with w fixed at 100 in each case.[13] As $E(a_i) = w/2$ in the industry, and the probability of finding a partner is maximized at $a_i = 2\delta$, it follows that $\delta = w/4$ is the value of the optimal overlap where partnering intensity is maximized. The upper left panel displays the degree distribution, which is generally decreasing with degree and therefore skewed, except for δ near $w/4$, the value of δ where partnering is most intense. We show a log-linear plot which indicates that the tail is exponential at most (in this representation the tails seem to be sub-linear for all values of δ, with the exception of the extreme $\delta = 5$ and $\delta = 45$ cases, indicating that the distribution has less than an exponential tail.) As seen in the lower left panel, the relationship between δ and expected degree also peaks when $\delta = w/4$.

The lower right panel is of slightly different nature, providing information about the direction of motion of the industry rather than its static state. If $\delta < w/4$, firms with relatively fewer ideas are relatively more successful in finding partners. Thus, looking

[13]The maximum probability is near 1, and the width of the region with probability near one increases with the number of firms: if n is large enough, virtually every firm will find a partner. If n is small, few firms will. Pisano (1990) discusses the "small number bargaining problem" as an explanation of a low degree. Our model provides a different explanation for low degree when there are few firms.

at the network that emerges from a single partnering episode, a negative correlation or covariance between degree and knowledge is to be expected. Knowledge-poor firms tend to have more partners, thus they tend to innovate more and grow faster than those rich in knowledge. This mechanism is different from decreasing returns in R&D or exhaustion of innovative potential: even without these effects, knowledge-poor firmsaccumulate faster than knowledge-rich firms, and the variance in knowledge held wil shrink. The reverse is true when $\delta > w/4$ and the correlation vanishes when $\delta = w/4$. We should point out that this has implications for any empirical work that implicitly assumes that firms have similar initial conditions with regard to knowledge. If this (simplifying) is assumption is incorrect, which seems likely, our results on correlation show that it may not be innocuous, and may introduce misleading correlation with other variables that proxy a firm's knowledge endowment.[14]

Proposition **??** permits one immediate conjecture regarding the dynamics of network formation. With time, all firms innovate: relatively quickly if they find partners; slowly if they do not. Firms who start with weak knowledge stocks find many partners and innovate quickly, meaning that their knowledge stocks grow rapidly. They soon have more than 2δ pieces of knowledge. They then have more difficulty finding partners, and as their knowledge grows more, the probability of finding a partner falls towards zero. Thus if the system begins with many knowledge-poor firms, the degree, and so the rate of innovation, will increase with time, and then decrease. If most firms are knowledge-rich at the outset, we will see a monotonic decrease in both degree and rate of innovation. These intuitions will be confirmed in Section**??**.

2.2.2 Clustering

The proportion of a firm's partners who are partners of each other measures the clustering of the firm's neighbourhood.[15] At the network level, the clustering coefficient is the average of the firms' neighbourhood clustering. Thus in a clustered graph, many closed triangles will exist. To examine the nature of clustering we ask whether triangle-closing links likely than random links. We can answer this question by looking at the probability of the link forming in two settings: first, that there is some i for which links ij and ik exist; and second in the absence of any condition. The second case is simply the probability that j and k find exactly δ matches in their knowledge. The first is the probability of finding δ matches conditional on the event $\{ij, ik\}$. How different are the two probabilities? Roughly speaking, if i is similar to j, and i is similar to k, then it is likely that j and k are similar to each other. If this similarity between a pair of agents increases the probability that a link forms between them, then the probability of triangle closing links will be higher than the probability of non triangle-closing links, and so clustering would be increased.[16]

[14] We thank a referee for pointing this out.

[15] The clustering of firm i's neighbourhood is the ratio of the number of ties among i's neighbours, $\sum_{j \neq k \in N_i^g} \chi_{jk}$, to the number of possible ties, $n_i^g(n_i^g - 1)/2$.

[16] Notice that this reasoning can be generalized beyond knowledge. Similarity in any property for which "similar to" is a transitive relationship will work the same way. Roller et al. (2007) for example, discuss the formation of RJVs in the context of similarity in terms of size, finding that firms of different sizes have disincentives to form RJVs. One relationship often investigated in alliance research which does not work this way is shared history, or relational embeddedness.

We use the characterisation used in section **??**, here for a trio of agents: a $3 \times w$ matrix where each row represents a position in the knowledge vector and each column represents one of the agents. Each row in the matrix is a three digit binary string, from the set $\{000, 001, 010 \ldots, 111\}$. The key to understanding the probability of links forming is to determine the combinations of these three-digit strings that must be present. This is the path laid out below.

The probability that jk forms is simply the probability that j and k share exactly δ pieces of knowledge, which is binomial and is written $\Pr\{jk\} = \binom{w}{\delta} \cdot 3^{w-\delta}/4^{w}$. However, because the formation of the third edge of a triangle is not independent of the existence of the other two, (there can be elements common to all 3 firms) we must calculate the conditional probability more carefully. Suppose ij and ik exist. Denote B the (random) number of elements common to all three i, j and k. If we assume that $B = b$, then up to a relabeling of the positions, the three firms j, i and k can be represented as in Table **??**.

positions	j	i	k	
1	1	1	1	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\leftarrow b$ instances of 111
b	1	1	1	
$b+1$	1	1	0	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\leftarrow \delta - b$ instances of 110 to make ij
δ	1	1	0	
$\delta+1$	0	1	1	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\leftarrow \delta - b$ instances of 011 to make ik
$2\delta - b$	0	1	1	
$2\delta - b + 1$				
$\vdots$				$\leftarrow \delta - b$ instances of 101, to make jk and
w				no instances of 111, 110 or 011 to preserve ij and ik

Table 2: Knowledge portfolios of i, j and k, and the possibility of a triangle.

The condition that ij and ik exist, with the additional condition that $B = b$ together fully determine the first $2\delta - b$ positions (given re-labelling), and demand that the strings 111, 110 and 011 do not appear in the rest of the matrix. Thus the link jk forms, conditional on $\{B = b, ij, ik\}$, if and only if the remaining $w - 2\delta + b$ positions (the empty box in table **??**) contain exactly $\delta - b$ instances of the pattern 101 and any combination of patterns 100, 010, 001, 000 in the remaining $w - 3\delta + 2b$ positions. This outlines the combinatorial problem, and following the derivation in the appendix, we show that the ratio of a triangle-closing to a random tie is given by

$$\frac{\Pr\{jk|ij, ik\}}{\Pr\{jk\}} = \frac{\sum_{b=\beta}^{\delta} \binom{w-2\delta+b}{\delta-b} \binom{w}{b} \binom{w-b}{\delta-b} \binom{w-\delta}{\delta-b} 2^{-6\delta+4b-w}}{4^{-w} 3^{w-\delta} \sum_{b'=\beta'}^{\delta} \binom{w}{\delta} \binom{w}{b'} \binom{w-b'}{\delta-b'} \binom{w-\delta}{\delta-b'} 8^{-w} 5^{w-2\delta+b'}}, \tag{5}$$

where $\beta = \max(0, (3\delta - w)/2)$ and $\beta\prime = \max(0, 2\delta - w)$.

Further analysis of this ratio is intellectually trying, but Figure **??** shows exact curves for several parameter values. The ratio of conditional to unconditional probability is always strictly greater than one. For a fixed w, it decreases and then increases rapidly in δ. For a given δ, the ratio decreases rapidly and then increases (slightly) with w, and when δ is close to w, the ratio gets extremely large. In the extreme case, if $\delta = w$, the conditional probability is 1, whereas the unconditional probability is $(1/4)^w$. As w increases the ratio approaches infinity; as w falls, it approaches 4. The probability of any link forming is very small, but no open triangles will exist.

The ratio of conditional to unconditional probabilities does not map directly into the clustering coefficient. However, it does indicate that the constraint on alliance formation, namely that partners have an optimal overlap in knowledge endowments, means that emergent networks will be strongly clustered, and that clustering will tend to decrease with the size of the knowledge space of the industry, and increase with the size of the optimal overlap. The following proposition states these results.

Proposition 3 *The probability that jk exists conditional on the existence of ij and ik is greater than the unconditional probability that jk exists. The constraint of an optimal overlap sustains the formation of triangles, possibly to a very large extent.*

The intuition underlying this result is that having a common neighbour, i, raises the probability that j and k have at least some elements in common, since they both have things in common with i. This gives them a head start, so to speak, in the search for common elements, relative to a pair of agents who have no common neighbour. The non-monotonicity of the relationships arise because the search for "hits" (places where both partners have knowledge) is formally equivalent to the search for misses, since each place in the knowledge vector will be either a hit or a miss. The asymmetry arises because the probability of hits and misses are different. These results are illustrated in Figure **??**.

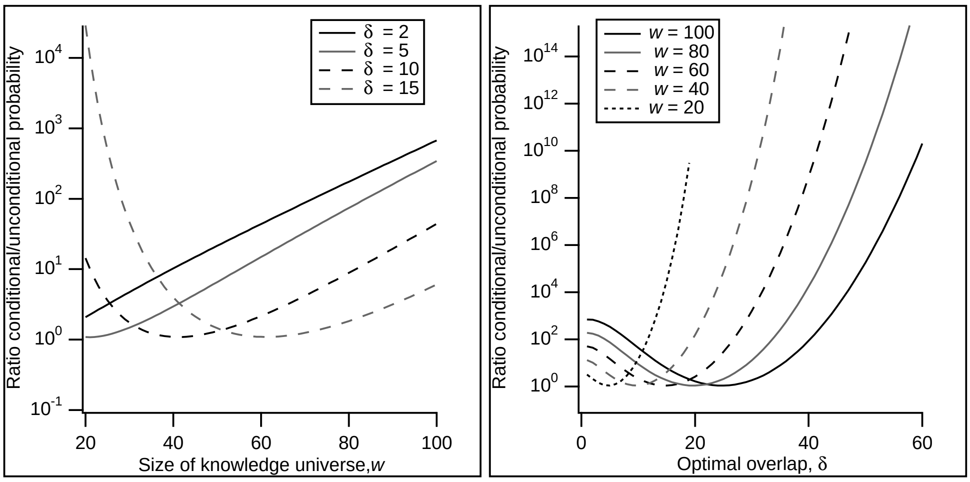

Figure 3: Ratio of conditional to unconditional probability. Left panel, as a function of w for 4 values of δ; right panel, as a function of δ for 5 values of w. In the right panel, both axes have been truncated to make the graphs more readable for smaller values of δ. Extrapolation of the curves over the entire abscissa ($[1, w]$) is not misleading.

The parameter δ gives the optimal amount of knowledge similarity. Similarity is relative, though to the total amount of knowledge in the economy (the union of the knowledge of all firms, which in the analysis above is measured by w). We can see from Figure **??** that the relationship between optimal similarity and clique-formation is non-monotonic, minimal for intermediate values of δ. We can show numerically (and seen in the right panel of figure **??**) that the minimum ratio occurs for $\delta = w/4$, and there is slightly larger than 1 (approximately 1.09). Moving away from $\delta = w/4$ so that firms need either a large or a small amount of overlap in what they know, triangle-closing link formation rapidly becomes over-whelmingly more likely than random link formation, and thus clustering relative to that of an equivalent random graph will be very high. In the extremes it is very difficult to find a partner, so degree will be small, but there will be essentially no open triangles.

What this analysis suggests then, is that a major part of the explanation for the high clustering typically observed empirically in strategic alliance networks, lies in the need for optimal amounts of similarity and dissimilarity between partners.

2.2.3 Small worlds

High clustering and low average path lengths are the defining features of small worlds. Analytic derivation of these statistics is not possible even for the reduced version of this model. The results so far suggest, though, that small worlds are likely to be present, at least for some parameter values. We can test that conjecture numerically, simply be creating a large number of networks and computing clustering and path length statistics. Setting $w = 100$, and $n = 100$, we generate 1000 networks, using the link creation rules described above, for different values of δ. We calculate average shortest path lengths and network clustering coefficients, both raw (d and cl) and normalized (d^* and cl^*) by an equivalent random graph, for each of those networks to show the presence of small worlds.

To cope with the issue of isolated firms, we focus on the largest component when discussing clustering and characteristic path length. The largest component is the largest subset of the industry such that there exists a finite path between any two firms in that subset. For structural properties such as clustering and path length particularly, in the literature it is customary to compare the results of any model with these properties in an equivalent random graph. "Equivalent" here is typically defined by network density, and the measures are simply adopted from random graph theory. Denoting $\bar{n} = \sum_{i \in \mathcal{N}} n_i^g / n$ the average degree, the standard approximations are $\bar{n}/(n-1)$ for the clustering coefficient and $\ln(n)/\ln(\bar{n})$ for the characteristic path length. These approximations are accurate for large n and non-extreme $\bar{n}$. However, for networks of the size we consider here they can be very misleading. Consequently, we normalize our output by a statistic calculated in the following way. For each network we wish to normalize, we record density and then generate 1,000 random networks with exactly that density. We then average clustering and path length over that sample. These we take as expected values for a random network, and then report the ratio of these statistics generated in the model, to these expected values.

In addition we show the skewness of the link distribution, and the correlation between knowledge levels and degree. These data are shown in table **??**.

As is clear from the table there are two regions in parameter space that exhibit small world properties. The ratio of normalized clustering to normalized path length (column 9) is large; and the network is relatively sparse in these regions (column 3). Small worlds

δ	Largest Comp.	Deg. $\bar{n}$	Corr Knowl-Deg.	Clustering cl	cl^*	Distances d	d^*	Ratio cl^*/d^*	Skew.
10	1.40	0.36	−0.04	0.00	0.00	0.36	0.26	0.00	1.41
13	5.43	1.44	−0.17	0.02	9.72	1.60	1.15	7.96	117.11
16	38.77	2.41	−0.42	0.08	15.95	3.55	1.05	15.35	978.97
19	81.28	4.42	−0.74	0.09	2.72	2.97	0.79	3.54	529.19
22	97.27	7.60	−0.69	0.10	1.33	2.49	0.99	1.35	67.61
25	99.33	9.22	0.01	0.09	1.03	2.34	1.01	1.02	0.22
28	95.24	7.39	0.69	0.09	1.34	2.53	0.98	1.38	−3.80
31	76.97	4.34	0.71	0.09	2.97	3.01	0.76	4.01	517.12
34	40.65	2.61	0.47	0.09	16.15	3.54	0.94	15.77	648.55
37	8.96	1.74	0.24	0.05	21.57	2.20	1.53	15.17	161.96
40	2.63	0.95	0.11	0.01	5.04	0.98	0.71	5.45	15.98

Table 3: Small world characteristics for the innovation networks; $\rho = 0$, $w = 100$.

exist when the optimal overlap, δ is neither large, nor small, nor close to the value $w/4$. Column 5 of Table **??** shows absolute clustering, which we see does not change for values of δ between 19 and 34. This means that changes in re-scaled clustering are driven by the denominator, that is, clustering in the equivalent random graph. This is determined almost analytically by degree, so the gap in the small world region (when $20 < \delta < 30$) is driven by the increase in degree in that region. In addition, the table shows that where the ratio of clustering to path length takes large values, the link distribution is also very skewed (column 11). Our simple model reproduces the observed structures of knowledge networks (small worlds with skewed link distributions) very neatly.

In Table **??** we also observe that the correlation between knowledge and degree changes sign with δ. Recall that on average here firms have $w/2$ pieces of knowledge. When firm knowledge is less than $\delta/2$ correlation is negative; when it is greater than $\delta/2$ it is positive. This reiterates the patterns shown in Figure **??**.

2.2.4 Repeated ties

The simple model also makes predictions about the likelihood that two firms engage in repeated ties.

We begin by observing that the simple model has a result that is too strong. If i and j form a partnership they have exactly δ pieces of knowledge in common. If they innovate successfully, the have $\delta + 1$ pieces in common. They never partner again, so repeated ties cease after a successful partnership. This is not necessarily the case in the more general model, in which the partnership criterion is relaxed, and we take that up below. But even if a partnership is unsuccessful, the probability of repeated ties would fall monotonically with time, simply because partner firms i and j may have partnerships with other firms. Suppose, for example, i has partnerships with both j and k. If the ik partnership discovers a fact known by j, then by that discovery, i and j now have $\delta + 1$ facts in common. Thus the probability of repeated ties decreases over time.

As we remarked, this very stark result is driven by the very strict requirement for partnership formation. When that is relaxed, as in the next section, we have more interesting dynamics with respect to repeated ties, which mimic the findings from the empirical literature.

2.3 The general model

The results above have been derived for a particularly stark version of the model, involving *i*) a very tight definition of a feasible partner (a firm that matches exactly δ pieces of knowledge) and *ii*) the assumption of identically and independently distributed knowledge endowments. We now turn to the numerical exploration of the more general model, in which both of these assumptions are relaxed. Regarding *i*), there remains an optimal distance, δ, but partners at distance near δ will also have innovation probabilities high enough that those links have positive expected value ($\rho > 0$). Maintaining the assumption of independence in knowledge endowments we will see that the analytical results generalize to this case. Regarding *ii*), there are two issues that are addressed. If a firm innovates over time, its knowledge endowment will increase, and thus the probability that it knows any particular idea grows over time. Second, the assumption that firms' knowledge stocks are independent of each other will be violated as soon as joint innovation takes place. If i and j innovate together, this will introduce some correlation into their knowledge stocks. It is not obvious how to model this *a priori*, so we relax the independence temporally, simulating a dynamic version of the model. We examine time series of degree, clustering, firms' propensity to repeat ties and correlation between knowledge and degree to see whether the change in knowledge stocks alone may have an impact on the dynamics of a network, and whether the analytic results are good approximations even when knowledge stocks are correlated due to a history of joint innovation.

2.3.1 Settings

All $n = 100$ firms begin with ideas beneath the technological frontier $w = 100$, with $\Pr\{z \in \theta_i\} = 1/2$. The frontier is pushed forward by the discovery of ideas beyond w. Formally, this is done by drawing the location of an innovation uniformly at random in $\{1, \ldots, w + 1\}$, where neither partner is already knowledgeable. This way, the frontier slowly expands, and the most knowledgeable firms are most likely to be the ones pushing the frontier: the technological potential of the industry can continue indefinitely.

Set a small autarchic innovation rate $1/\lambda$, using $\lambda = 200$. Relax the restriction that ρ equals 0, and instead consider an inverted-U with a strictly positive base so that

$$f(y) = \frac{1}{\lambda\rho^2}\left(-y^2 + 2\delta y - \delta^2 + 2\rho^2\right), \ \rho > 0. \tag{6}$$

This implies that $f(\delta) = 2/\lambda$ (twice the innovation rate of internal R&D). Finally set $c = 1/\lambda$, and let $f(\delta - \rho) = f(\delta + \rho) = c$, using ρ to control the width of the inverted U. As alliances form if $f(y) \geq c$, the base width of the inverted-U (2ρ) and c play equivalent roles: a larger c is equivalent to a smaller ρ, both reducing the number of partnerships that can form. The equilibrium network is $g = \{ij : f(y_{ij}) \geq c\} = \{ij : |y_{ij} - \delta| \leq \rho\}$. When $\rho = 0$ the restricted model obtains, with $f(y) = 2/\lambda$, when $y = \delta$ and 0 otherwise.

There remain two independent parameters, δ and ρ, which we vary from 0 to 50 and 0 to 25 respectively. For each point in the parameter space, we generate 1000 independent replications, and retain average values.

2.3.2 Relaxing optimal overlap

The analytical results in Section ?? above suggested how the network structure (in particular degree) in a single period would respond to different parameters and states of the agents. In the first subsection below we show that these same properties hold with a relaxed overlap condition.

δ	Largest	Deg.	Corr	Clustering		Distances		Ratio	Skew.
	Comp.	$\bar{n}$	Knowl-Deg.	cl	cl/cl^*	d	d/d^*	cl^*/d^*	
7	69.68	5.25	−0.57	0.36	10.12	2.41	0.65	16.36	24038.18
10	95.58	15.30	−0.88	0.58	4.04	1.94	0.98	4.14	3154.95
13	99.77	36.84	−0.96	0.71	1.93	1.63	1.00	1.93	0.58
16	100.00	63.56	−0.96	0.81	1.26	1.36	1.00	1.26	−0.36
19	100.00	84.24	−0.89	0.90	1.06	1.15	1.00	1.06	−1.44
22	100.00	94.58	−0.76	0.97	1.01	1.04	1.00	1.01	−2.73
25	100.00	97.53	−0.19	0.99	1.00	1.01	1.00	1.00	−3.71
28	100.00	95.09	0.69	0.97	1.01	1.04	1.00	1.01	−3.33
31	100.00	84.29	0.88	0.91	1.06	1.15	1.00	1.07	−1.79
34	99.89	63.06	0.96	0.82	1.29	1.36	1.00	1.29	−0.60
37	98.66	35.87	0.97	0.74	2.10	1.65	1.00	2.09	0.39
40	90.73	16.11	0.90	0.68	4.75	1.94	0.98	4.86	1262.20
43	64.58	6.68	0.67	0.58	14.42	2.27	0.67	23.19	13557.86

Table 4: Small world characteristics for the innovation networks; general case $\rho = 10$, $w = 100$.

We can see that Tables ?? and ?? are very similar, and with the more generous condition on alliance profitablity the model again produces the stylized facts. We observe small worlds with skewed link distributions, though the small world region is pushed towards the extremes of the δ range: small worlds exist when δ is either very large or very small, and when they exist, the link distribution is very skewed. The effect of relaxing the overlap constraint implies that more pairs of firms are able to form profitable partnerships. But small worlds can only exist when networks are sparse. The expected size of overlap between two arbitrary firms is $w/4$. If δ is in an intermediate range, and the inverted -U permits alliances with overlap some distance from δ many partnerships form and the network is dense. Thus small worlds can exist only for extreme values of δ, in which case only pairs with unusually small (or large) amounts of shared knowledge can partner.

2.4 A dynamic model

The second assumption we made for the restricted, analytical model was that firms' knowledge endowments were identically and independently distributed. We relax that here, but do so by implementing the dynamic version of the model. This way, the non-independence arises naturally from joint innovation, as might be expected empirically. The settings remain as above, and each period firms form all profitable (in expected value) alliances, attempt to innovate, and then dissolve the alliances. We examine the evolution of network structures and the incidence of repeated ties.

2.4.1 Network structures

As pointed out above, the time series should follow a pattern predictable from the behaviour of the restricted model. In essence, as firms innovate their knowledge stocks relative to δ increase. As there is no inherent source of inertia in the model, the alliance formation process begins afresh each period but with firms having, on average, more knowledge. Relative to the amount of knowledge held by firms, δ falls, so the time series pattern should resemble a gradual transition from a large δ to a small δ in the snapshot results above. This simple pattern is complicated, however, by the fact that innovation generally, and joint innovation particularly, imply that firms' portfolios of knowledge become increasingly correlated with time. This applies both to pairs of partners and to the population generally. As a direct consequence eventually it becomes harder and harder to find a partner. The general pattern is the result of the combination of these two effects.

If the first effect dominates, degree should rise and then fall as we move, through internal R&D and then through joint innovation, from a context of knowledge scarcity where it is difficult to find partners, to a context of knowledge abundance where firms have "too many" ideas and it is again difficult to find partners. But if firms begin with much knowledge relative to δ, only the tail of this process is observed. Clustering should follow a similar pattern: high levels of clustering early in the process since only a few agents are able to form links but these tend to be grouped in small cliques. It will increase as degree increases, but excess clustering will fall, mechanically, with degree, Finally, distance will increase and then decrease, again driven by changes in degree. As small worlds are characterised by low distances and high clustering, the extent to which small worlds are present will rise and fall with time.

The series shown in Figure **??** confirm the conjectures above. We merely add the observation that small world behaviour, as seen in the ratio of clustering to distance, is driven almost entirely by normalized clustering, as normalized distance is always close to one.

At the firm level, it is worth asking who is involved in alliances. Again, mechanisms identified in the stark, one-period model apply. If firms start with too much knowledge on average (when δ is low) the correlation between degree and knowledge is negative and remains so. If firms start with too little, on the other hand (if δ is large), initially those with above average knowledge levels find partners easily, but as the population as a whole accumulates knowledge, this changes, and over time it is those firms with relatively little knowledge who have the larger degree (the negative correlation seen in the upper right panel).

Through almost the entire history of the system the skewness of the degree distribution is positive, and is very large for some periods. This indicates a distribution with a relatively heavy right hand tail. Changes in the magnitude of skewness are closely connected to changes in density (average degree). Initial endowments and the nature of innovation imply that many firms exhaust the pool of their potential partners at roughly the same time, which is seen in the rapid increase in skewness (a large group of firms suddenly has zero degree). The subsequent gradual decrease arises as the remaining firms gradually exit the alliance game.

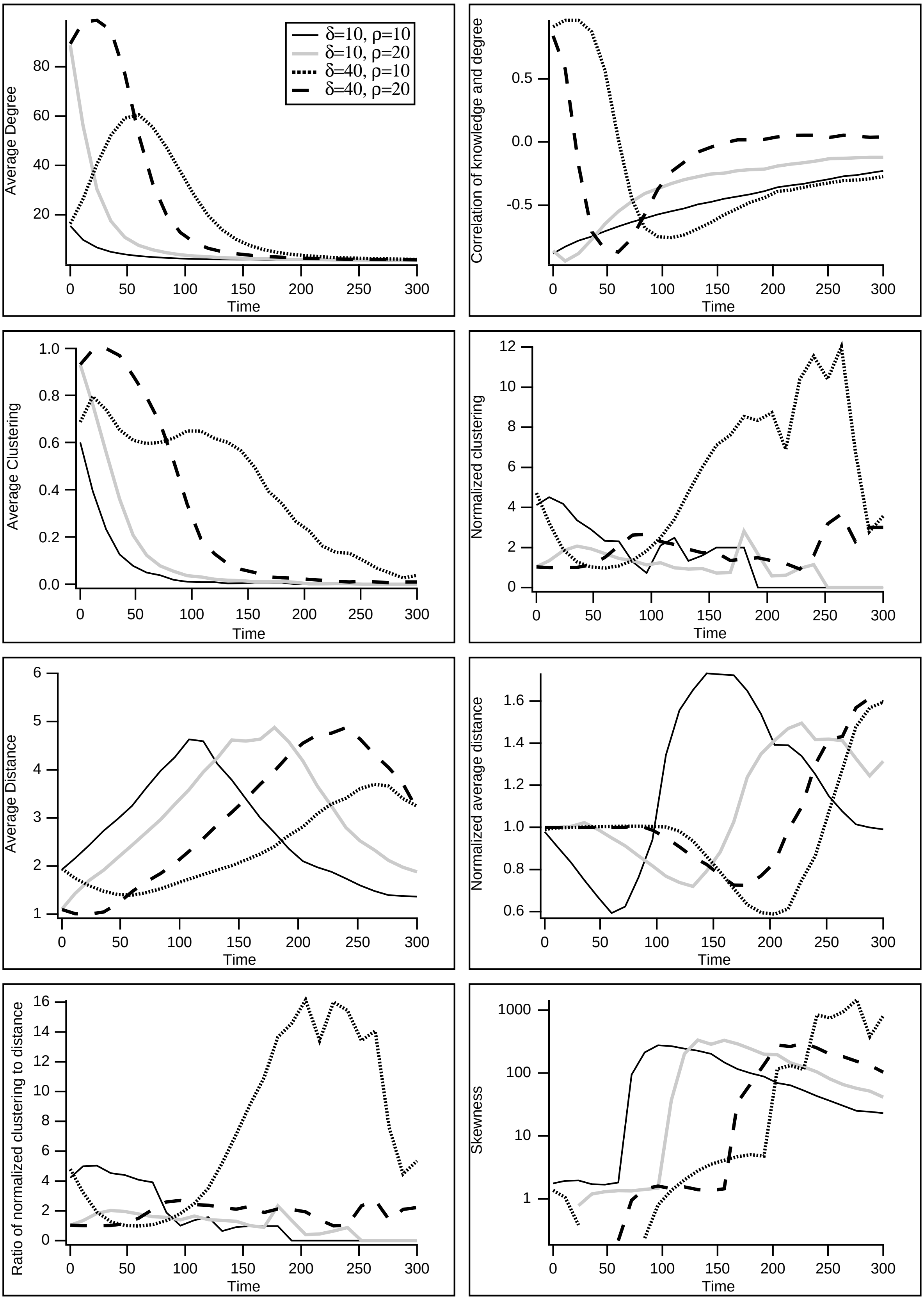

Figure 4: Time series for several structural parameters. The legend in the top left panel applies to all panels.

2.4.2 Repeated ties

An issue of some concern in the empirical literature on alliances is the propensity for a firm to revisit old alliance partners. In particular, one explanation of a partnership formation is relational embeddedness: whether or not the partnership has a history of successful collaboration (see for example Dyer and Chu, 2003; Gulati, 1995; Podolny, 1994; or Walker et al. 1997). Repeated ties are strong ties which offer high quality exchange of (particularly tacit) information (Uzzi, 1996), and the can act a social control (Powell, 1990 or Uzzi, 1996 for example). Many papers find an increasing relationship (see Rowley et al. 2000, for example) between length of history and propensity to re-partner, though Chung et al. (2000) and Gulati (1995) both find that the length of a partnership history has a non-monotonic effect on the probability that the partnership re-forms. The explanations in the literature have to do with learning how to cooperate, developing shared routines and so on. Figure **??** shows this effect in our model. The horizontal axis measures the number of times a partnership has been formed (which we call history length); the vertical axis shows the proportion of partnerships that re-form, given a history length. What we observe is that even in the absence of any learning to cooperate, trust generation etc., a relationship between history and repeated tie formation emerges.

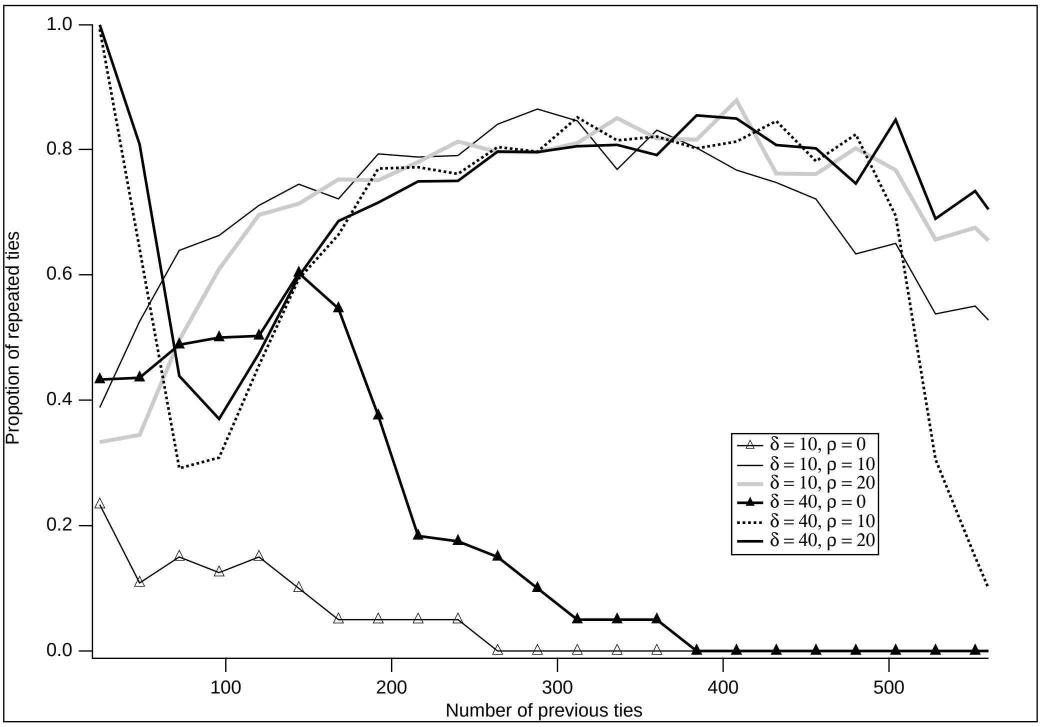

Figure 5: Proportion of repeated ties conditional on the number of previous ties (length of history)

If it were the case that the accumulation of knowledge had no effect on the profitability of alliances, in our model all alliances would repeat forever, and the proportion of repeated ties for any history length, would be 1. This is not the case, and the patterns in Figure **??** are driven by partnerships dissolving as the member firms accumulate knowledge. We can see three different patterns in the figure. For $\rho = 0$ the proportion of ties that repeat falls with history length (when $\delta = 40$ there seems to be a slight rise initially). When the optimal overlap is low ($\delta = 10$), we have a quadratic relationship: initially increasing history length increases the probability that a tie will repeat, and after a certain point, increasing history length decreases that probability. When the optimal overlap is very large ($\delta = 40$) there are two phases: very short histories repeat, but the value of a history falls rapidly with its length. In the second phase, in inverted-U appears, as in the small δ case. The initial fall is explained by the fact that a large group of firms (firms that have amounts of knowledge near the mode of the distribution) enter the alliance game roughly at the same time, and create large numbers of alliances within that group. The group as a whole accumulates knowledge rapidly, and at roughly the same rate. As a group again, they rapidly have "too much" knowledge, and can no longer find partners, and they exit the population of alliances: their ties no longer repeat. The rest of the space is explained by the more gradual entry and exit of firms into the population of alliances.

3 Discussion and conclusion

One implication of this model is that as an industry matures, and as the innovation opportunities are found and taken up, strategic alliance activity will respond first by increasing, as firms are better able to find partners that complement their own abilities, and then decreasing, as the overlap between firms' competences rises. As a heuristic examination we can look at alliance history in the biotech industry. Figure **??** shows the average number of alliances per firm between 1986 and 2005, increasing and then falling. It also shows that network clustering also rises and then falls, with roughly similar timing.[17] We should hasten to say that this is not meant to discount other possible explanations for this pattern, such as, for example, that financial successes or difficulties may change firms' alliance behaviour (Lerner et al. 2002, for example), but the similarity between the observed patterns and the patterns generated by our model (see Figure **??**) is suggestive.

Discussions of alliances emphasize that an important reason firms form alliances is to find complementary knowledge assets. Theoretical discussions of knowledge complementarities have made the convincing case that firms' knowledge assets, and how they fit together, play a major role in the success of knowledge-creating alliances. Empirical studies of the structure of the network of strategic alliances though, while acknowledging the importance in principle of knowledge complementarities, have generally had difficulty incorporating them directly into statistical analyses of partner choice.[18] In this paper we have argued that the important role of knowledge complementarity in partner choice can be used to

[17]Data are from the Recombinant Capital *recap* database of worldwide alliances in biotech. Both curves show three year moving averages. The "All Alliances" series are normalized by the number of firms in the industry; the "R&D Alliances" series is normalized by the number of firms that were involved in an R&D alliance within the previous three years, and which, in the clustering panel, contributed positively to the clustering coefficient.

[18]Some work of Gulati provides an exception. Gulati (1995) considers the effects of strategic complementarities between a pair of firms, defined in terms of skills, resources and historical development, on their propensity to form an alliance. Gulati and Gargiulo (1999) consider strategic interdependence: the distance between two firms, defined in terms of national origin (meant to capture geographic clustering of capabilities) and industrial subsegment (meant to capture complementarities across different technological niches). They find that a bigger distance between two firms in these dimensions increases the probability

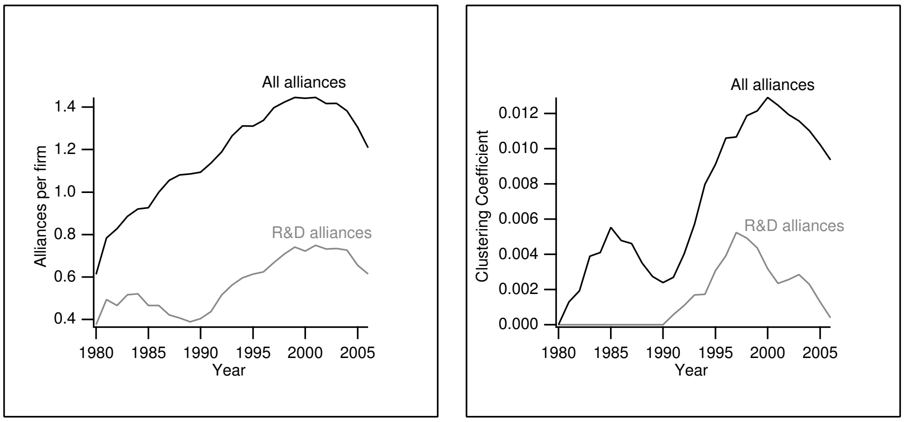

Figure 6: Number of alliances per year, per firm in biotech; Clustering in the biotech industry network (three-year moving averages); All alliances (black curves) and R&D alliances (grey curves).

explain resultant network structures. This suggests that it would be desirable to control for knowledge complementarity in empirical work regarding alliance formation.

In a very simple model we have shown that standard observations about alliance networks, or knowledge networks more generally, can be reproduced by focussing on the central issue at the heart of those networks, namely the production of new knowledge. While not denying the importance to firms of considerations such as relational or social embeddedness, we show that when innovation is the goal of an alliance and success demands knowledge complementarities, these may be enough to determine network structure. The model also shows the way in which innovation in an industry can in and of itself reduce the incentives to form alliances. As firms learn more and more, the value of partnerships decreases as it becomes difficult to find partners at the optimal distance. Innovation in isolation becomes more effective. Innovation and technical progress can continue, but alliance activity will decrease.

Pushing slightly beyond the bounds of the model, we can use it to address several issues.

The goal of many R&D alliances is to secure complementary knowledge or competences, so the opportunity cost of forming an alliance is the knowledge that could be acquired by other means, such as for example, querying public knowledge sources. Typically as an industry ages, tacit knowledge tends to be replaced with codified knowledge. Codified knowledge diffuses relatively easily, particularly in the context of a dominant design. Face-to-face interactions, necessary for tacit knowledge transmission, become less central to information flows, since other means exist to transmit codified knowledge. The efficiency of acquiring knowledge from public sources increases, and so the opportunity cost of alliance formation increases. Thus for an alliance to be profitable, partners must be closer to the optimal distance in knowledge space, since sub-optimal matches may be dominated by other means of knowledge acquisition (market acquisition such as licensing or sub-contracting for example). In the model this corresponds to a reduction in the width of the inverted-U (a

that they will form an alliance. In both papers, the notion of knowledge complementarity is relatively rough. Nohria and Garcia-Pont (1991) also examine strategic complementarities, but defined not in terms of knowledge but more in general structural features of a firm.

reduction in ρ).

Another way that ρ can be used is to capture differences between exploration and exploitation. One could expect that when firms are largely engaged in exploration, in general they are less specific about the complementary knowledge they need, in part because they are not sure what they are looking for. This can be captured in the model through a wide inverted-U, or a large value of ρ.

Some innovation projects are modular in that partners need not have heavy interactions to make them successful. Firm i does its part, firm j does its part, and in the end the two parts are put together. By contrast, some projects demand heavy interaction throughout. In the latter case, firms must be able to interact efficiently over the course of the project, and so must have a large stock of common knowledge to be able to communicate well. In the former case, less information is passed during the project, so less common foundation is necessary. This consideration is captured by the magnitude of the optimal overlap (δ) — it must be large in the second case, and can be small in the first.

Finally, to understand a single period of alliance formation the model may be quite useful, and applied relatively directly. However, as firms innovate the knowledge universe increases in size (w grows), while by assumption δ, the optimal overlap, remains unchanged. How the optimal overlap changes as the total amount of knowledge in the industry increases is unclear. If the optimal overlap has a smaller growth rate than does the aggregate knowledge level, which could be an effect of increased codification, then empirical behaviour is likely to be well-represented by our time series results.

References

[1] Ahuja, G. (2000) "Collaboration networks, structural holes and innovation: A longitudinal study" *Administrative Science Quarterly*, 45, 425-455.

[2] Ahuja, G. and R. Katila (2001) "Technological acquisitions and the innovation performance of acquiring firms: A longitudinal study" *Strategic Management Journal*, 22, 197-220.

[3] Audretsch, D.B. and M.P. Feldman (1996) "Knowledge spillovers and the geography of innovation and production" *American Economic Review,* 86, 630-640.

[4] Baptista, R. (2000) "Do innovations diffuse faster within geographical clusters?" *International Journal of Industrial Organization*, 18, 515-535.

[5] Baum, J.A., A.V. Shipilov and T.J. Rowley (2003) "Where Do Small Worlds Come from?" *Industrial and Corporate Change*, 12, 697-725.

[6] Burt, R. S. (1992), *Structural Holes: The Social Structure of Competition,* Harvard University Press, Cambridge, Mass.

[7] Capello, R. (1999) "Spatial transfer of knowledge in high technology milieux: learning versus collective learning processes" *Regional Studies,* 33, 353-365.

[8] Chung, S., S. Singh and K. Lee (2000) "Complementarity, Status Similarity and Social Capital as Drivers of Alliance Formation" *Strategic Management Journal*, 21, 1-22.

[9] Coleman, J.A. (1988) "Social Capital in the Creation of Human Capital" *American Journal of Sociology*, 94, S95-S120.

[10] Cowan, R., N. Jonard and J-B. Zimmermann (2007) "Bilateral Collaboration and the Emergence of Innovation Networks", *Management Science*, 53(7), 1051-1067.

[11] Dyer, J.H and W. Chu (2003) "The role of trustworthiness in reducing transaction costs and improving performance: Empirical evidence from the United States, Japan and Korea" *Organization Science* 14(1), 57-68.

[12] Dyer, J.H. and K. Nobeoka (2000) "Creating and Managing a High Performance Knowledge Sharing Network: The Toyota Case" *Strategic Management Journal*, 21, 345-367.

[13] Feldman, M. (1999) "The new economics of innovation, spillovers and agglomeration: a review of empirical studies" *Economics of Innovation and New Technology*, 8, 5-26.

[14] Garcia-Pont, C. and N. Nohria (2002) "Local versus Global Mimetism: The Dynamics of Alliance Formation in the Automobile Industry" *Strategic Management Journal*, 23, 307-21.

[15] Grant, R.M. (1996) "Towards a Knowledge-Based Theory of the Firm" *Strategic Management Journal*, 17, 109-122.

[16] Gulati, R. (1995) "Social Structure and Alliance Formation Patterns: A Longitudinal Analysis" *Administrative Science Quarterly*, 40, 619-652.

[17] Gulati, R. and M. Gargiulo (1999) "Where Do Inter-Organizational Networks Come from?" *American Journal of Sociology*, 104, 1439-1493.

[18] Jackson, M.O. and A. Wolinsky (1996) "A Strategic Model of Social and Economic Networks" *Journal of Economic Theory*, 71, 44-74.

[19] Johnson, D.K. and M. Mareva (2002) "It's a Small(er) World: The role of Geography and Networks in Biotechnology Innovation" Colorado College Department of Economics and Business Working Paper 2006-03.

[20] Kogut, B., W. Shan and G. Walker (1992) "Competitive Cooperation in Biotechnology: Learning through Networks?" in N. Nohria and R. Eccles (Eds.) *Networks and Organizations: Structure, Form and Action*, Harvard Business School Press, Boston.

[21] Lerner, J., H. Shane and A. Tsai, (2003) "Do equity financing cycles matter? evidence from biotechnology alliances" *Journal of Financial Economics* 67, 411-446.

[22] Mowery, D.C., Oxley, J.E. and B.S. Silverman (1996), "Strategic Alliances and Interfirm Knowledge Transfer" *Strategic Management Journal*, 17, 77-91.

[23] Mowery, D.C., Oxley, J.E. and B.S. Silverman (1998) "Technological Overlap and Interfirm Cooperation: Implications for the Resource-based View of the Firm" *Research Policy*, 27, 507-523.

[24] Nakamura, M., Shaver J.M. and B. Yeung (1996) “An Empirical Investigation of Joint Venture Dynamics: Evidence from U.S.-Japan Joint Ventures” *International Journal of Industrial Organization*, 14, 521-541.

[25] Newman, M.E. (2001) “The Structure of Scientific Collaboration Networks” *Proceedings of the National Academy of Science USA,* 98, 404-409.

[26] Nohria, N and C. Garcia-Pont (1991). “Global Strategic Linkages and Industry Structure” *Strategic management Journal*, 12, Special Issue, pp. 105-124.

[27] Nooteboom, B. (2000) *Learning and Innovation in Organizations and Economics* Oxford University Press.

[28] Pisano, G. (1990)“The R&D Boundaries of the Firm: An Empirical Analysis” *Administrative Science Quarterly* 35(1), 153-176.

[29] Podolny J. (1994) “Markeet uncertainty and the social character of economic exchange” *Administrative Science Quarterly* 39, 458-483.

[30] Powell, W. (1990) “Neither market nor hierarchy: Network forms of organization” *Research in Organizational Behavior* vol 12, Staw, B.M, and L.L. Cummings (eds.) JAI Press: Greenwich Conn. 295-336.

[31] Powell,W.W., D.R. White, K.W. Koput and J. Owen-Smith (2005) “Network Dynamics and Field Evolution: The Growth of Inter-organizational Collaboration in the Life Sciences” *American Journal of Sociology*, 110, 1132-1205.

[32] Prevezer, M. and P. Swann (1996) “A Comparison of the Dynamics of Industrial Clustering in Computing and Biotechnology” *Research Policy*, 25, 1139-1157.

[33] Riccaboni, M. and F. Pammolli (2002) “On Firm Growth in Networks” *Research Policy*, 31, 1405-1416.

[34] Roller, L-H, R. Siebert and M. Tombak (2007) “Why firms form (or do not form) RJVs” *Economic Journal* 117, 1122-1144.

[35] Rowley,T. D. Behrens and D. Krackhardt (2000) “Redundant governance structures: An analysis of structural and relational embeddedness in the steel and semiconductor industries” *Strategic Management Journal*, 21, 369-386.

[36] Saxenian, A. (1991) “The origins and dynamics of production networks in Silicon Valley” *Research Policy*, 20, 423-437.

[37] Schoenmakers W. and G. Duysters (2006) “Learning in strategic technology alliances” *Technology analysis and strategic management,* 18, 245-264.

[38] Uzzi, B. (1997) “Social Structure and Competition in Inter-firm Networks: The Paradox of Embeddedness” *Administrative Science Quarterly*, 42, 35-67.

[39] von Hippel, E. (1998) *The sources of innovation* Oxford: Oxford University Press.

[40] Walker, G., B. Kogut and W. Shan (1997) "Social Capital, Structural Holes and the Formation of an Industry Network" *Organization Science*, 8, 108-125.

[41] Watts, D.J. (1999) "Networks, dynamics and the small world phenomenon" *American Journal of Sociology*, 105, 493-527.

[42] Watts, D.J. and S. Strogatz (1998) "Collective Dynamics of Small-World Networks" *Nature*, 393, 440-442.

4 Appendix

Properties of the degree distribution from Equations ?? and ??.

1. Degree distribution:
   The degree distribution is equivalent to the unconditional probability that i finds l partners:
   The unconditional probability obtains directly as $\Pr\{n_i^g = l\} = \sum_{a=\delta}^{w} \Pr\{n_i^g = l | a_i = a\} \Pr\{a_i = a\}$ when $l > 0$ and $\Pr\{n_i^g = 0\} = \Pr\{a_i < \delta\} + \sum_{a=\delta}^{w} \Pr\{n_i^g = 0 | a_i = a\} \Pr\{a_i = a\} = 1 - \sum_{l>0} \Pr\{n_i^g = l\}$. From this we can write:
   $$\Pr\{n_i^g = l\} = \begin{cases} \binom{n-1}{l}/2^w \cdot \sum_{a=\delta}^{w} \left(\binom{a}{\delta}/2^a\right)^l \left(1 - \binom{a}{\delta}/2^a\right)^{n-1-l} \binom{w}{a}, & l > 0, \\ 1 - \sum_{l>0} \Pr\{n_i^g = l\}, & l = 0. \end{cases} \tag{7}$$

2. Expected degree:
   The unconditional probability that i finds l partners is written in equation **??** above. From this, the mathematical expectation of the degree distribution can be directly computed, as $E(n_i^g) = \sum_{l=0}^{n-1} l \cdot \Pr\{n_i^g = l\}$. This is expanded to
   $$E(n_i^g) = \sum_{l=1}^{n-1} l \left\{ \binom{n-1}{l}/2^w \cdot \sum_{a=\delta}^{w} \left(\binom{a}{\delta}/2^a\right)^l \left(1 - \binom{a}{\delta}/2^a\right)^{n-1-l} \binom{w}{a} \right\}.$$

3. Covariance between degree and knowledgeability:
   Observe that $\operatorname{cov}(a_i, n_i^g) = E(a_i n_i^g) - E(a_i) E(n_i^g)$. Using the conditional expectation we see that
   $$E(a_i n_i^g) = E(E(a_i n_i^g | a_i = a)) = (n-1)/2^w \cdot \sum_{a=\delta}^{w} \binom{w}{a} \binom{a}{\delta} a/2^a,$$
   Substitution gives the result:
   $$\begin{aligned} \operatorname{cov}(a_i, n_i^g) &= (n-1)/2^w \cdot \sum_{a=\delta}^{w} \binom{w}{a}\binom{a}{\delta} a/2^a \\ &\quad - w/2 \sum_{l=1}^{n-1} l \left\{ \binom{n-1}{l}/2^w \cdot \sum_{a=\delta}^{w} \left(\binom{a}{\delta}/2^a\right)^l \left(1 - \binom{a}{\delta}/2^a\right)^{n-1-l} \binom{w}{a} \right\}. \end{aligned}$$

Derivation of the ratio of triangle-closing to random links:

The condition that ij and ik exist, with the additional condition that $B = b$, together fully determine the first $2\delta - b$ positions (given re-labelling), and demand that the strings 111, 110 and 011 do not appear in the rest of the vector. Thus the link jk forms, conditional on $\{B = b, ij, ik\}$, if and only if the remaining $w - 2\delta + b$ positions (the empty box in table **??**) contain exactly $\delta - b$ instances of the pattern 1 0 1 and any combination of patterns 1 0 0, 0 1 0, 0 0 1, 0 0 0 in the remaining $w - 3\delta + 2b$ positions. Thus

$$\Pr\{jk|B = b, ij, ik\} = \binom{w - 2\delta + b}{\delta - b} (1/5)^{\delta - b} (4/5)^{w - 3\delta + 2b}. \tag{8}$$

We now "uncondition" on B, using $\Pr\{jk|ij, ik\} = \sum_{b=\beta}^{\delta} \Pr\{jk|B = b, ij, ik\} \Pr\{B = b|ij, ik\}$, for which $\Pr\{B = b|ij, ik\}$, the probability of having b instances of pattern 111 conditional on ij and ik existing, is needed. In the sum, the limits on b are determined by the condition that $w - 3\delta + 2b > 0$, so $\beta = \max(0, (3\delta - w)/2)$. It is obtained using Bayes' formula, which is written

$$\Pr\{B = b|ij, ik\} = \frac{\Pr\{ij, ik|B = b\} \Pr\{B = b\}}{\sum_{b'=\beta'}^{\delta} \Pr\{ij, ik|B = b'\} \Pr\{B = b'\}},$$

with $\Pr\{B = b'\} = \binom{w}{b'} (1/8)^{b'} (7/8)^{w-b'}$ for any $\beta' \leq b' \leq w$, with $\beta' = \max(0, 2\delta - w)$, and $\Pr\{B = b'\} = 0$ otherwise. Finally $\Pr\{ij, ik|B = b'\}$ must be computed (note that in the sum at the denominator, $\Pr\{ij, ik|B = b'\} = 0$ if $b' > \delta$). To get $\{ij, ik|B = b'\}$, firms' portfolios must be such that, over $w - b'$ positions, there must be $\delta - b'$ times pattern 1 1 0, $\delta - b'$ times pattern 0 1 1, and in the remaining positions any other pattern exclusive of 1 1 1. Thus

$$\Pr\{ij, ik|B = b'\} = \binom{w - b'}{\delta - b'} \binom{w - \delta}{\delta - b'} \left(\frac{1}{7}\right)^{\delta - b'} \left(\frac{1}{7}\right)^{\delta - b'} \left(\frac{5}{7}\right)^{w - 2\delta + b'}. \tag{9}$$

Then

$$\Pr\{B = b|ij, ik\} = \frac{\binom{w}{b} \binom{w-b}{\delta-b} \binom{w-\delta}{\delta-b} 8^{-w} 5^{w-2\delta+b}}{\sum_{b'=\beta'}^{\delta} \left(\binom{w}{b'} \binom{w-b'}{\delta-b'} \binom{w-\delta}{\delta-b'} 8^{-w} 5^{w-2\delta+b'}\right)}. \tag{10}$$

Using the fact that $\Pr\{jk|ij, ik\} = \sum_{b=\beta}^{\delta} \Pr\{jk|B = b, ij, ik\} \Pr\{B = b|ij, ik\}$, one finally obtains

$$\Pr\{jk|ij, ik\} = \frac{\sum_{b=\beta}^{\delta} \binom{w-2\delta+b}{\delta-b} \binom{w}{b} \binom{w-b}{\delta-b} \binom{w-\delta}{\delta-b} 2^{-6\delta+4b-w}}{\sum_{b'=\beta'}^{\delta} \left(\binom{w}{b'} \binom{w-b'}{\delta-b'} \binom{w-\delta}{\delta-b'} 8^{-w} 5^{w-2\delta+b'}\right)}. \tag{11}$$

The aim of this calculation was to derive $\Pr\{jk|ij, ik\}/\Pr\{jk\}$, the ratio of a triangle closing to a random tie, which is written

$$\frac{\Pr\{jk|ij, ik\}}{\Pr\{jk\}} = \frac{\sum_{b=\beta}^{\delta} \binom{w-2\delta+b}{\delta-b} \binom{w}{b} \binom{w-b}{\delta-b} \binom{w-\delta}{\delta-b} 2^{-6\delta+4b-w}}{4^{-w} 3^{w-\delta} \sum_{b'=\beta'}^{\delta} \binom{w}{\delta} \binom{w}{b'} \binom{w-b'}{\delta-b'} \binom{w-\delta}{\delta-b'} 8^{-w} 5^{w-2\delta+b'}}. \tag{12}$$