A quantitative approach of innovation

Anne Plunket
Professeur d'économie
anne.plunket@universite-paris-saclay.fr
Université Paris-Sud, Université Paris Saclay

1er semestre

Master IREN - Industries de Réseau et Economie Numérique

Chapter 4. Networks and innovation

1. What are social networks ?

1.1. Why study networks ?

- **Society and nature are structured as networks**
  - \- brains (e.g. neural networks), organizations (e.g., who reports to whom, who benefits from relations with whom), economies (e.g. who sells to whom), ecologies (e.g. who eats whom), medicine (e.g. how diseases spread) ...
- The fact that we are all **connected, constrained and influenced** by our fellow humains in this world has a number of implications...
  - Our choices and decision-making may be influenced by others or influence others
  - Don't we make rational decisions on whom to contact and whom to get acquainted to ?
- **The role of social networks** $\Rightarrow$ **The power of networks**

1.1. Why study networks ?

- **The role of social networks** ⇒ **The power of networks**
  - **In politics** - « New York Times », Dec. 17,2018, « Some of the popular images and themes the russians posted on social media » by Scott Shane
    When Russia targets Americans on social media, it has political goals : in 2016, to damage Hillary Clinton and help elect Donald J. Trump ; ...

    With the message:

    > “At least 50,000 homeless veterans are starving dying in the streets, but liberals want to invite 620,000 refugees and settle them among us. We have to take care of our own citizens, and it must be the primary goal for our politicians!”

    640,390 ‘Shares’

  - **In politics again** - « New York Times », Dec. 14, 2018 « After Yellow Vests Come Off, Activists in France Use Facebook to Protest and Plan ».
  - **In Economics and innovation** : Technology and research networks facilitate the diffusion of knowledge and innovation / Open innovation networks to master the complexity of knowledge

1.1. What is a network ?

- A network is a ***set of nodes*** or actors, also called *vertices* with ***connections*** between them called edges, ties or ***links***.
  - Networks are a way of thinking about ***social systems*** that focus attention on the ***relationships among the entities*** making up the system
  - How highly connected systems operate ?
  - A system taking the form of a network can be represented as a **graph**

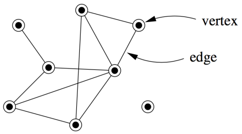


Fig. 1.1 *A small example network with eight vertices and ten edges.*

  - The nodes in a network can be ***individuals***, but it can also be collectivities such as teams, ***firms***, cities, countries or whole species.
- **Nodes have attributes**, that is, characteristics that can be categorical traits (e.g., male, female, location) or quantitative attributes (e.g., age, revenu.)

1.3. Critical issues

- Critical issues studied ?
  - How social network structures impact behavior, performance ?
  - Which network structures are likely to emerge in a society ?
- In order to answer these questions we can consider :
  - **centrality** - which individuals are best connected to others or have most **influence**
  - **connectivity** - whether and how individuals are connected to one another through the network
- Whether actors are collectivities or individuals should not be confused with levels of analysis
  - **The node level** : do individuals with more links have more power ?
  - **The dyad level** : pairwise relations between actors - do my collaborators tend to collaborate among each other ?
  - **The network level** : do well-connected networks tend to diffuse ideas faster ?

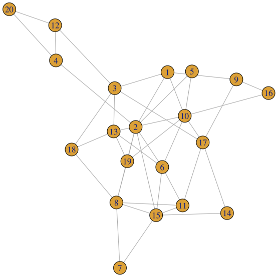
1.3. Critical issues

- To understand networks we need to
  - Find **statistical properties** that characterize the structure and behavior of networks
    - e.g. properties such as **path lengths** (= distance from one node to an other) and **degree** (= # of direct connections) distributions
  - Create **models of networks** to understand the meaning of these properties :
    - their emergence and how they interact (r.g.,- homophily - preferential attachment)
  - **Predict the behavior** of networks on the basis of structural proporties

2. What are the properties of networks ?

1. A random graph as a benchmark

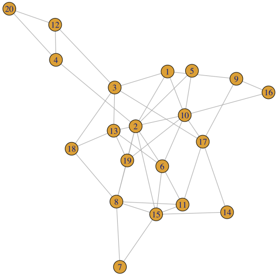
- Random-graphs as **benchmarks** to study real-world networks (Jackson, 2008)
- The links in **a random graph are formed given a completely random process** (Erdös and Rényi) - each link is formed with a given probability p and the formation is independent across links



This is a graph with 20 nodes formed with probability of $p = 0.2$

1.1. Two statistical properties : degree and degree distribution

1. **The degree** $d_i(g)$ of node i in network g is the number of direct links that a node has.



library(igraph)
g = erdos.renyi.game(20, 0.2)
plot(g)
degree(g,v = 3)=5

1.2. Two statistical properties : degree and degree distribution

2. **The degree distribution** of a network is the relative frequencies of nodes that have different degrees.
   - $P(d)$ is the fraction of nodes that have degree d under a degree distribution P.
   - The degree distribution of a random networks is a binomial distribution or a Poisson distribution

   Example : $n = 50$ nodes with an expected degree of 1 for each node (= probability 0.02 to form a node) - number of isolated nodes and the frequency distribution of degrees is a poisson approximation (Jackson, 2008, MIT Press, Social and economic networks) .

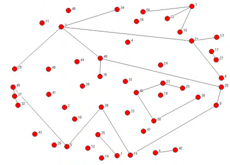


Figure 1.2.3. A Randomly Generated Network with Probability .02 on each Link

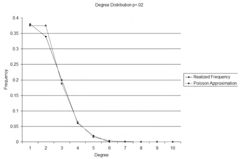


Figure 1.2.3 Frequency Distribution of a Randomly Generated Network and the Poisson Approximation for a Probability of .02 on each Link

g = erdos.renyi.game(20, 0.02)
degree.distribution(g) // Degree distribution is the cumulative frequency of nodes with a given degree

2. « small-world networks »

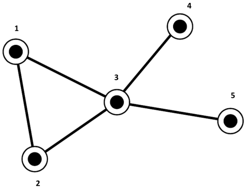
- Real-world networks are different from **random graphs**
- Real-world networks behave as **« small worlds »**
  - Experiment by Stanley Milgram in 1960s : quantify the typical distance between actors in a social networks = measure the typical geodesic distance
  - letters passed from person to person were able to reach a designated target individual in only a small number of steps - *six degrees of separation* = average length
    - 1/4 of the letters reached their targets with a median number of links of 5 and the maximum of 12
  - Demonstration of the small-world effect = most pairs of vertices in most networks seem to be connected by a short path through the network

- Following Watts and Strogatz (1998), **a small world network is a network with**
  1. small average degrees
  2. high clustering
  3. small average distances

2. « small-world networks »

1. **Degree distribution** - the average degree is very small and there is large variation in the degrees
   - Skewed distribution : While the majority have a small number of links, some have a high number of connections (World Wide Web (average = 4 links versus 1000 links (Albert, Jeong and Barabasi, 1999), Economists networks (average = 2 co-authors vs 50 co-authors)) - why ?

2. **Clustering** - significant level of clustering
   - If vertex A is connected to B and B is connected to C, then there is a high probability that A will also be connected to C
     What fraction of my friends are friends of each other ?



2. « small-world networks »

Positive aspects

Clustering (= cohesion) increases the **information transmission capacity** in a network

- **Speed and likelihood of information access** : dense connectivity of individual clusters ensures that information introduced into a cluster will quickly reach other firms in the cluster
- Make information exchange **meaningful and useful**. The internal density can increase the dissemination of alternative interpretations of problems and their potential solutions ; and facilitate collective understanding and stimulates collective problem solving = cf **community of practices** (Brown and Duguid, 1991).
  - Make firms more willing and able to exchange information (Ahuja, 2000).
  - Coleman (1988) and Granovetter (1992) suggest that densely clustered networks give rise to trust, reciprocity norms which leads to a high level of **cooperation** and creates self-enforcing **informal governance** mechanisms. It facilitates the transfer of tacit, embedded knowledge (Zander and Kogut, 1995).

2. « small-world networks »

Risks

- **Yet Dense clusters provide many redundant paths** to the same actors and sources of information and knowledge. Cohesion can lead to norms of adhering to established standards and conventions, which can potentially stifle experimentation and creativity (Uzzi and Spiro, 2005). **This limits innovation.**
  - Uzzi and Spiro (2005) studying the structure of artist's collaboration network in Broadway found that large-scale structures influence their creativity and financial/artistic performance

2. « small-world networks »

Measure

- **Transitivity or clustering coefficient** for any node i in a network g is

  weighted overall clustering coefficient measure (Borgatti, et al. 2002 ; Newman et al. 2002)

$$\text{Clustering}_w = \frac{3x(\text{number of triangles in the graph})}{(\text{number of connected triples})}$$

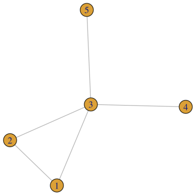
  - The numerator is the number of pairs of individuals with a common acquaintance (a third individuals they are both linked to).
  - The denominator is the total number of triads, in which at least two links exist.
  - where a triangle is a set of three nodes (e.g. i,j,k), each of which is connected to both of the others and a « connected triple » is a set of three nodes in which at least one is connected to both the others (e.g. i is connected to j and k but j and k need not be connected)

- This measure indicates the proportion of triples for which transitivity holds (i is connected to j and k then by transitivity j and k are connected)
- The factor of three in the numerator ensures that the measure lies strictly in the range of zero and one because each triangle implies three connected triples.

2. « small-world networks »

Measure

```
library(igraph)
g <- rbind(c(1,2), c(1,3), c(2,3), c(3,4), c(3,5) )
g <- graph.data.frame(g, directed=FALSE)
plot(g)
transitivity(g, type=c("globalundirected"))
transitivity(g, type=c("localundirected"))
```



2. « small-world networks »

3. **Average distances are very small.**
   - **Geodesic distance** $d(i,j)$ between two nodes i and j in network g is the **length of the shortest path** between them
   - Exemple : $d(1,5) = 2$
   - Despite low overall density of ties, actors in a network are linked with each other through a relatively small number of intermediaries.

In sum : Small worlds are

- *Sparsely connected* with each actor having few ties relative to the number of actors in the industry
- *Locally clustered* into cliques within which partners of partners are also frequently partners.
- *Dense cliques of actors* are spanned by relationships that cut across the cliques, acting as conduits for information and control.

2. « small-world networks »

Empirics

- Small worlds have been used to study **collaboration networks** between firms and individuals (Baum, Shipilov, Rowley, 2003 ; Fleming, King, Juda, 2007)
- There is considerable **uncertainty associated with establishing collaborations** which stems from the challenge associated with obtaining information about the competences, needs, intentions and resources of potential partners.

⇒ Small world structures are interesting because they have great efficiency in moving information, innovations, experience ... that enable organizational learning, adaptation and competitive advantage.

- Imperfect information can raise search costs and the risk of exposure to opportunistic behavior

3. Formation of networks : dyadic ties, evolution and performance

How do networks emerge and what is the impact on performance ?

3. Formation of networks : dyadic ties, evolution and performance

- **How social networks emerge and evolve over time ?**
  - Focus on research that explicates the existence, creation, persistence, and dissolution of social relationships among social actors and their implications for the evolution of networks
- Most theories on networks are based on **sociology** and predictions apply to the **impact of networks of individuals and organisations on innovation or creativity** in management and economics.
  - **Embeddedness in social networks** is viewed as an explanation of the actors' behavior and achievement.
- Based on the way **dyadic ties** are formed, persist and dissolve, the theory emphasizes **three mechanisms**
  - Assortative mechanisms
  - Relational mechanisms
  - Proximity mechanisms

1. Assortative mechanisms

- **Assortative mechanisms are based on actors' attributes** as social relationships rely on the compatibility and complementarity of actors' attributes.
  - Dyadic associations between actors depend on similarities or dissimilarities of actors attributes and their consequent propensity to form connections
- **The role of homophily**
  - = The greater the similarity between two individuals the more likely they are to establish a connection.
  - Similarities in the human attributes of potential alters, such as age, gender, religion, ethnicity, values, intelligence, and education characterize friendships and marriage
- **The role of Heterophily**
  - Collaborations among diverse people - the diversity of social ties can be seen in coauthorship (Moody, 2004) and alliance networks formed with dissimilar collaborators (Powell, Koput, Owen-Smith, 2005)
    - Powell et al. (2005) examined the determinants of attachment among universities, venture capital firms, public research organizations, and large pharmaceutical firms during the first ten years of the biotechnology field
    - Compared with interpersonal ties, interorganizational links may be more prone to heterophily

2. Relational mechanisms

Relational versus structural embeddedness

- **Relational mechanisms** originate form information about **actors' relationships**
  - Determine the shape and structure of the network on the formation of new connections (Stuart and Sorenson, 2007)

⇒ Two types of network embeddedness :

- **Relational embeddedness** focuses on **the quality of dyadic ties** between firms (Gulati and Gargiulo, 1999).
  - Reduce search costs and alleviate risks of opportunism : prior direct ties provide channels through which each partner can learn about the competences and reliability of the other.
- **Structural embeddedness** focuses on the role of **third parties** (Granovetter, 1985)
  - People do not just have relationships with each other but also with the same third parties
  - ⇒ Shift from direct ties to indirect channels of information, reputation
  - Third party ties can promote good behavior by fostering a concern for local reputation and trustworthiness
  - Role of prior direct and indirect connection linking individuals on future connections

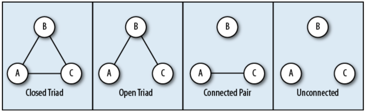
2. Relational mechanisms

- **The role of repetition**
  - Tendency of repeated ties : over time collaborators show a significant propensity to work with persons with whom they have worked in the past (Uzzi and Spiro, 2005) :
  - Repeated ties are correlated with/indicators of the **strength of a relationship**, trust, reduction of coordination costs and facilitator of information exchange.

- **Triad and network closure** : the proposition that actors separated by one intermediary are the most likely to become connected in subsequent (= closure mechanism) time periods

  When interfirm ties are formed in response to relational embeddeness, it increases the *strength* of ties and *the closure* in the network

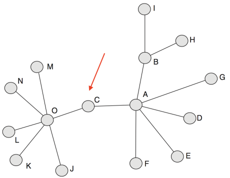
  = Social networks are characterized by clustering : they tend to have a **high density of closed triads** = people tend to become friends with the friends of their friends (e.g. inventors, see Fleming et al. 2007)



2. Relational mechanisms

Dynamics of degree and centrality

- Individuals tend to be at the **center** of their network (stars) while others remain on the **periphery or are isolates** [⇒] **Network centrality**
  - Different aspects of an actor's connectedness and status include betweenness, closeness, degree, prestige, attractiveness, rank and brokerage (Borgatti and Everett 2006).
  - **Betweenness centrality** calculates how many times an actor sits on the geodesic (i.e. the shortest path) linking two other actors together ⇒ It measures how much **potential control** an actor has over the flow of information



2. Relational mechanisms

Dynamics of degree and centrality

- Freeman, 1979 mesure :

$$\text{Betweeness Centrality} = \sum_{j<k} g_{jk}(n_i)/g_{jk}$$

  - where $g_{jk}(n_i)$ refers to the number (n) of geodesics (= shortest path) linking fims j and k that contain focal firm i. The term $g_{jk}(n_i)/g_{jk}$ captures the probability that firm i is involved in the shortest path between j and k. Bet. centr. is the sum of these probabilities over all pairs of firms.
- **Degree centrality** - the number of direct ties an actor possesses - has an impact on the evolution of the actor's social network :
  - Skewed degree distribution : most actors have only a few ties, while a small number have extraordinarily many (Borgatti and Everett 1999, Albert and Barabasi 2002).
- **Positive association between an actor's central position** in the network and **goal achievement including creativity** (Burt, 2004 ; Uzzi and Spiro, 2005), rate of alliance formation (Powell et al. 2005), political influence (Fernandez and Gould, 1994)

2. Relational mechanisms

Dynamics of degree

- The role of **preferential attachment** purportedly occurs because actors looking for new connections use an actor's degree as a proxy for his or her fitness.
  - Interoganizational collaboration networks (Kogut et al. 2007)
  - Not only are high-degree actors more likely to form attachments than noncentral actors, but
  - they are also more likely to form attachments with each other—popular actors attach to other popular actors, whereas lower-degree actors tend to attach to other low-degree actors.
- **Not always true** : Powell et al.'s (2005) analysis of the interorganizational collaboration networks found that in many cases firms exhibited a preference for novelty over preferential attachment—with well-established, highly connected firms collaborating with younger, less-connected organizations

3. Proximity mechanisms

- Proximity mechanisms explains networks and network change at the level of actors' social and cultural environments.
  - Hypothesis : interaction increases with proximity such as geographic/physical propinquity
- Proximity :
  - favors social attachment
  - influences the persistence of social ties by moderating the effort required to maintain relationships
- **Results** show that social interaction is organized around social, legal and physical entities around which joint activities are organized (e.g. workplaces, clubs, groups, associations) = it is not physical geography alone

Empirics : a dyadic approach

Empirics

Empirical analysis of networks, proximity and innovation : a dyadic approach

Cassi L. and A. Plunket, (2014), Proximity, network formation and inventive performance : in search of the proximity paradox, Annals of Regional Science, 53(2), 395-422

Empirics : a dyadic approach

Empirics

- Aim : assess the role of proximity on the formation and performance of inventive collaborations
  - **Question 1.** : Is proximity a "facilitator" of collaboration and knowledge diffusion ?
  - **Question 2.** : Does too much proximity harm innovation ?
- Data : 12068 patents in genomics, 4406 applicants and 24708 inventors with an EU address and co-inventors from 1990-2006
- Unit of analysis : dyads of inventors - 1988 dyads based on 906 patents
- Theory : explore the respective impact of geographical proximity in addition to social, technological and organizational proximity
- **Research strategy** :
  - Considering a network of inventors and studying their links (dyads) given their proximity and network position
  - **Determinants of collaboration** : Rare event logit : co-inventor tie formation = f(proximity and network variables, year dummies)
  - **Determinants of inventive performance** : negative binomial forward patent citations = f(proximity and network variables, year dummies)

Empirics : a dyadic approach

Review of the literature

- **Geographical, social proximity and organizational proximity are coordination mechanisms and substitutes**
  - **Geographically proximity** and networks : Organizational, social and geographical proximity are substitutes (Boschma, 2005 ; Cassi & Plunket, 2015)
  - **Social proximity within network** : strong cohesive ties - end-up collaborating with your partners' partners
    - The position of actors in the network (Coleman, 1988 ; Burt, 1992 ; Fleming et al. 2007, Baum et al., 2012. . .)
    - **Closure positions** : Strong cohesive ties (Coleman, 1988)
      Share social proximity - Promote trust and collaboration / Redundant ties : similar knowledge bases and technological skills
  - **Organizational proximity** : because transaction costs are greatly reduced, the likelihood of collaborating is increased when two individuals work or patent for the same organization.

Empirics : a dyadic approach

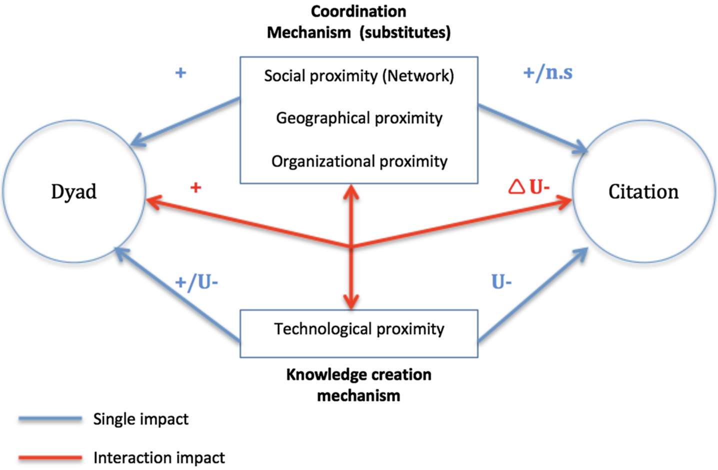


Fig. 1 Hypothesis testing

Empirics : a dyadic approach

Review of the literature

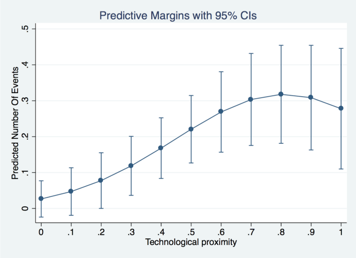
- **Technological proximity** = knowledge creation mechanism
  Optimal level of cognitive proximity : inverted U shape (Mowery, et al. 1998 ; Nootebom et al. 2007)
  Technological proximity is computed as an uncentered correlation of two vectors f_{ik} and f_{jk} representing each inventor i and jtechnological position defined in terms of k IPC 4 digit—International Patent Classification—classes (Jaffe 1989) The index is ranging from zero to one, depending on the degree of overlap between the co-inventors' prior patent IPC codes

$$T_{ij} = \frac{\sum_{k=1}^{K} f_{ik} f_{jk}}{\sqrt{\sum_k f_{ik}^2 f_{jk}^2}}$$

  - Technological distance allows a **profitable recombination of different knowledge bases**, but actors need some proximity in order to understand each other and interact efficiently.

Empirics : a dyadic approach

Figure 3. Predicted number of forward citations



Empirics : a dyadic approach

Table 2 – Estimation of Tie formation and Citations

VARIABLES	(1a) Tie Formation ReLogit	(1b) Tie Formation ReLogit	(1c) Citations NegBin	(2a) Tie Formation ReLogit	(2b) Citations NegBin	(3a) Tie Formation ReLogit	(3b) Citations NegBin
Technological proximity	1.301+	2.030***	6.128*	1.701***	6.076*	1.676***	6.125*
	[1.75]	[13.11]	[2.10]	[9.94]	[2.09]	[9.74]	[2.11]
Technological proximity sq	0.553		-3.778+		-3.765+		-3.800+
	[1.00]		[-1.79]		[-1.79]		[-1.80]
Geographical proximity	0.592***	0.592***	-0.007	0.554***	-0.009	0.601***	-0.020
	[38.35]	[38.33]	[-0.18]	[33.01]	[-0.21]	[35.34]	[-0.30]
Closure				3.471***	0.070	1.446***	0.137
				[24.38]	[0.25]	[7.04]	[0.37]
Closure x geographical proximity						-0.491***	0.023
						[-10.40]	[0.26]
Same applicant	2.458***	2.462***	0.726*	1.674***	0.720*	1.640***	0.724*
	[20.32]	[20.37]	[2.37]	[12.12]	[2.33]	[11.79]	[2.32]
Same type	-0.046	-0.045	1.116***	-0.070	1.116***	-0.042	1.115***
	[-0.71]	[-0.69]	[3.80]	[-1.01]	[3.81]	[-0.58]	[3.80]
Degrees - Avrg	0.072	0.072	0.886**	0.233***	0.846**	0.251***	0.848**
	[1.47]	[1.48]	[3.13]	[4.12]	[2.71]	[4.39]	[2.71]
Degrees - Abs.diff.	-0.026	-0.027	-0.383*	-0.526***	-0.371*	-0.563***	-0.373*
	[-0.32]	[-0.33]	[-2.51]	[-5.84]	[-2.33]	[-6.18]	[-2.36]
Border	-1.323***	-1.323***	-0.811+	-1.196***	-0.816+	-1.198***	-0.823+
	[-13.35]	[-13.36]	[-1.78]	[-11.32]	[-1.79]	[-10.69]	[-1.80]
Originality			1.613*		1.605*		1.603*
			[1.97]		[1.97]		[1.97]
# inventors per patent			-0.572		-0.578		-0.592
			[-1.51]		[-1.53]		[-1.59]
Experience - Avrg	-0.151**	-0.151**		-0.147*		-0.153**	
	[-2.80]	[-2.81]		[-2.53]		[-2.60]	
Experience - Abs.diff	0.221**	0.218**		0.449***		0.459***	
	[2.81]	[2.78]		[4.95]		[5.01]	
Stock - Abs.diff	-0.124*	-0.125*		-0.182**		-0.176**	
	[-2.06]	[-2.08]		[-2.76]		[-2.64]	
Stock - Avrg	-0.096	-0.097		-0.153		-0.142	
	[-0.81]	[-0.82]		[-1.11]		[-1.04]	
Constant	-4.224***	-4.431***	-7.246***	-4.261***	-7.182***	-4.041***	-7.205***
	[-14.47]	[-22.36]	[-4.14]	[-18.87]	[-4.18]	[-17.98]	[-4.14]
Observations	22,854	22,854	1,988	22,854	1,988	22,854	1,988
Log Likelihood	-3831.02	-3831.48	-855.5	-3359.7	-855.5	-3306.28	-855.4
D.F.	22	21	20	22	21	23	22
Chi2			64.09		64.58	.	65.44
Alpha (overdispersion test)			2.576***		2.576***		2.575***
			[12.42]		[12.41]		[12.37]

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1
Note: Tie formation are estimated through Rare Event logistic with robust standard errors in parenthesis (clustered over inventors and controls)
Citations are estimated with a negative binomial model with robust standard errors in parenthesis (clustered over patents)
Regressions include year dummies
Except for categorical variables, all variables are in logs.

Empirics : a dyadic approach

Results

- Geographical proximity and network position per se do influence tie formation but do not influence technological performance
- Proximity paradox is partly supported and affects only technology
- Technological proximity has an inverted u-shape suggesting the existence of an optimal level
- Limitations : only one industry (genomics) ; only one type of connection (= patent coauthorship) ; other types of social ties are ignored due to a lack of data

4. Network structure and firm performance

Does the structure of networks have an impact on firm performance ?

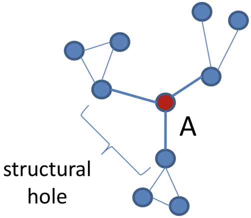
4.1. Firms' network strategy

The role of open versus closed network positions

- **Should firms occupy densely interconnected « closed » network positions ?**
  - **Coleman's (1988) closure** argument : firms are better off occupying densely interconnected, closed network positions in which their partners are also partners
  - It provides **coordination and integration** benefits by facilitating the ease of exchange and commonness of information among firms (Coleman, 1988)
  - **Favors exploitation** which involves using and refining existing knowledge to improve organizational functioning by reducing variability in the quality or efficiency of current strategies, competencies, technologies, and procedures. (Baum, Cowan and Jonard, 2014)
  - The solution space is well defined, and **search is local and highly specific**.

4.1. Firms' network strategy

- **Should firms occupy sparsely connected « open » network positions ?**
  - It confers access and control benefits through conveyance of **diverse information and resources** as well as **brokerage opportunities** (Burt, 1992)
  - **Burt's (1988) structural hole** argument : firms embed themselves in sparsely connected, **open network positions** comprised of **disconnected** partners, nonredundant partners



  - A structural hole is a lack of connection between two nodes that is bridged by a **broker**
  - **Favors exploration**

4.1. Firms' network strategy

- **Structural hole position is ideal**
  - for gaining access to diverse sources of information and knowledge, facilitating identification of emerging opportunities and threats, alternative future options, and the location of complementary knowledge (Mitsuhashi, 2003 ; Powell et al ., 1996).
  - for experimentation to identify emerging innovations and alternative future options, new ways of doing things and new things to do.
  - The **solution space is ill-defined, search is wide**, and a premium is placed on newer, more diverse information. (Baum, Cowan and Jonard, 2014)

4.2. Empirics : Firms' network embeddeness and performance

Empirical analysis of networks and innovation

Schilling M. and Phelps C. (2007), Interfirm Collaboration Networks : The impact of large-scale network structure on firm innovation, Management Science, 53 (7), 1113 - 1126

4.2. Empirics : Firms' network embeddeness and performance

Two questions

- **Does the structure of an industry-level interfirm network influence the rate of knowledge creation among firms in the network ?**
- **What structural properties will enhance firm innovation ?**
  - The structure of alliance networks influences their potential for knowledge creation.
  - Dense local clustering provides information transmission capacity in the network by fostering communication and cooperation but what about nonredundant connections ?
- Forming alliances is costly and constrained : trade-off between forming dense clusters to facilitate rapid exchange and integration of knowledge, versus forging links to create short paths to a wider range of firms.

4.2. Empirics : Firms' network embeddeness and performance

- When dense clusters are **sparsely connected** to each other, they enable to create and preserve the requisite **variety of knowledge** in the broader network
- The **diversity of knowledge** distributed across clusters in the network provides the requisite **variety for novel recombination**.
- **Two key large-scale network properties** are examined and their impact on innovative output of members of the network.
  - **Clustering** = the dense connectivity of clusters creates transmission capacity in a network (Burt, 2001) enabling large amounts of information to rapidly diffuse
  - **Reach** = the shortest path lengths to a wide range of firms. It ensures that diverse information sources can be tapped.
    Reach = direct and indirect relationships
  - **Hypothesis** : networks with high clustering and high reach will significantly enhance the creative output of member firms.

4.2. Empirics : Firms' network embeddeness and performance

Reachability

Variables

- Firm's **clustering** coefficient = the proportion of its partners that are themselves directly linked to each other.
  - The clustering coefficient for the overall network is the average of this measure across all firms in the network
- **Firm's reachability :** The more firms that can be reached by any path from a given firm, the more knowledge that firm can potentially access
  - The likelihood, speed and integrity of knowledge transfer between firms directly depend on the path length separating those two firms
  - The diffusion of information and knowledge is more rapid within short average path lengths (Watts, 1999) : reach more information, quickly and with less information distorsion
    - **Average path length** = average number of links that separates each pair of actors/firms in the network

4.2. Empirics : Firms' network embeddeness and performance

Reachability

- **Reachability** depends on the distance between any two firms = it depends on the geodesic distance between them (Borgatti et al. 2002)

$$\text{Reachability}_{ij} = \sum_j \frac{1}{d_{ij}}$$

d_{ij} = minimum distance (geodesic distance) from i to j

- A network's average distance-weighted reach is this measure averaged across all firms in the network

$$\text{Network reachability} = \left(\sum_n \sum_j \frac{1}{d_{ij}} \right) / n$$

n is the number of nodes (= firms) in the network.

- The network's average reachability is a meaningful measure of the **overall size and connectivity** of a network.

4.2. Empirics : Firms' network embeddeness and performance

- **Data** : Panel of US firms that are part of alliance networks of 11 high-technology manufacturing industries : aerospace equipment, automotive bodies and parts, chemicals, computer and office equipment, household audiovisual equipment, pharmaceuticals...
- Network of alliances based on publicly reported strategic alliances
- 1106 firms involved in 3663 alliances forming 5306 dyads
- **Dependent variable** : number of granted patent applications for firm i in year t (aggregating subsidiary patents)
- **Independent variables** :
  - Clustering coefficient
  - Reachability
  - Clustering x Reach = predict that the combination of clustering and reach will have a positive impact on member firm innovation

4.2. Empirics : Firms' network embeddeness and performance

- Model specification : Patents is a count variable. Negative binomial (= generalization of poisson model)
  Regression at the firm, industry and time level with year, firm and industry FE

$$
\begin{aligned}
&Patents_{it+1(2,3)} \\
&= f(Clustering_{jt}, Reach_{jt}, Clustering * Reach_{jt}, \\
&\quad R\&D\ Alliance\ \%_{jt}, R\&D\ Intensity_{jt}, Centrality_{it}, \\
&\quad Local\ Efficiency_{it}, Centralization_{jt}, Density_{jt}, \\
&\quad Presample\_Patents_{it}, Automotive, Chemicals, \\
&\quad Computers, Audiovisual, Medical, Petroleum, \\
&\quad Pharmaceuticals, Semiconductors,
\end{aligned}
$$

Table 1 Network Size and Component Structure, Averages over 1992–2000

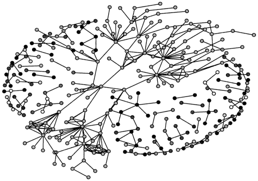
Industry	Average number of firms from industry in alliances[a]	Average number of alliances per firm	Average network size (nodes)[b]	Percent in main component (%)
Aerospace	9	3.05	28	46
Automotive	15.67	3.43	53.2	37
Chemicals	45.17	2.97	199.8	11
Computers and office equipment	79.67	4.48	347	45
Household audiovisual equipment	9	1.5	28.3	10
Measuring and controlling	22.67	1.96	48.33	21
Medical equipment	66.17	1.66	172.33	7
Petroleum refining and products	5.3	2.65	24.83	18
Pharmaceuticals	218.33	2.54	510	64
Semiconductors	58.67	3.51	204	55
Telecommunication equipment	44.83	6.53	266.33	54

[a]This number includes only those firms with the designated primary SICs; it does not include partners in the network that are not in those SICs.
[b]Includes all U.S. firms in the network, including both those with the designated primary SICs and their alters, regardless of SIC.

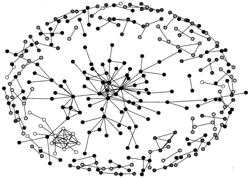
4.2. Empirics : Firms' network embeddeness and performance

Figure 1 Network Size and Component Structure (Common Shade of Gray Indicates Firms in Same Component)

Computers, 1996



Computers, 1997



4.2. Empirics : Firms' network embeddeness and performance

Table 2 **Panel Negative Binomial Regression Models with Fixed and Random Effects** ($N = 1{,}106$; $Obs = 3{,}444$)

	$Patents_{it+1}$			$Patents_{it+2}$			$Patents_{it+3}$		
	1	2	3	4	5	6	7	8	9
Fixed effects									
Constant	1.136** (0.354)	0.582 (0.359)	0.604 (0.360)	1.257** (0.327)	1.663** (0.333)	1.614** (0.324)	1.433** (0.337)	1.859** (0.369)	1.825** (0.368)
Presample Patents	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)
Density	−0.248 (1.154)	−0.624 (1.358)	−0.527 (1.468)	−0.411 (1.529)	−2.220 (1.808)	−2.637 (1.843)	−2.012 (1.861)	−1.598 (2.509)	−1.674 (2.134)
Centralization	−0.014 (0.008)	−0.014 (0.008)	−0.012 (0.008)	−0.018** (0.006)	−0.016* (0.007)	−0.035** (0.006)	0.019** (0.007)	0.019** (0.007)	0.019* (0.007)
Ind. R&D Intensity	2.739 (2.668)	2.867 (2.522)	2.877 (2.581)	0.741 (2.366)	−0.088 (2.373)	−0.246 (2.327)	−7.126** (2.478)	−6.754** (2.504)	−6.754** (2.504)
R&D Alliance %	−0.112 (0.275)	0.223 (0.275)	0.222 (0.289)	0.068 (0.217)	−0.131 (0.223)	−0.188 (0.191)	−0.040 (0.248)	−0.305 (0.264)	−0.312 (0.304)
Efficiency	−0.199** (0.068)	−0.189** (0.072)	−0.190** (0.073)	−0.303** (0.091)	−0.321** (0.095)	−0.327** (0.087)	−0.267** (0.097)	−0.272** (0.089)	−0.270** (0.088)
Betweenness	0.003 (0.006)	0.003 (0.005)	0.003 (0.005)	0.005 (0.006)	0.004 (0.007)	0.002 (0.006)	−0.001 (0.009)	−0.001 (0.010)	−0.001 (0.010)
Clustering		0.420** (0.136)	0.507* (0.235)		0.346** (0.127)	−0.141 (0.196)		0.234 (0.183)	−0.319 (0.279)
Reach		0.010** (0.003)	0.011** (0.003)		−0.012** (0.003)	−0.020** (0.004)		−0.007* (0.003)	−0.009* (0.004)
Clustering × Reach			−0.015 (0.030)			0.081** (0.023)			0.014* (0.007)
Log Likelihood	−4,646.65	−4,637.32	−4,637.12	−4,597.46	−4,586.78	−4,577.98	−4,468.75	−4,464.64	4,464.46
Random effects									
Constant	1.118** (0.309)	0.542 (0.339)	0.541 (0.339)	0.984** (0.307)	1.342** (0.303)	1.256** (0.290)	0.920** (0.296)	1.333** (0.331)	1.214** (0.321)
Presample Patents	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)
Density	1.444 (0.900)	0.250 (1.092)	0.243 (1.166)	0.527 (1.197)	−1.872 (1.394)	−2.451 (1.352)	−1.454 (1.434)	−1.286 (1.618)	−1.538 (1.654)
Centralization	−0.021** (0.006)	−0.020** (0.007)	−0.021** (0.007)	−0.021** (0.006)	−0.020** (0.006)	−0.027** (0.005)	0.016* (0.006)	0.017* (0.007)	0.013* (0.006)
Ind. R&D Intensity	0.887 (2.429)	1.030 (2.408)	1.027 (2.424)	−0.357 (2.231)	−0.818 (2.151)	−0.590 (2.135)	−8.029** (2.278)	−7.987** (2.343)	−8.101** (2.460)
R&D Alliance %	0.014 (0.230)	0.383 (0.214)	0.384 (0.222)	0.208 (0.215)	−0.017 (0.187)	−0.090 (0.158)	0.106 (0.220)	−0.139 (0.233)	−0.153 (0.274)
Efficiency	−0.342** (0.062)	−0.336** (0.069)	−0.336** (0.069)	−0.396** (0.079)	−0.436** (0.081)	−0.435** (0.073)	−0.297** (0.087)	−0.307** (0.080)	−0.312** (0.078)
Betweenness	0.008 (0.005)	0.007 (0.004)	0.007 (0.005)	0.003 (0.005)	0.004 (0.005)	0.001 (0.005)	−0.000 (0.008)	−0.001 (0.008)	−0.001 (0.008)
Clustering		0.554** (0.106)	0.548** (0.212)		0.485** (0.116)	−0.101 (0.186)		0.152 (0.159)	−0.422 (0.344)
Reach		0.008** (0.003)	0.008* (0.003)		−0.013** (0.003)	−0.022** (0.003)		−0.008* (0.003)	−0.011** (0.004)
Clustering × Reach			0.001 (0.028)			0.082** (0.019)			0.043* (0.020)
a	0.707** (0.047)	0.716** (0.047)	0.710** (0.048)	0.675** (0.047)	0.684** (0.048)	0.690** (0.480)	0.650** (0.046)	0.652** (0.046)	0.652** (0.046)
b	0.358** (0.021)	0.360** (0.022)	0.360** (0.022)	0.321** (0.019)	0.328** (0.020)	0.334** (0.02)	0.291** (0.018)	0.290** (0.018)	0.293** (0.018)
Log likelihood	−8,520.70	−8,509.78	−8,509.78	−8,425.33	−8,407.95	−8,392.95	−8,198.66	8,194.98	−8,193.03

Notes. All models include firm, time period, and industry effects. Standard errors are in parentheses.
$^*p < 0.05$, $^{**}p < 0.01$ (two-tailed tests for all variables).

4.2. Empirics : Firms' network embeddeness and performance

Conclusion

- Clustering enables even globally sparse network to achieve high information transmission capacity through locally dense pockets of closely connected firms
- Reach increases the quantity and diversity of information available to firms in the network by bringing the information resources of more firms within relatively close range
- networks that have both the high information transmission capacity enabled by clustering, and the high quantity and diversity of information provided by reach, should facilitate greater innovation by firms.
- Results show :
  - Both local density (= clustering) and global efficiency (=reach) can exist simultaneously, and this combination enhances innovation
- Limitation : do not control for firm attributes that could shape the flow of knowledge, do not control for the type of knowledge transfered (tacit or rather explicit and mature, or complex...), do not consider whether collaborations last more than three years, connectivity could be biased downwards.

References

- Baum, J.A.C., Shipilov, A. V & Rowley, T.J., 2003. Where do small worlds come from ? Ind Corp Change, 12(4), pp.697–725.
- Borgatti S., Everett M.G. (2013), Johnson J. Analyzing Social Networks, Sage
- Cassi L. and A. Plunket, (2014), Proximity, network formation and inventive performance : in search of the proximity paradox, Annals of Regional Science, 53(2), 395-422
- Goyal S., Connections. An introduction to the economics of networks,Princeton University Press
- Easley D. and Kleinberg J. (2010), Networks, Crowds, and Markets : Reasoning about a Highly Connected World, Cambridge University Press http://www.cs.cornell.edu/home/kleinber/networks-book/
- Rivera, M.T., Soderstrom, S.B. & Uzzi, B., 2010. Dynamics of Dyads in Social Networks : Assortative, Relational, and Proximity Mechanisms. Annual Review of Sociology, 36(1), pp.91–115.

References

- Schilling M. and Phelps C. (2007), Interfirm Collaboration Networks : The impact of large-scale network structure on firm innovation, Management Science, 53 (7), 1113 - 1126
- M.E.J. Newman, The structure and function of complex networks, SIAM Review, 45(2), 167-256
- References for networks in R
  - `https://sna.stanford.edu/rlabs.php`
  - `https://www.sci.unich.it/~francesc/teaching/network/`
  - `https://assemblingnetwork.wordpress.com/2013/05/29/network-basics-with-r-and-igraph-part-i-of-iii/`
  - Kolaczyk E. and Csardi G. (2009), Statistical analysis of network data with R, Use R !, Springer