R&D Networks: Theory and Empirical Evidence

Mauro Napoletano

OFCE-Sciences Po,

SKEMA and Scuola Superiore Sant'Anna

mauro.napoletano@sciencespo.fr

Outline

1. **Networks in economics: a brief introduction**
2. R&D networks: a model based on knowledge recombination
3. Efficiency and Equilibrium Analysis of R&D networks
4. Simulations of network dynamics
5. Empirical R&D networks

mauro.napoletano@sciencespo.fr

Networks in Economics: a brief introduction

- Most activities that are relevant in economics (e.g. production, trade in goods or in financial assets, exchange of information or ideas) occur through the **interaction** of **heterogeneous** economic agents (households, firms).
- Furthermore, most interactions between a pair of economic agents, typically generate some *indirect effects* (either positive or negative) on some other economic agents. Economist refer to the above indirect effects with the term **externalities**
- It follows that the **structure of the network** of interactions among agents shapes the unfolding of economic activities.
- In addition, the network also determines the transmission of externalities within the economic environment, thus magnifying or dampening the consequences of economic actions and or interactions.

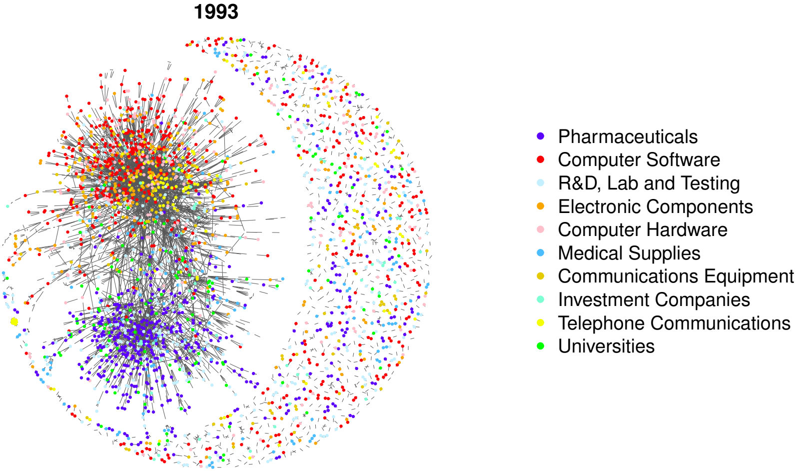
Examples of applications of graph theory in economics

- Financial Networks and Systemic Risk: see e.g. Allen and Gale (2000), Battiston et al. (2009, 2016), Acemoglu et al. (2015), Caballero and Simsek (2013), Eliott, Golub and Jackson (2014).
- Input-Output Networks: see e.g. Bak, Scheinkman and Woodford (1993), Long and Plosser (1983), Acemoglu et al. (2012, 2014).
- **R&D Networks**: see e.g. Goyal and Joshi (2003), Goyal and Moraga-Gonzales (2001), Cowan and Jonard (2003), König et al. (2011, 2012)
- For a survey or other applications see Jackson (2010), Goyal (2007), Bramoullè, Galeotti and Rogers (2017).

R&D networks

- R&D collaborations among firms (both formal and informal) have increased in importance especially in high-technology industries (e.g. pharmaceutical, computer industry).
- This is because the knowledge base used in this industry has become too large to be mastered by a single firm, and there has then been a process of division of inventive labor.
- R&D collaborations allow firms to directly combine the knowledge, skills and physical assets needed to innovate. In addition, they provide access to indirect spillovers, by serving as conduits through which knowledge and information can spread across firms (Ahuja, 2000; Powell et al., 2005).
- As a consequence, collaborations do not occur between two isolated actors, but rather involve firms that are already part of a network of partnerships with other firms.

R&D networks: an example



- The increasing importance R&D collaborations has also spurred research about the structural features of the network of R&D collaborations, and on their impact on industry innovative performance.
- Relevant questions
  1. Does the position in the network grant a firm a higher innovation rate?
     - ★ Powell, Koput and Smith-Doerr (1996), Ahuja (2000): innovativeness increases with the degree centrality of the firm in the network
  2. Which network structures are efficient?
  3. Which network structures emerge under endogenous partner selection? In addition, are these endogenous structures efficient?
- We shall address the above issues via a model of an industry where innovation is governed by knowledge sharing: innovation depends on firm i's knowledge and the knowledge of its neighbors.

Outline

1. Networks in economics: a brief introduction
2. **R&D networks: a model based on knowledge recombination**
3. Efficiency and Equilibrium Analysis of R&D networks
4. Simulations of network dynamics
5. Empirical R&D networks

The Model

- Sufficiently homogeneous knowledge base in the industry
- Two-stage game
  1. Firms form pairwise R&D collaborations with other firms in the same industry,
  2. Firms introduce innovations in the market by using the knowledge accumulated through collaborations.
- Firms exploit R&D collaborations to innovate.
- Each innovation returns a value $V > 0$ (Aghion & Howitt '98).
- A firm's knowledge growth is proportional to its own knowledge and the knowledge levels of its R&D partners (see e.g. Kogut and Zander '92, Weitzman '98) '04).

$$\dot{x}_i(t) = \gamma x_i(t) + \beta \sum_{j=1}^{n} a_{ij}(t) x_j(t), \quad \gamma > 0, \beta > 0,$$

The knowledge growth function: key properties

1. The previous equation implies that the growth of knowledge is *cumulative* (e.g. Dosi, 1988), i.e. larger knowledge stocks (of the firm of its partners) spur higher knowledge growth.
2. Knowledge growth depends both on spillovers emanating from the direct neighbors of i, as well as on spillovers from indirect neighbors which affect the stock of knowledge of direct neighbors.
3. This implies that the topology of the whole network of R&D collaborations - including all direct and indirect paths along which knowledge can flow between firms - influences knowledge dynamics within the firm.

- Returns $R(x_i(t))$ from innovations at time t are an increasing and concave function of the knowledge stock $x_i(t)$ of firm i

$$R(x_i(t)) = V \ln x_i(t), \quad V > 0.$$

- Each collaboration involves an increasing cost per unit of time $\tilde{c}_0 + \tilde{c}_1 t$ (e.g. Copeland and Fixler, 2009; Jones, 1995).
- When forming or severing a collaboration a firm evaluates its total discounted profits from collaborations, taking the network as given:

$$\tilde{\pi}_i(G_i, \tilde{\mathbf{c}}, \rho) = \int_0^{\infty} \left(V \ln x_i(t) - d_i(\tilde{c}_0 + \tilde{c}_1 t) \right) e^{-\rho t} dt,$$

- $\rho > 0$: discount rate; $G_i \subseteq G$ is the connected component of firm i.

- The asymptotic knowledge growth rate for firm i is a linear function of the largest eigenvalue, $\lambda_{\mathrm{PF}}(G_i)$, of the adjacency matrix associated to component G_i

$$\lim_{t\to\infty} \frac{\dot{x}_i(t)}{x_i(t)} = \lambda_{\mathrm{PF}}(G_i) + \gamma.$$

- We can exploit the above result to obtain a more tractable expression for firm's discounted profits from collaborations

$$\tilde{\pi}_i(G_i, \tilde{\mathbf{c}}, \rho) \approx \frac{V(\lambda_{\mathrm{PF}}(G_i) + \gamma)}{\rho^2} - \frac{\tilde{c}_1 d_i}{\rho^2}.$$

- Or (affine transformation)

$$\boxed{\pi_i = \lambda_{\mathrm{PF}}(G_i) - c d_i}$$

Firm's Payoff Function: Key Properties

- $\lambda_{\mathrm{PF}}(G)$ is the same for all firms in the same component G
  - strong externality effect: formation/severance of a collaboration affects all firms in the same component.
- However, firms' stock of knowledge, profits and marginal revenues/profits depend on the position of the firm in G.
- $\lambda_{\mathrm{PF}}(G)$ (utility) increases with the number of walks in G ("multiconnectivity", see Powell et al., 2005).
  - It is best for a firm to reach the other firms through many walks but to have not too many links to pay for.
- Other properties
  - (i) $\lambda_{\mathrm{PF}}(G') \geq \lambda_{\mathrm{PF}}(G)$ if $ij \notin E$ and $\lambda_{\mathrm{PF}}(G') \leq \lambda_{\mathrm{PF}}(G)$ if $ij \in E$, with inequalities being strict if the graph is connected.
  - (ii) $\lambda_{\mathrm{PF}}(G) \leq \lambda_{\mathrm{PF}}(K_n) = n - 1$.
  - (iii) $|\lambda_{\mathrm{PF}}(G') - \lambda_{\mathrm{PF}}(G)| \leq 1$.
  - (iv) $|\lambda_{\mathrm{PF}}(G') - \lambda_{\mathrm{PF}}(G)|$ decreases with n in many networks structures

Comparing Payoff Functions (1/2)

Jackson and Wolinsky (1996) introduces a utility function of the form

$$\pi_i = \sum_{i=1}^{n} \delta^{d(i,j)} - cd_i$$

where $0 \leq \delta \leq 1$ and $d(i,j)$ is the length of the shortest path from node i to node j. The cost term in our utility function is the same as in Jackson and Wolinsky (1996). The difference is in the benefit term: while the latter utility function only considers the shortest path we take into account all walks across.

Comparing Payoff Functions (2/2)

Bala and Goyal (2000), introduced a utility function of the form

$$\pi_i = |G_i| - cd_i,$$

where $|G_i|$ is the size of the connected component of agent $i \in G_i$, that is the number of agents who can be reached by agent i in the network G. This utility function takes into account all agents that agent i. The higher is the number of agents that are in its connected component the higher is the utility of agent. The difference between the Bala and Goyal utility function and our profit function is that our model assumes that all the characteristics of the topology of the component, and not only its size, contribute to determine the profit of the firm.

Outline

Efficiency

- We define social welfare as the sum of firms' individual profits

$$
\begin{aligned}
\Pi(G,c) &= \sum_{i=1}^{n} \pi_i(G_i,c) \\
&= \sum_{i=1}^{n} (\lambda_{\mathrm{PF}}(G_i) - cd_i) \\
&= \sum_{i=1}^{n} \lambda_{\mathrm{PF}}(G_i) - 2mc.
\end{aligned}
$$

- Let $\mathcal{G}(n)$ denote the set of graphs with n nodes, then the efficient graph G^* is defined as

$$G^* = \underset{G \in \mathcal{G}(n)}{\operatorname{argmax}} \quad \Pi(G,c).$$

Proposition: Let $\mathcal{H}(n, m)$ denote the set of connected graphs having n nodes and m links. If $c \in (0, 1)$ then $G^* \in \mathcal{H}(n, m)$.

If all firms are in one connected component $G = G_1 = G_2 = ... = G_n$ we have

$$\Pi(G, c) = n\lambda_{\text{PF}}(G) - 2mc$$

Corollary: The graph $G^* \in \mathcal{H}(n, m)$ has a stepwise adjacency matrix.

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Definition: (Brualdi,'85) A stepwise matrix **A** is a matrix with elements a_{ij} satisfying the condition: if $i < j$ and $a_{ij} = 1$ then $a_{hk} = 1$ whenever $h < k \leq j$ and $h \leq i$.

Proposition Let G^* be the efficient graph for a given number n of firms and cost of collaboration c.

(i) If $c \in [0, 1)$ then G^* is a connected nested split graph, i.e. a graph with a stepwise adjacency matrix

(ii) If $c \in [0, \frac{n}{2n-1}]$ ($\sim [0, 0.5]$ for large n) then G^* is unique and given by the complete graph K_n.

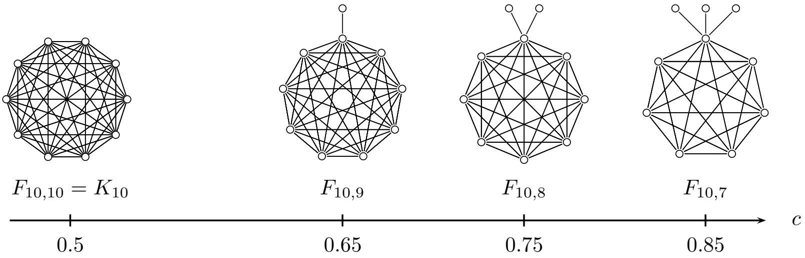
(iii) If $c \in [\frac{n}{2n-3}, 1)$ ($\sim [0.5, 1)$ for large n) and $n > 3$ then K_n is not efficient and G^* is asymmetric.

(iv) If $c \in [1, \infty)$ then G^* is a nested split graph with possibly some isolated nodes.

Definition: A stepwise matrix $\mathbf{A}$ is a matrix with elements a_{ij} satisfying the condition: if $i < j$ and $a_{ij} = 1$ then $a_{hk} = 1$ whenever $h < k \leq j$ and $h \leq i$ (Brualdi,'85).

Efficiency

- Density of the efficient graph decreases with collaboration costs
- As costs increase the graphs becomes asymmetric (for $c > 0.5$) and then disconnected (for $c > 1$)
- Disconnected efficient graphs always feature one non trivial component and firms that are completely excluded from the network



Equilibrium Concepts

Improving Path (see Jackson and Watts 1998):

- A link is *removed whenever at least one of the players strictly gains* from the change,

  $\forall ij \in G_t \; \pi_i(G_t) \geq \pi_i(G_t - ij)$ and $\pi_j(G_t) \geq \pi_j(G_t - ij)$ and

- A link is created whenever *neither player is harmed* by the creation and *at least one of them strictly gains*,

  $\forall ij \notin G_t$ if $\pi_i(G_t + ij) > \pi_i(G_t)$ then $\pi_j(G_t + ij) < \pi_j(G_t)$.

Pairwise Stability (Jackson and Wolinsky, 1996)

The graph G is pairwise stable if:

(i) for all $ij \in E(G)$, $\pi_i(G) \geq \pi_i(G - ij)$ and $\pi_j(G) \geq \pi_j(G - ij)$ and

(ii) for all $ij \notin E(G)$, if $\pi_i(G + ij) > \pi_i(G)$ then $\pi_j(G + ij) < \pi_j(G)$.

Equilibrium Networks

Proposition Let c denote the marginal cost of a link and n the total number of firms in the network G. The following network structures are **pairwise stable**(Jackson and Wolinsky, 96) in the space (c, n):

- The complete graph K_n (for $c = 0, \ \forall n$ is the unique stable graph)
- The empty graph, $\bar{K}_n$, (for $c > 1 \ \forall n$)
- The graph consisting of $d \geq 2$ disconnected cliques $K_k^1, ..., K_k^d$ with $n = kd$
- The spanning star $K_{1,n-1}$
- The dominant-group architecture $D_{n,k}$ is stable if $n = k + 1$ but it is not stable if $n \geq k + 2$.
  - Any network with two or more isolated nodes is not stable in the cost range $c \in [0, 1)$.

Example of Equilibrium Networks

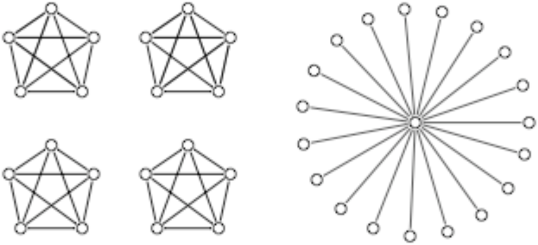


Figure: An example of two possible equilibrium networks. A set of disconnected cliques of the same size (left) and a spanning star (right).

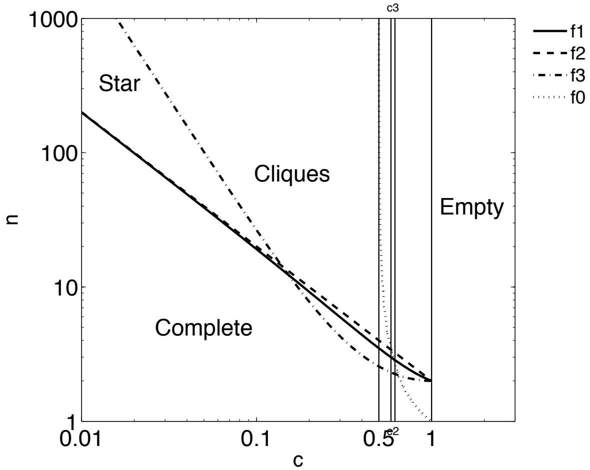


Figure: Characterization of stable networks for combinations of cost c and network size n. Both axes are in logarithmic scale.

On the possibility and impossibility of efficient equilibrium networks

- Equilibrium analysis reveals a region of small industry size and small cost of collaboration in which stable graphs are efficient
  - To the left of the curve $g(c)$ the complete graph is the unique efficient graph.
  - In addition, below the solid curve $f_1(c)$ the complete graph is also stable.
- In contrast, the efficient graph does not belong to the set of possible equilibria in a large region of collaboration costs if the size of the industry is large enough.
  - This occurs in the region above curve $f_1(c)$ and to the left of curve $g(c)$

- Furthermore, stable graphs differ in important respects from efficient graphs
  - Stable networks are too much asymmetric (the star) or disconnected (the cliques) when both symmetry and connectedness are socially desirable (i.e. for large n and low costs of collaboration).
  - Stable networks can be both symmetric and disconnected (the cliques) when both asymmetry and connectedness are socially desirable (i.e. for large n and high collaboration costs).

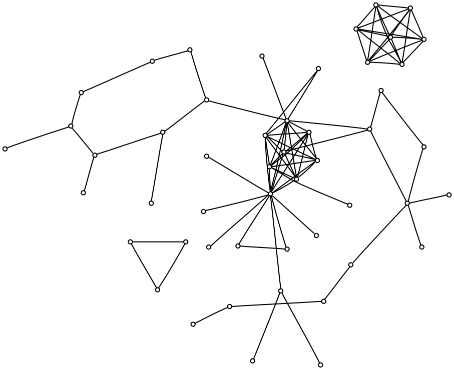
Outline

Simulation Analysis of Network Dynamics

Goals:

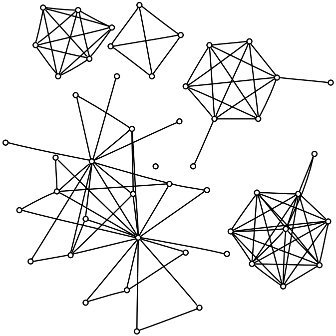
- To enlarge the set of network equilibria
- Analyze which stable network structures are actually reached by the R&D network evolution process.
- Analyze the topological properties of the selected network properties and confront them with stylized facts on empirical R&D networks.

Example: Equilibrium Network for $\alpha = 0.0$ and $c = 0.15$



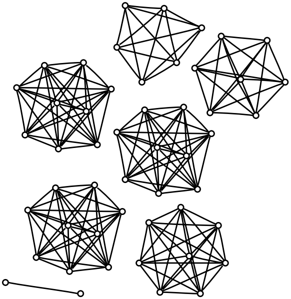
- Heterogeneous degree distribution (hubs), giant component
- High severance cost prevent firms from further deleting links

Example: Equilibrium Network for $\alpha = 0.1$ and $c = 0.15$



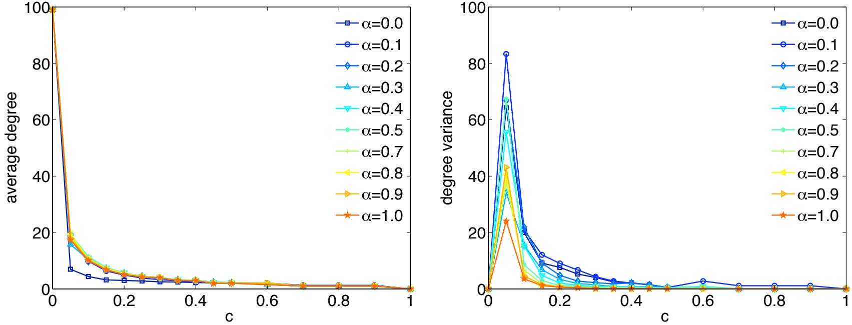
- Stronger clustering
- More disconnected components

Example: Equilibrium Network for $\alpha = 1.0$ and $c = 0.15$

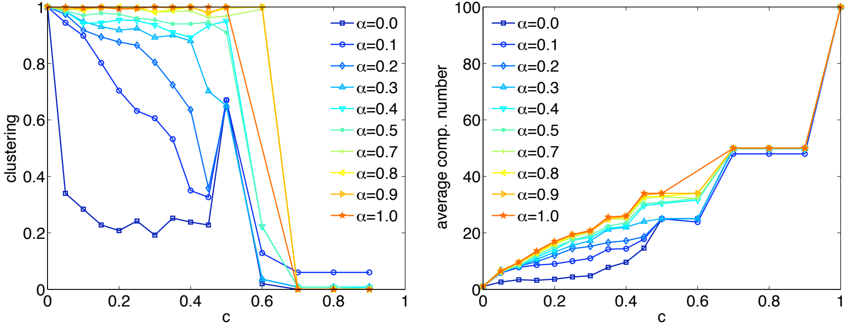


- The smaller severance costs, the larger the tendency to form disconnected cliques

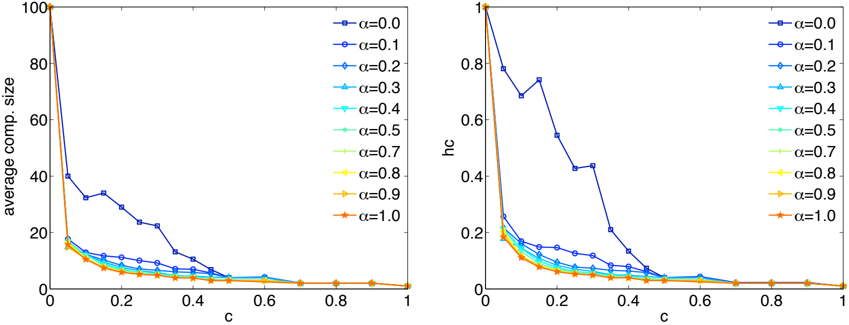
Average Degree and and Degree Variance



Clustering and Components Number



Component Size and Concentration



Simulation Analysis: Main findings

- Sparse equilibrium networks organized in clusters of highly interconnected firms are a robust feature of our model
- Moreover, low values of collaboration costs and high values of severance costs lead to equilibrium structures featuring a small number of large components with a highly dispersed degree distribution.
- As collaboration costs increase and severance costs decrease we observe that equilibrium networks tend to be organized in size-homogeneous cliques having few connections among them

Outline

From theoretical to empirical R&D networks

- Theoretical works provide several testable predictions about the dynamics and the efficiency of network structures. Do real networks confirm or infirm these predictions?
- Which properties of R&D networks are invariant across sectors and at different scales of aggregation?
- How do R&D network properties evolve over time?

Investigating empirical R&D networks

- Perform a **cross-sectoral** analysis of the structural properties of R&D networks in manufacturing sectors
- Track the **evolution** of the networks for 25 years (from 1984 to 2009).
- Investigate via regression analysis the mechanisms that underlie alliance formation and the emergence of network properties.

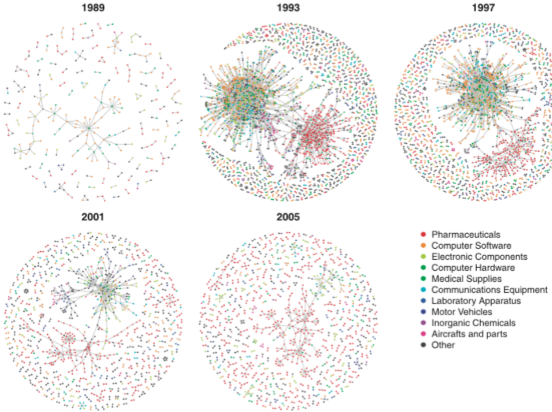
The dataset

- **SDC Platinum** alliance database, by Thomson Reuters.
- 21572 publicly announced R&D partnerships, from 1984 to 2009.
- All kinds of economic actors: manufacturing firms, investors, banks, universities, etc. We focused on manufacturing firms.
- Every firm is associated with a SIC (Standard Industrial Classification) code.
- We check all firm names and control for all legal extensions (e.g. ltd and inc) and other recurrent keywords (e.g. bio, tech, pharma, and lab) that could affect the matching between entries referring to the same firm. We keep as separated entities the subsidiaries of the same firm located in different countries. The raw data set contains 16, 313 firms, which are reduced to 9499 after running such an extensive standardization procedure.

Building the empirical R&D networks

- We link two firms every time an alliance is announced. When an alliance involves more than two companies (*consortium*), all the firms are connected in pairs.
- We assume that every alliance lasts 3 years, consistent with empirical work on alliances average duration (Deeds & Hill, 1999; Phelps, 2003).
- **Pooled R&D network**: contains all the companies, independently of their industrial sector.
- **Sectoral R&D sub-networks**: we consider all the alliances involving at least one company belonging to the selected industry.
- We obtain 26 snapshots (from 1984 to 2009) of the Pooled and the sectoral R&D networks.

R&D networks: some basic facts



- The network size increases to a peak in 1997 ("rise phase") and then shrinks again ("fall") phase. **Universality** across industries.

Evolution of sectoral networks

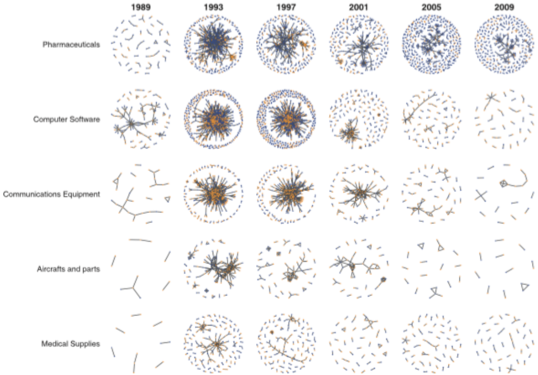


Figure 2. Evolution of five selected sectoral R&D networks. Snapshots in 1989, 1993, 1997, 2001, 2005, and 2009 for five selected sectoral R&D networks: Pharmaceuticals, Computer Software, Communication Equipment, Aircrafts and parts, Medical Supplies. Blue nodes represent the firms strictly belonging to the examined sector, while orange nodes represent their alliance partners belonging to different sectors.

Network size and density

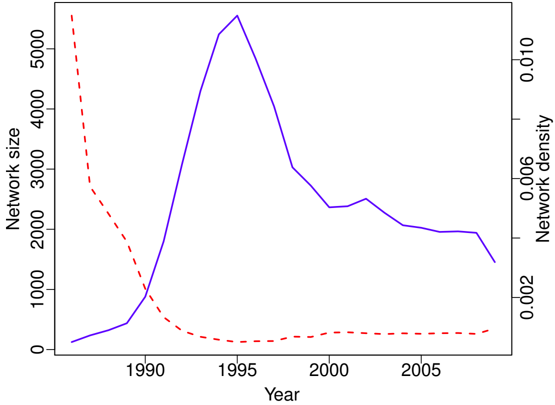


Figure: Evolution of size (solid line) and density (dashed line) in the pooled R&D network

Network connected components

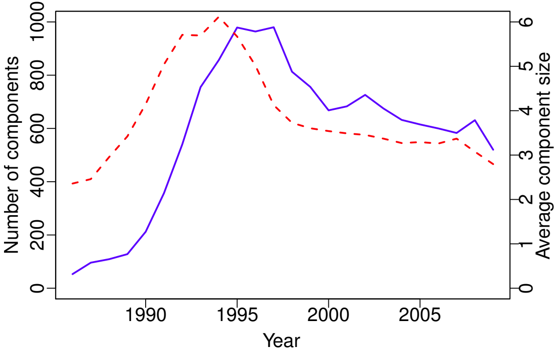
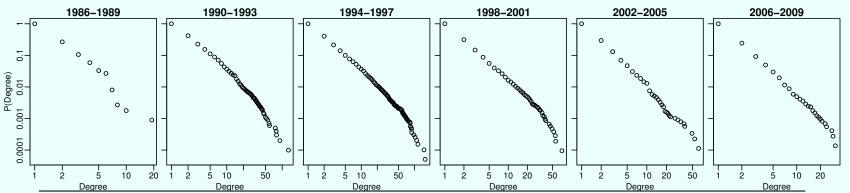


Figure: Evolution of the number of connected components (solid line) and average size of connected components (dashed line) in the pooled R&D network.

Degree Heterogeneity: Results



	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Mean	1.51	2.52	2.51	1.87	1.70	1.49
SD	1.22	4.30	4.98	2.77	2.11	1.45
Skewness	4.90	9.35	11.28	9.26	10.56	7.92
Kurtosis	47.30	158.40	206.69	133.70	200.25	104.84
KS test *p*-Value	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$
Hill tail estimator	3.04	2.31	2.34	2.61	2.78	3.05

- All the R&D networks are characterised by dispersed, skewed and fat-tailed degree distributions, as shown by their first four moments.
- A higher average degree is associated with more dispersed, skewed and fat-tailed degree distributions: more alliance **activity** $\Longrightarrow$ more alliance **inequality**.

Assortativity: Results

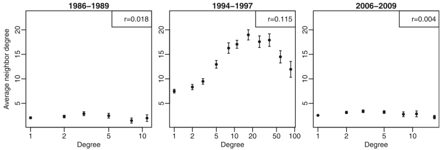


Figure 5. Average neighbors' degree as a function of firm's degree in the pooled R&D network. The whiskers measure ±2*standard deviation from the mean. *Note*: On the top-right corner of each plot we report the corresponding value of the assortativity mixing coefficient in the sub-period under analysis.

Table 7. Assortativity mixing coefficient in the pooled and the sectoral R&D networks (SIC codes are in brackets)

Sector	1986–1989	1990–1993	1994–1997	1998–2001	2002–2005	2006–2009
Pooled	0.018	0.128	0.115	0.210	0.262	0.004
Pharmaceuticals	0.271	0.328	0.336	0.306	−0.046	−0.051
Computer Software	−0.113	0.010	0.009	0.104	0.505	−0.027
Electronic Components	0.209	0.112	0.068	0.201	0.316	0.595
Computer Hardware	−0.106	−0.020	−0.047	0.033	0.186	0.317
Medical Supplies	−0.100	0.545	0.476	0.259	0.045	−0.004
Communications Equipment	0.188	0.057	0.036	0.312	0.407	0.303
Laboratory Apparatus	−0.046	0.312	0.207	0.396	0.487	0.059
Motor Vehicles	−0.151	0.337	0.317	0.383	0.386	0.864
Inorganic Chemicals	0.245	−0.052	0.233	0.297	0.422	−0.153
Aircrafts and parts	0.173	0.195	0.417	0.291	0.423	0.556

- Both Pooled and Sectoral R&D networks are **assortative**, especially at the peak of the rise phase. Inverted U-shaped relation between node degree and average neighbors' degree.

Small World Property: Results

Table 8. Small world quotient of pooled and sectoral R&D networks, for the *giant component*

Sector	1986–1989	1990–1993	1994–1997	1998–2001	2002–2005	2006–2009
Pooled	1.38	42.93	90.43	32.08	9.81	2.55
Pharmaceuticals	0.00	19.83	44.87	20.75	4.06	2.36
Computer Software	0.90	17.56	39.34	12.08	4.86	0.35
Electronic Components	1.47	12.13	20.53	11.72	6.79	2.06
Computer Hardware	0.42	14.52	25.56	10.22	4.50	0.00
Medical Supplies	0.00	2.82	7.03	1.46	0.39	0.23
Communications Equipment	1.78	8.28	15.34	8.10	4.19	1.46
Laboratory Apparatus	0.00	5.25	5.58	2.95	1.84	0.64
Motor Vehicles	0.99	4.17	7.15	4.28	3.09	1.18
Inorganic Chemicals	1.29	3.66	6.38	1.24	1.23	0.00
Aircrafts and parts	0.83	4.50	5.38	3.02	1.93	2.09

- R&D networks exhibits a rise-and-fall dynamics in the small world property, both at the pooled and at the sectoral level.
- Small worlds plus degree inequality $\rightarrow$ core-periphery structures?

Hierarchical structures: Results

Nestedness

Table 9. Nestedness coefficients for the pooled and the sectoral R&D networks

Sector	1986–1989	1990–1993	1994–1997	1998–2001	2002–2005	2006–2009
Pooled	0.791	0.996	0.999	0.996	0.988	0.994
Pharmaceuticals	0.706	0.983	0.993	0.990	0.988	0.994
Computer Software	0.823	0.991	0.997	0.979	0.930	0.805
Electronic Components	0.778	0.979	0.991	0.980	0.950	0.820
Computer Hardware	0.804	0.991	0.995	0.982	0.921	0.662
Medical Supplies	0.680	0.838	0.926	0.701	0.750	0.775
Communications Equipment	0.464	0.955	0.988	0.966	0.893	0.585
Laboratory Apparatus	0.728	0.879	0.945	0.911	0.799	0.732
Motor Vehicles	0.630	0.835	0.949	0.898	0.813	0.651
Inorganic Chemicals	0.561	0.925	0.924	0.762	0.678	0.805
Aircrafts and parts	0.688	0.848	0.874	0.827	0.801	0.570

The values are averaged in six sub-periods.

- Rise and fall of **nested neighborhood structures**, in correspondence with the presence of small worlds, in the pooled and sub-sectoral R&D network.

Network regression analysis

- The analysis of the properties of R&D networks provides hints about the mechanisms of formation of alliances in the network
- We further test the validity of different models using the regression model developed in Nesta et al. (2010)
- Tested Models

H1 **Accumulative Advantage**. The probability of forming an alliance with a firm increases with the centrality of that firm in the network.

H2 **Structural Homophily (or Diversity)**. The probability of an alliance between two firms increases with their similarity (diversity).

H3 **Multiconnectivity**. The probability of forming an alliance with a firm increases if that firm allows to reach other firms in the network through multiple independent paths.

Network regression analysis: variables

Table 10. Nomenclature, type, and meaning of all the variables employed in our econometric model

Variable	Type	Meaning
Dependent variable		
Alliance formation	Binary	Formation of a link between the two considered firms in the considered year.
Independent variables		
1. Accumulative advantage		
Joint centrality	Positive real	Logarithm of the arithmetic mean of the degree centrality of the two firms (for incumbent dyads).
Incumbent centrality	Positive real	Logarithm of the degree centrality of the incumbent firm (for mixed dyads).
2. Structural homophily and diversity		
Same nation	Binary	1 if the two firms are registered in the same nation.
Same SIC	Binary	1 if the considered firms have the same SIC code.
Technological distance	Positive real	Distance between the two firms in a technology space (measured through patent similarity).
Centrality inequality	Positive real	Logarithm of the ratio of the degree centralities—the largest divided by the smallest—of the two firms (only for incumbent dyads).
Inverse path length	Positive real	Inverse of the length of the shortest network path connecting the two firms (only for incumbent dyads).
Common neighbors	Positive integer	Number of the alliance partners that the two firms have in common (only for incumbent dyads).
3. Multiconnectivity		
Average multi-path growth	Real	Arithmetic mean of the largest eigenvalues of the connected components to which the two firms belong (for incumbent dyads)
Incumbent multi-path growth	Real	Largest eigenvalue of the connected components to which the incumbent firm belongs (for mixed dyads).
Control variables		
Year dummies	Binary	23 dummy variables (the observation period consists of 24 years) for the time-fixed effect models.
Past alliances	Positive integer	Number of alliances established between the two firms in all previous years (only for incumbent dyads).
$(\text{Past alliances})^2$	Positive integer	Square of the variable *Past alliances* (only for incumbent dyads).
Sector alliances	Positive real	Arithmetic mean of the number of alliances in the industrial sectors of the two firms, in the year preceding the observation.

The observation unit consists of a firm dyad, i.e. every potential pair of firms in the network.

Network regression analysis: results

Table 12. Estimates of the model coefficients on the incumbent dyads (standard error in parentheses)

Risk set A	Rise phase (1986–1997)			Fall phase (1998–2009)		
Incumbent dyads	Model 1A	Model 2A	Model 3A	Model 4A	Model 5A	Model 6A
(Intercept)	−4.766**	−4.375**	−4.337**	−9.577**	−7.168**	−7.127**
	(0.998)	(0.978)	(0.979)	(0.264)	(0.288)	(0.494)
Joint centrality	1.552**	1.016**	1.043**	1.402**	1.003**	1.006**
	(0.036)	(0.048)	(0.049)	(0.077)	(0.105)	(0.109)
Same nation		0.368**	0.370**		0.039	0.039
		(0.054)	(0.054)		(0.120)	(0.120)
Same SIC		0.659**	0.657**		0.353*	0.352*
		(0.066)	(0.065)		(0.146)	(0.146)
Technological distance		−2.315**	−2.305**		−3.439**	−3.440**
		(0.111)	(0.111)		(0.272)	(0.272)
Centrality inequality		−0.099**	−0.114**		−0.412**	−0.412**
		(0.036)	(0.036)		(0.087)	(0.087)
Inverse path length		2.362**	2.443**		2.134**	2.139**
		(0.115)	(0.114)		(0.223)	(0.228)
Common neighbors		0.016	0.010		−0.040	−0.041
		(0.010)	(0.011)		(0.029)	(0.029)
Average multi-path growth			−0.080**			−0.003
			(0.019)			(0.030)
Past alliances	0.250	0.135	0.126	0.602	0.518	0.517
	(0.452)	(0.414)	(0.411)	(0.709)	(0.795)	(0.796)
(Past alliances)^2	−0.155	−0.088	−0.084	−0.138	−0.116	−0.116
	(0.272)	(0.231)	(0.228)	(0.321)	(0.396)	(0.397)
Sector alliances	0.001	−0.001*	−0.001*	−0.001	−0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Year dummies	(Yes)	(Yes)	(Yes)	(Yes)	(Yes)	(Yes)
AIC	17712.974	15404.617	15390.905	4359.410	3741.857	3743.846
BIC	17891.093	15649.530	15646.951	4527.007	3972.303	3984.767
Log Likelihood	−8840.487	−7680.308	−7672.452	−2163.705	−1848.928	−1848.923
Likelihood ratio test	0	2320.358	2336.070	0	629.554	629.564
Deviance	17680.974	15360.617	15344.905	4327.410	3697.857	3697.846
Number of observations	505,067	505,067	505,067	261,666	261,666	261,666

Models 1A and 4A test the effect of accumulative advantage only; models 2A and 5A test the effect of accumulative advantage and structural homophily (diversity); models 3A and 6A test the effect of accumulative advantage, structural homophily (diversity), and multiconnectivity. Models 1A, 2A, and 3A are related to the rise phase (1986–1997) of the R&D network, while models 4A, 5A, and 6A are related to its fall phase (1998–2009). The likelihood ratio test of the models 2A and 3A is computed with respect to model 1A; the likelihood ratio test of the models 5A and 6A is computed with respect to model 4A.

$^{**}P < 0.01$; $^{*}P < 0.05$.

Conclusions of empirical analysis

1. Many network properties robustly hold *across sectors* and *across scales* of aggregation (pooled vs. sectoral networks)
   - heterogeneous degree distribution;
   - assortativity;
   - small world properties;
   - presence of hierarchical structures.
2. The result that the pooled and sectoral networks are organized into hierarchical architectures (and nested structures in particular) confirm the predictions of the knowledge recombination model described in the first part of the lecture.
3. The regression analysis of alliance formation indicates that the dynamics of the networks was driven by mechanisms of accumulative advantage, structural homophily and multiconnectivity. The break from the rise to fall phase is explained by the loss in significance of multiconnectivity.

- M. Tomasello, M.Napoletano, A. Garas and Frank Schweitzer, 2016, *The rise and fall of R&D networks*, Industrial and Corporate Change, doi: 10.1093/icc/dtw041 .
- M. D. König, S. Battiston, M. Napoletano and Frank Schweitzer, 2012, *The Efficiency and Stability of R&D Networks*, Games and Economic Behavior Volume 75, Issue 2, Pages 694-713
- M. D. König, S. Battiston, M. Napoletano and Frank Schweitzer, 2011,*Recombinant Knowledge and the Evolution of Innovation Networks*, Journal of Economic Behavior and Organization, Volume 79, Issue 3, Pages 145-164
- M. D. König, S. Battiston, M. Napoletano and Frank Schweitzer, 2008, *On Algebraic Graph Theory and the Dynamics of Innovation Networks*, Networks and Heterogeneous Media, 3(2):201–219

mauro.napoletano@sciencespo.fr