Modeling the R&D collaborations of firms

Prof. Frank Schweitzer

Science / AAAS

Economic Networks: The New Challenges
Frank Schweitzer, *et al.*
Science **325**, 422 (2009);
DOI: 10.1126/science.1173644

The need to combine two perspectives

- **Micro:** *Socioeconomic perspective*
  - strategic behaviour of single agents' ⇔ network architecture
- **Macro:** *Physics/Computer science perspective*
  - statistical regularities of the network as a whole
- **Data-driven modeling**: infer interaction rules of agents

more details in: F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, D. R. White (2009). Economic Networks: What Do We Know And What Do We Need To Know?, Advances in Complex Systems 12 (2009) 407

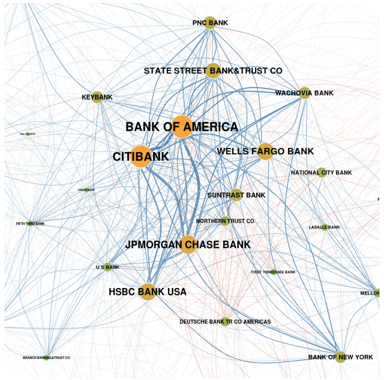
Example 1: Financial Network

Systemic risk

- focus on dynamics *along* links
  - redistribution of load
  - feedback on nodes
  - failure cascades

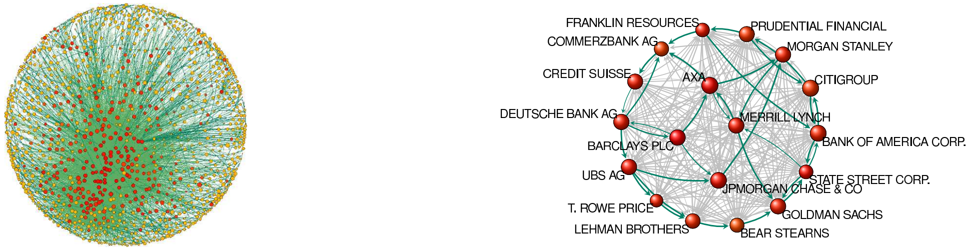
OTC activties

- data: US national banks, 1998/Q4 - 2012/Q4
- links serve a purpose:
  - risk diversification
  - weighted, directed
- core-periphery structure



Nanumyan V, Garas A, Schweitzer F (2015) The Network of Counterparty Risk: Analysing Correlations in OTC Derivatives. PLoS ONE 10(9): e0136638

Example 2: Ownership Network

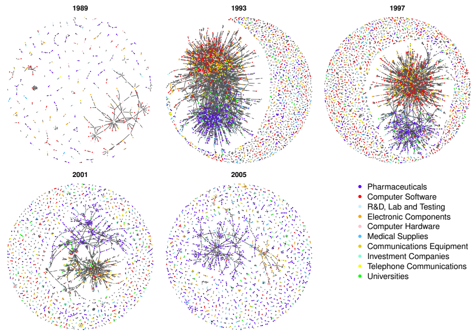


(left) SCC (1318 nodes, 12191 links). Node size scales logarithmically with operation revenue.
(right) Zoom on some major TNCs in the financial sector. Some cycles are highlighted.

- 75% of the ownership of the SCC firms stays within the SCC
  - propagation of financial distress increases systemic risk
  - cross-ownership decreases competition ⇒ market failure

S. Vitali, J. Glattfelder, S. Battiston: The network of global corporate control, PLoS ONE (2011)

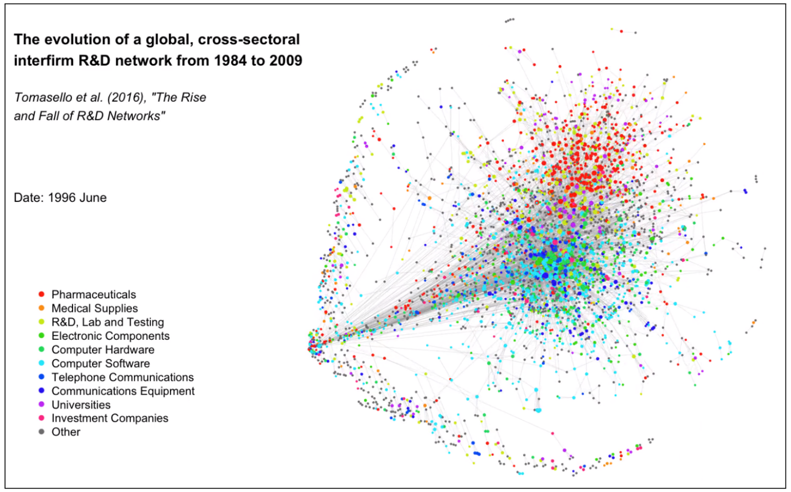
Example 3: R&D networks



- **Data:** SDC Platinum (1984-2009): 14.561 firms, 21.572 alliances
- **Life cycle of R&D networks:** collaborations have a finite lifetime

M. Tomasello, M. Napoletano, A. Garas, F. Schweitzer: *The Rise and Fall of R&D Networks*, **Industrial and Corporate Change** vol 26/4, pp 617-646 (2016)

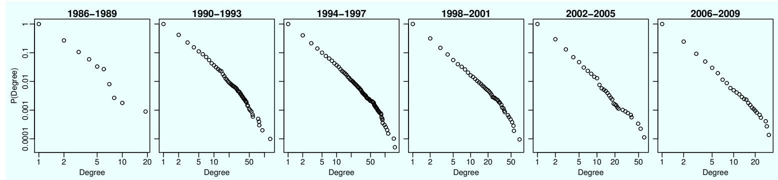
More papers ⇒ **www.sg.ethz.ch** ⇒ **Projects** ⇒ **R&D Networks**



What do we know about R&D networks?

- **Theory: Agent-based models of strategic link formation**
  - *predictions for R&D networks:* stable, sparse, highly clustered, highly skewed degree distributions, core-periphery structures
    - M. König, S. Battiston, M. Napoletano, F. Schweitzer: *On Algebraic Graph Theory and the Dynamics of Innovation Networks*, **Networks and Heterogeneous Media 3(2):201-219 (2008)**
    - M. König, S. Battiston, M. Napoletano, F. Schweitzer: *Recombinant knowledge and the evolution of innovation networks*, **Journal of Economic Behavior & Organization 79:145-164 (2011)**
    - M. König, S. Battiston, M. Napoletano, F. Schweitzer: *The efficiency and stability of R&D networks*, **Games and Economic Behaviour 75(2):694-713 (2012)**
- **Empirics: Structure and dynamics, cross-sectoral analysis**
  - *quantification:* degree heterogeneity, assortativity, small-world properties, modularity, core-periphery structure, nestedness
    - M. Tomasello, M. Napoletano, A. Garas, F. Schweitzer: *The Rise and Fall of R&D Networks*, **Industrial and Corporate Change 26(4):617-646 (2017)**
    - T. Scholl, A. Garas, F. Schweitzer: *The spatial component of R&D networks*, **Journal of Evolutionary Economics 28(2):417-436 (2018)**

Results: Heterogeneity

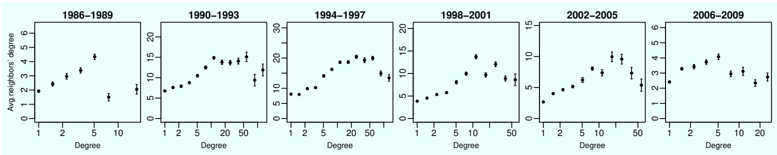


	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Mean	1.51	2.52	2.51	1.87	1.70	1.49
SD	1.22	4.30	4.98	2.77	2.11	1.45
Skewness	4.90	9.35	11.28	9.26	10.56	7.92
Kurtosis	47.30	158.40	206.69	133.70	200.25	104.84
KS test *p*-Value	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$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

- R&D networks are characterised by dispersed, skewed and fat-tailed degree distributions (see first four moments)
- A higher average degree is associated with more dispersed, skewed and fat-tailed degree distributions
  - more alliance **activity** ⟹ more alliance **inequality**

ETH zürich

Results: Assortativity

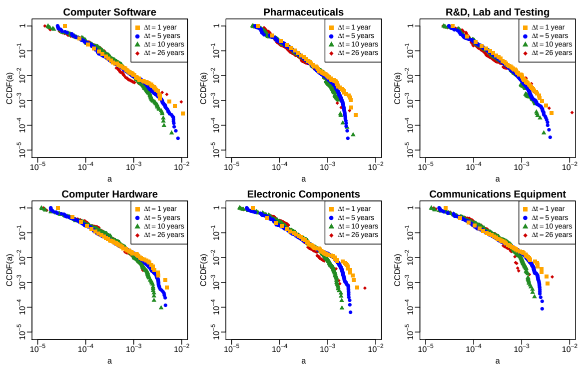


Assortativity mixing coefficient	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Aggregate	0.167	0.110	0.119	0.195	0.170	0.035
Computer Software (737)	-0.103	-0.074	-0.067	-0.029	-0.002	-0.105
Pharmaceuticals (283)	0.005	0.172	0.119	-0.049	-0.047	-0.043
R&D, Lab and Testing (873)	-0.024	-0.032	0.011	0.132	0.185	0.025
Computer Hardware (357)	-0.188	-0.179	-0.192	-0.133	-0.103	-0.145
Electronic Components (367)	-0.174	-0.151	-0.194	-0.094	0.023	0.267
Communications Equipment (366)	-0.233	-0.149	-0.147	-0.143	-0.077	-0.312
Medical Supplies (384)	-	-0.165	-0.155	0.106	-0.184	-0.108
Telephone Communications (481)	-0.273	-0.178	-0.097	-0.035	-0.036	-0.279
Universities (822)	-	-0.133	-0.102	0.026	0.152	0.078

- aggregate R&D network is **assortative**, sectoral R&D sub-networks tend to be **disassortative**
- Emergence of a **twofold behaviour** at the micro-perspective. Inverted U-shaped degree correlation.

ETH zürich

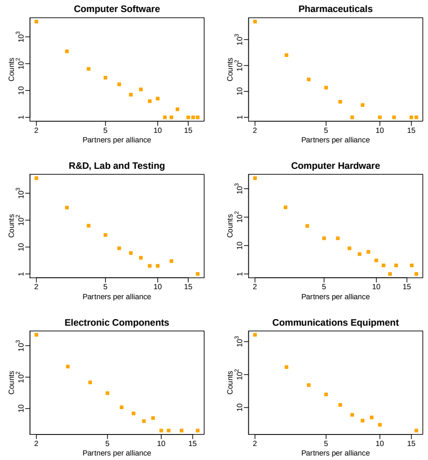
Input: Agents' activity a_i



- **Activity**: propensity to engage in a collaboration
  - $\Rightarrow$ broad and right skewed distribution

ETH zürich

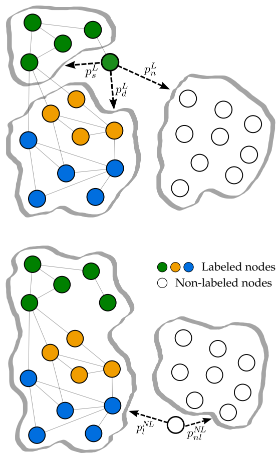
Input: Collaboration size m



- broad and right skewed **collaboration size** distribution
- most R&D alliances $\Rightarrow$ 2 partners
- significant presence of consortia

ETH zürich

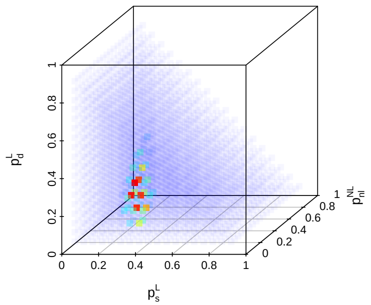
Agent based model of link formation



- **agent** i: two fixed properties
  - **activity** a_i: propensity to engage in a collaboration $\Rightarrow$ from data
  - **label** l_i: membership in a *circle of influence/group* ($\rightarrow$ *color*)
- **dynamics**:
  1. activation
  2. choose m *collaboration partners* (m taken from data)
     - **Incumbent:** (labeled node): $p_s^L + p_d^L + p_n^L = 1$
     - **Newcomer:** (non-labeled node): $p_{nl}^{NL} + p_l^{NL} = 1$
     - within labeled groups, partners are chosen wrt their degree
  3. form fully connected clique of size m, label propagation
- **Our task:** determine p_d^L, p_s^L, p_{nl}^{NL}

Calibration 1/2

- First exploration: *network formation* parameter space
- M. L. approach ⇒ parameter combination giving the *best match* with reality, w.r.t.:
  - average degree $\langle k \rangle$;
  - average path length $\langle l \rangle$;
  - global clustering coefficient (transitivity) C.
- *Optimal simulated network*: $p^* \equiv (p_s^{*L}, p_d^{*L}, p_n^{*L}, p_{nl}^{*NL}, p_l^{*NL})$
  - errors $\varepsilon_{\langle k \rangle}$, $\varepsilon_{\langle l \rangle}$, ε_C have to be smaller than ε_0

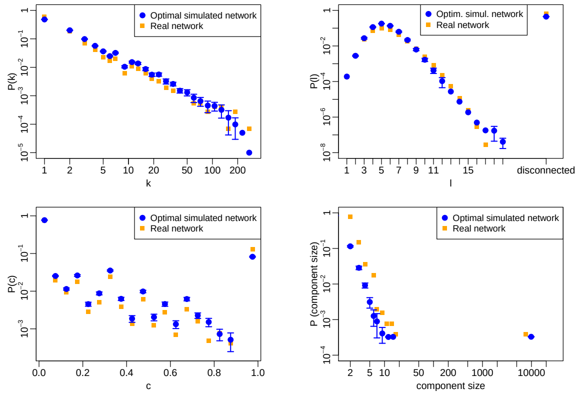


Calibration 2/2

	ε^0	$\langle k \rangle^*$	$\langle l \rangle^*$	C^*	p_s^{*L}	p_d^{*L}	p_n^{*L}	p_l^{*NL}	p_{nl}^{*NL}
Pooled R&D network	2%	2.76	5.33	0.098	0.30	0.30	0.40	0.75	0.25
Pooled R&D network w/ patents	2%	3.48	5.02	0.111	0.45	0.20	0.35	0.90	0.10
Sectoral R&D networks									
Pharmaceuticals (SIC 283)	2%	3.13	4.95	0.097	0.35	0.35	0.30	0.80	0.20
Computer hardware (SIC 357)	6%	5.37	3.59	0.175	0.55	0.30	0.15	0.90	0.10
Communications equipment (SIC 366)	2%	4.83	3.76	0.210	0.75	0.15	0.10	0.80	0.20
Electronic components (SIC 367)	2%	4.76	3.83	0.174	0.65	0.20	0.15	0.90	0.10
Computer software (SIC 737)	3%	3.56	4.27	0.141	0.55	0.20	0.25	0.95	0.05
R&D, laboratory and testing (SIC 873)	3%	3.30	5.22	0.200	0.40	0.40	0.20	0.20	0.80

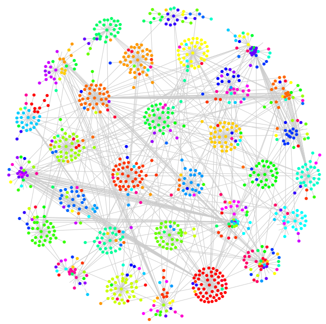
- *Incumbents* tend to form links with *other incumbents* ⇒ $p_s^{*L} + p_d^{*L} > 70\%$
  - p_s^L, p_d^L: Prob. of a **labeled** node to select a node w/ the **same** label / a **different** label
  - preference for *same circle of influence* ⇒ $p_s^{*L} \geq p_d^{*L}$ (pooled R&D network and technologically dynamic sectors more balanced).
- *Newcomers* tend to link with *incumbents* in R&D networks ⇒ $p_l^{*NL} > p_{nl}^{*NL}$
  - p_{nl}^{NL}: Prob. of a **non-labeled** node to select a **non-labeled** node
  - exception: *Newcomers* tend to link with *newcomers* in R&D: test&lab, ⇒ $p_l^{*NL} < p_{nl}^{*NL}$.

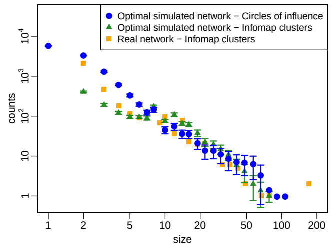
Validation: Reproduce network distributions



Reproduced distributions:
- degree
- path length
- local clustering coefficient
- component size

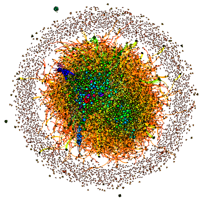
Validation: Reproduce clusters

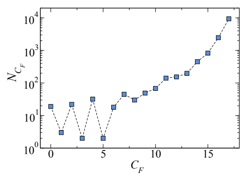




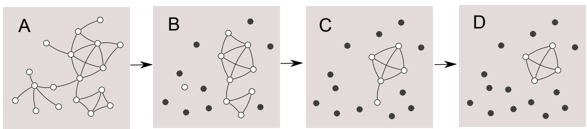
- Visualization of the optimal simulated R&D network (only the largest 30 clusters identified by Infomap). Each color corresponds to a different label.
- 1,600 **empirically detected clusters**.
- 4,900 **simulated labels**. **Overlap** = 89%.

Quantifying the network position of firms



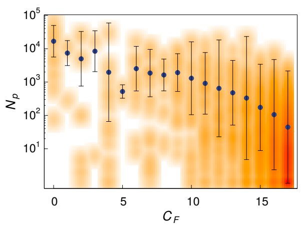


- cumulative R&D network: **core-periphery structure**
- **weighted** k**-core decompositon**: quantifies importance of nodes
  - *coreness*: $C_C^i = k_s^{\max} - k_s^i$ (*the lower C_C, the higher centrality*)



Correlation with success?

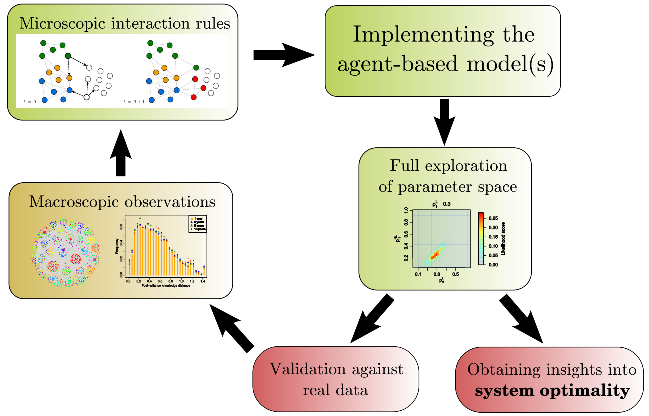
- **Success of R&D alliances:** number of patents
  - *NBER patent db*: 3 mio patents granted in US (1974-2006)
  - *overlap with SDC*: 1.5 mio patents, 6.500 firms, 23 years
  - **# patents/coreness:** Kendall correlation: **-0.843**



A low coreness value indicates success

G. Vaccario, L. Verginer, A. Garas, M. V. Tomasello, F. Schweitzer: Embeddedness and Success (2021)

Methodology: Data-driven modeling



Physics Today, Februar 2018 (free access)

Conclusions

1. **Collaboration networks:** entry/exit, rewiring of links
   - *co-authorship network:* 226.000 agents, 1.5 mio links, 110 years
   - *R&D alliance network:* 14.000 agents, 21.000 links, 26 years

2. **Agent-based modeling:** individual firm
   - **no need** to include all microscopic details ⇒ probabilistic approach
   - *interaction model:* alliances, knowledge exchange
     - **calibration:** probability link formation, **validation:** macro pattern
     - null model to **test for strategies** in firm behavior

3. **New insights from ABM:** quantification, policy design
   - *Role of heterogeneity:* quantify, study impact
     - agents' activity, preference, membership in structural communities
   - *Mechanism design:* optimize
     - identify suboptimal states, explore *incentives*, policy design