Company networks and their correlations beyond nearest neighbors

Diego Rybski
Hernán D. Rozenfeld, Stefan Hennemann

International Conference on Econophysics 2011
East China University of Science and Technology
Third Teaching Building
5.6.2011 – Session C

PIK
Potsdam Institute for Climate Impact Research

Outline

1. Network fluctuation analysis (kFA)
2. Data
3. Company networks
4. Spatial embedding

1. Network fluctuation analysis (kFA)

Motivation

Degree correlations

- likelihood that nodes of given degree are connected
- assortative/disassortative mixing:
  positive/negative correlations
- measures:
  - Pearson correlation coefficient
  - average nearest neighbor degree
  - conditional probability $p(k_1, k_2)$
  - ...

disassortativity tightly related to *fractality* of complex networks

Average nearest neighbor degree

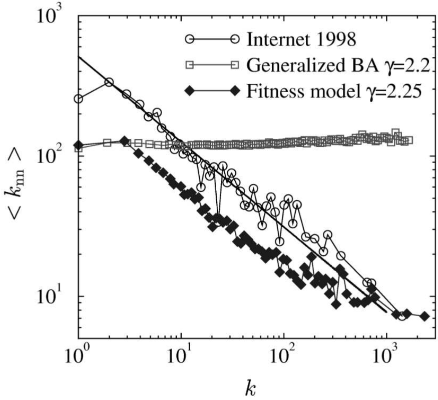


FIG. 2. The average connectivity $\langle k_{nn} \rangle$ of the nearest neighbors of a node depending on its connectivity k for the 1998 snapshot of the Internet, the generalized BA model with $\gamma = 2.2$ (Ref. [8]), and the fitness model (Ref. [18]). The solid line has a slope -0.5. The scattered results for very large k are due to statistical fluctuations.

Figure: R. Pastor-Satorras et al. PRL 2001.

Conditional probability $p(k_1, k_2)$

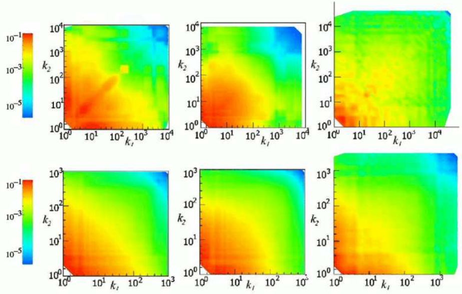


FIG. 1 (color online). The joint degree distribution $P(k_1, k_2)$ of WWW (top row) and Internet at the router level (bottom row) before renormalization (left), after renormalization forbidding multiple links (center), and including multiple links (right).

Figure: L.K. Gallos et al. PRL 2008.

Motivation

But:

- only correlations between nearest neighbor nodes, i.e. distance 1
- much of the rich topological information gets lost
- **how can correlations be measured at larger distances?**

Fluctuation Analysis

We propose a Degree Fluctuation Analysis (kFA):

Fluctuation Analysis

We propose a Degree Fluctuation Analysis (kFA):

1. Consider the shortest path between the nodes i and j (if it is not unique, we consider an arbitrary one).

Fluctuation Analysis

We propose a Degree Fluctuation Analysis (kFA):

1. Consider the shortest path between the nodes i and j (if it is not unique, we consider an arbitrary one).
2. Extract the sequence of degrees (k_l) along this path of length d_{ij}.

Fluctuation Analysis

We propose a Degree Fluctuation Analysis (kFA):

1. Consider the shortest path between the nodes i and j (if it is not unique, we consider an arbitrary one).
2. Extract the sequence of degrees (k_l) along this path of length d_{ij}.
3. Calculate the average $K_{ij} = \langle k_l \rangle$ of the sequence (k_l).

Fluctuation Analysis

We propose a Degree Fluctuation Analysis (kFA):

1. Consider the shortest path between the nodes i and j (if it is not unique, we consider an arbitrary one).
2. Extract the sequence of degrees (k_l) along this path of length d_{ij}.
3. Calculate the average $K_{ij} = \langle k_l \rangle$ of the sequence (k_l).
4. Find the shortest paths between all pairs of nodes and determine the corresponding averages, $K_{ij} \ \forall \ i \neq j$.

Fluctuation Analysis

We propose a Degree Fluctuation Analysis (kFA):

1. Consider the shortest path between the nodes i and j (if it is not unique, we consider an arbitrary one).
2. Extract the sequence of degrees (k_l) along this path of length d_{ij}.
3. Calculate the average $K_{ij} = \langle k_l \rangle$ of the sequence (k_l).
4. Find the shortest paths between all pairs of nodes and determine the corresponding averages, $K_{ij} \ \forall \ i \neq j$.
5. Calculate the fluctuation function $F(d) = \sigma(K_{ij}|d)$, the conditional standard deviation of the K_{ij} at distance d.

In analogy to Fluctuation Analysis in time series analysis

Long-range (anti-) correlations

If the covariance, $C(d) \sim \langle (k_i - \langle k \rangle)(k_j - \langle k \rangle) | d \rangle$, between the degrees at distance d scales as

$C(d) \sim d^{-\gamma}$	for positive correlations (assortative) or
$C(d) \sim -(d^{-\gamma})$	for negative correlations (disassortative)

then we expect

$$F(d) \sim d^{\alpha_k} ,$$

where $\alpha_k = -\gamma/2$.

Fluctuation exponent differs by 1 from usual Hurst-like exponent: $\alpha = \alpha_k + 1$.

Long-range (anti-) correlations

If the covariance, $C(d) \sim \langle (k_i - \langle k \rangle)(k_j - \langle k \rangle) | d \rangle$, between the degrees at distance d scales as

$C(d) \sim d^{-\gamma}$	for positive correlations (assortative) or
$C(d) \sim -(d^{-\gamma})$	for negative correlations (disassortative)

then we expect

$$F(d) \sim d^{\alpha_k},$$

where $\alpha_k = -\gamma/2$.

Fluctuation exponent differs by 1 from usual Hurst-like exponent: $\alpha = \alpha_k + 1$.

$\alpha_k > -1/2$	positive correlations (assortative)
$\alpha_k = -1/2$	uncorrelated
$\alpha_k < -1/2$	negative correlations (disassortative)

Long-range (anti-) correlations

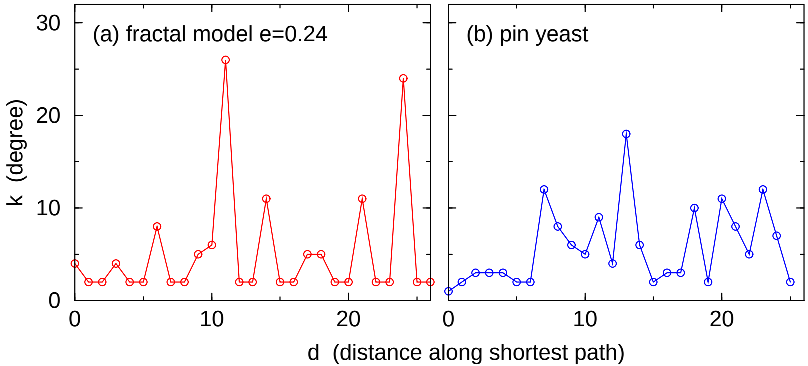


Figure: Degrees of nodes versus distance along shortest paths for (a) fractal network model and (b) pin yeast network.

1. Barabási-Albert model

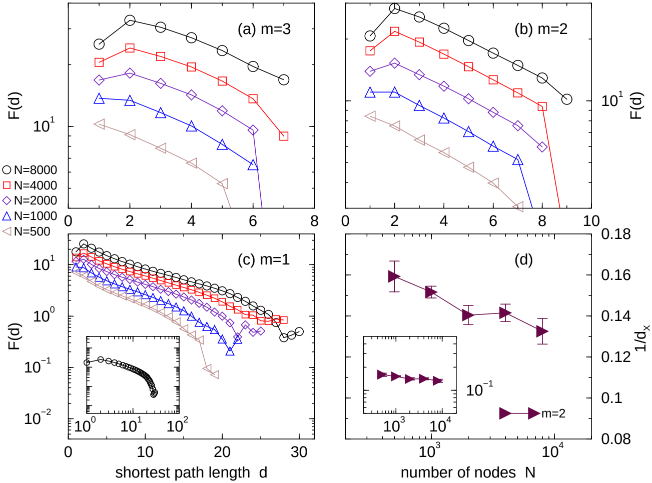


Figure: Degree fluctuation functions for the BA model. (a-c) show $F(d)$, (d) shows for $m = 2$ the slopes of exponential fits as a function of the network sizes. 100 configurations.

2. Cayley tree at percolation transition

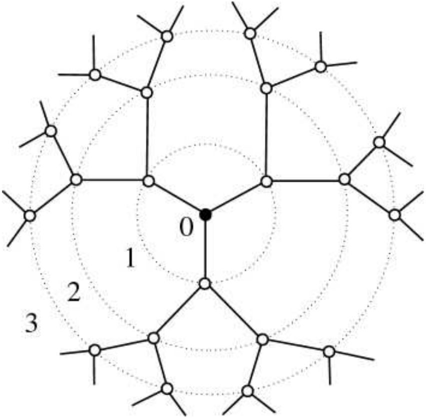


Figure: $z = 3$ (source: wikipedia)

percolation transition: $p_c = \frac{1}{z-1}$

topological dimension of giant component: $d_f = 2$

2. Cayley tree at percolation transition

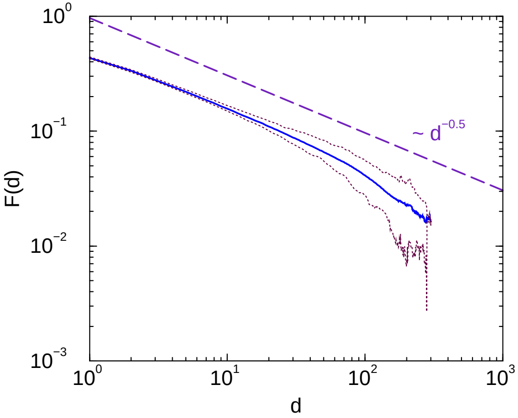


Figure: Degree fluctuation function of the Cayley tree at percolation transition ($z = 3$, $n = 150$). Dotted maroon lines: quantiles enclosing 90% (100 configurations).

3. Fractal network model

C.M. Song, S. Havlin, and H.A. Makse, nature physics, 2006.

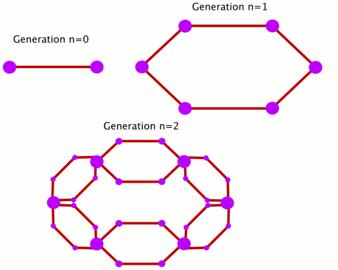


Fig. 1. Construction of a pure fractal network. Example of network model with parameters $n = 0, 1, 2; m = 2; x = 2; e = 0$.

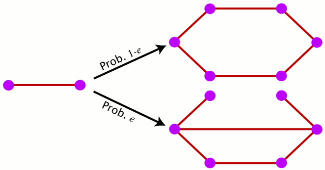


Fig. 2. Construction of network. With probability e the link between hub remains, otherwise, with probability $1 - e$ it is replaced for another link between new nodes.

Figure: generation n, m new nodes, x new links, probability e (source: H.D. Rozenfeld and H.A. Makse, 2009)

fractal dimension: $d_{\mathrm{f}} = \frac{\ln(2m+x)}{\ln(3-2e)}$

degree distribution: $p(k) \sim k^{-\left(1+\frac{\ln(2m+x)}{\ln m}\right)}$

3. Fractal network model

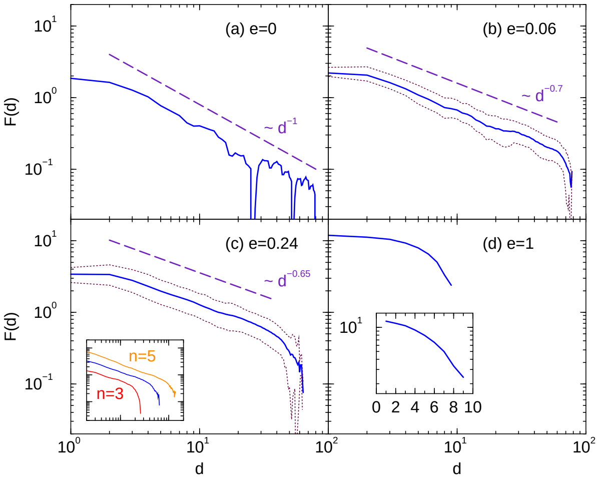


Figure: Degree fluctuation functions for the fractal model ($n = 4$). Dotted maroon lines: quantiles enclosing 90% (250 configurations). Inset in (c): $n = 3$ and $n = 5$ (25 configurations).

3. Fractal network model

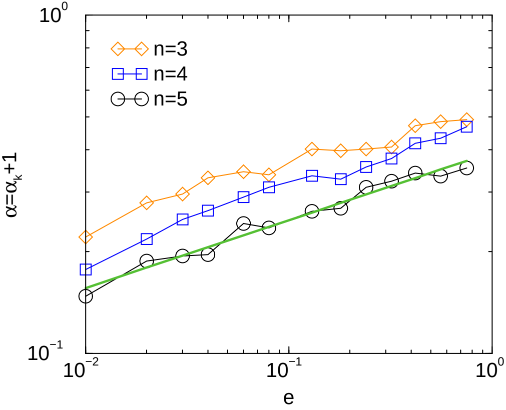


Figure: Exponents $\alpha = \alpha_k + 1$ of the fractal network model. Power-law fit for $n = 5$: $\alpha \sim e^{\epsilon}$, $\epsilon \approx 0.2$.

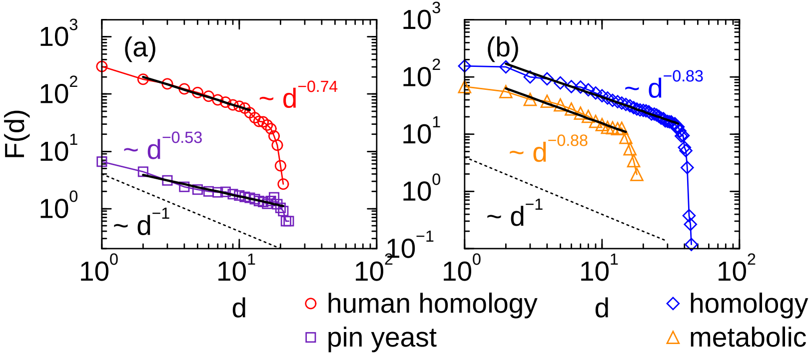
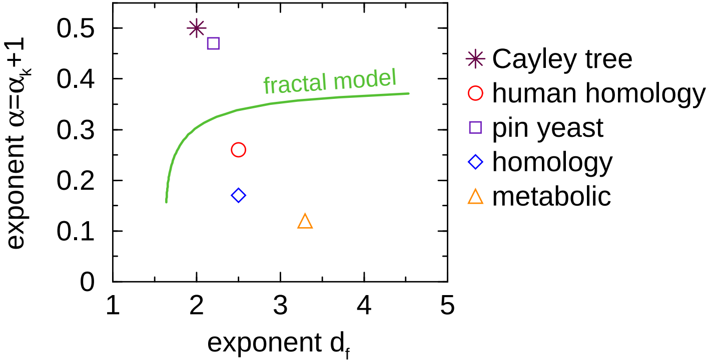


Figure: Degree fluctuation functions for real-world networks.

Comparison with fractal dimension



network	*k***FA**	**box covering** (Song et al.)
human homology	$\alpha_k \simeq -0.74$	$d_\mathrm{f} \simeq 2.5$
pin yeast	$\alpha_k \simeq -0.53$	$d_\mathrm{f} \simeq 2.2$
homology	$\alpha_k \simeq -0.83$	$d_\mathrm{f} \simeq 2.5$
metabolic	$\alpha_k \simeq -0.88$	$d_\mathrm{f} \simeq 3.3$

Findings

- BA model: exponential decay
- Cayley tree: $\alpha_k = -1/2$ (uncorrelated)
- fractal network model:
  - $e = 0$: $\alpha_k = -1$ (long-range anti-correlated)
  - $e = 1$: exponential decay
- real-world networks: power-law decay
- fluctuation exponent complementary information to fractal dimension

Findings

- BA model: exponential decay
- Cayley tree: $\alpha_k = -1/2$ (uncorrelated)
- fractal network model:
  - $e = 0$: $\alpha_k = -1$ (long-range anti-correlated)
  - $e = 1$: exponential decay
- real-world networks: power-law decay
- fluctuation exponent complementary information to fractal dimension

- vary parameters m and x
- analytical description
- also other spatial correlations:
  - various network properties (clustering, betweenness, ...)
  - additional information available (time of addition, activity, ...)

2. Data

Data

Motivation

In economic geography it is believed that economic activity is influenced by

- space, i.e. spatial distance
- network, i.e. network distance
- topic, i.e. distance in terms of content

Ergo:
small research fields via defined research topics
collaboration networks via co-authorship

Data

Data collection (*Hennemann S. et al., submitted 2011*)

publications from isi web of science (2004-2008)
network of institutions with at least 1 common publication

keyword	**nodes**	**links**	**diameter**
bluetooth	439	515	13
image-compression	599	690	12
heart-valve	835	1589	17
tissue-engineering	2505	7443	10

keyword	**universities**	**res. inst.**	**companies**	**hospitals**
bluetooth	0.57	0.13	0.27	0.03
image-compression	0.69	0.13	0.14	0.04
heart-valve	0.42	0.18	0.07	0.34
tissue-engineering	0.44	0.19	0.17	0.20

further information: countries, geo-tagging

3. Company networks

Measuring long-range correlations

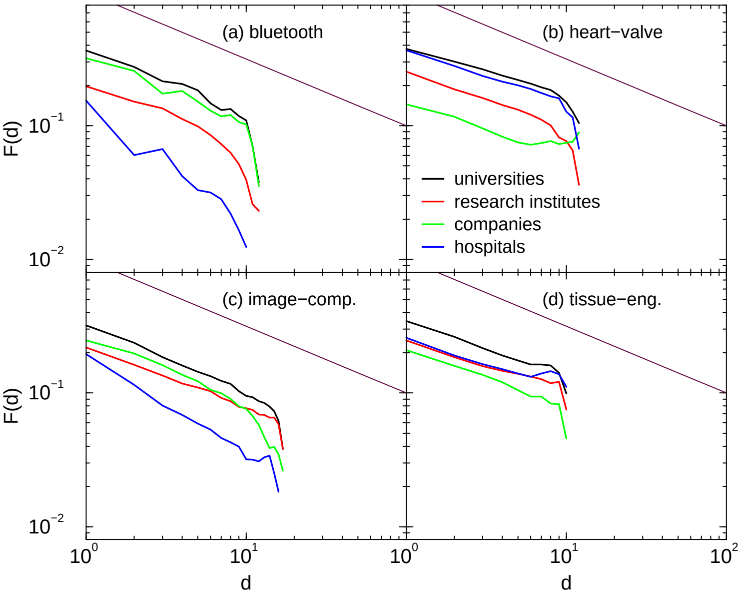
Applying fluctuation analysis

("fluctuation" in space – not in time)

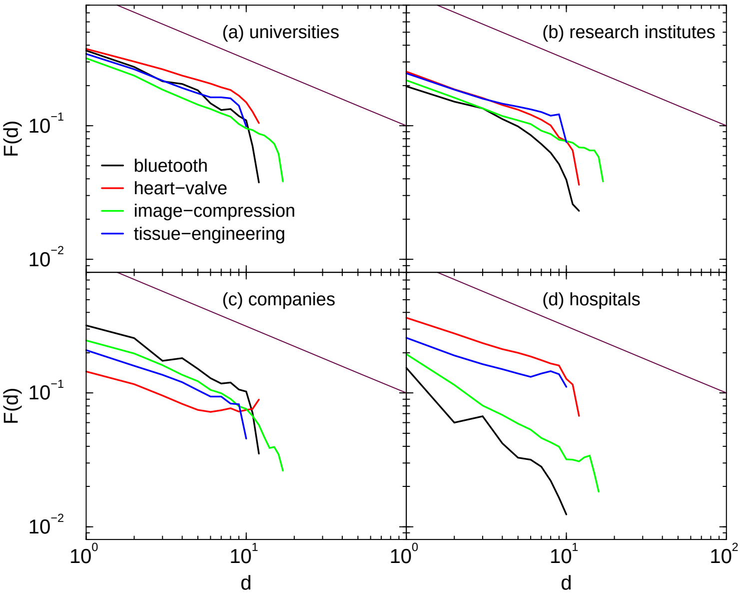
- each node is associated to 1 of 4 types
  (universities, research institutes, companies, hospitals)
- pick one type
- replace the attribute of each node with
  - "1" if type matches
  - "0" if it does not match
- network fluctuation analysis on the 0/1-nodes attributes
- repeat for all 4 types

correlation patterns are reflected in fluctuation functions

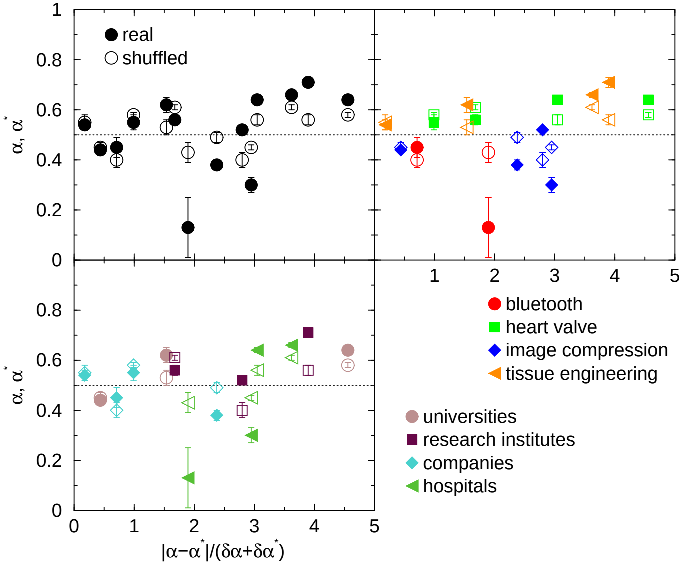
Results: fluctuation functions



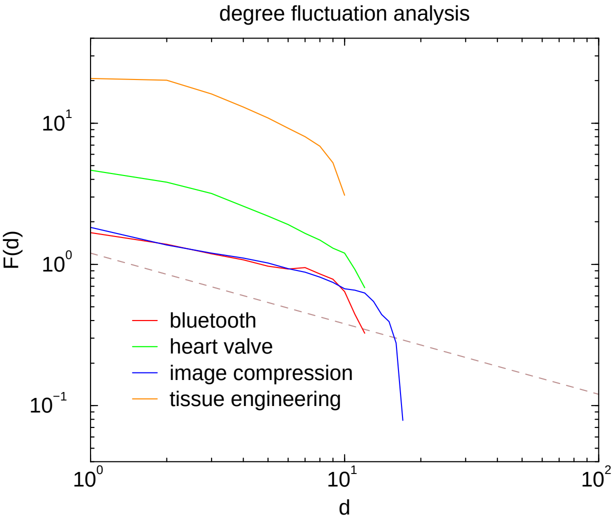
Results: fluctuation functions



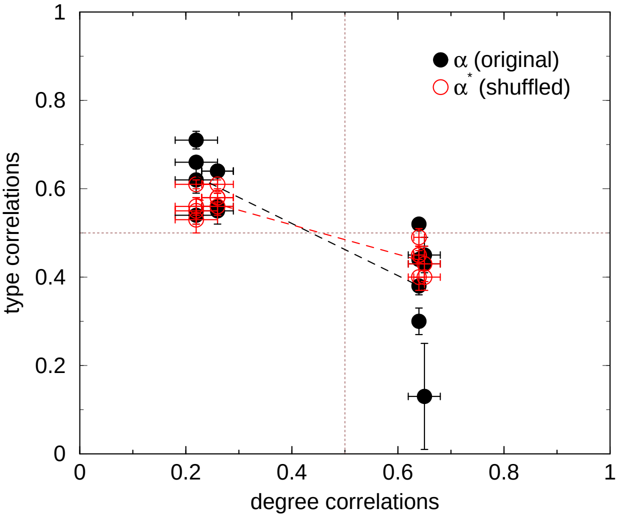
Results: fluctuation exponents



Results: degree fluctuation functions



Results: exponent cross-plot



Findings

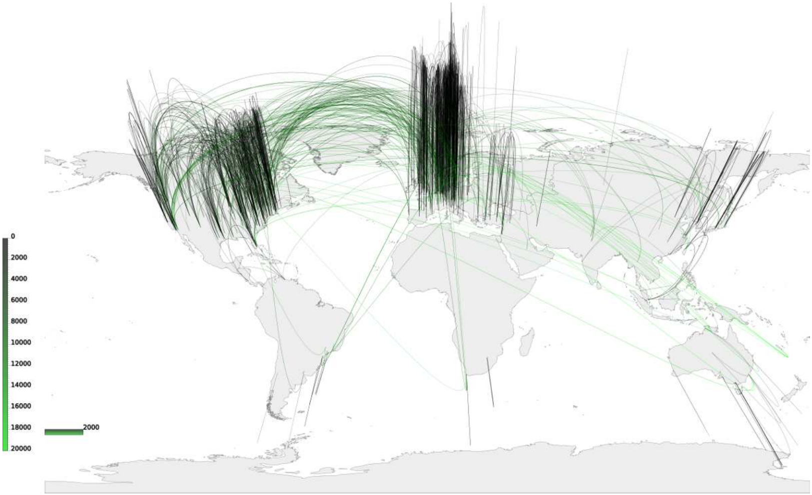
- scaling in the spatial fluctuations of institutions
- both, positive and negative correlations
- strength seems to depend on the network (i.e. topic)
- shuffling reduces correlations
- also correlations in the degree
- fluctuation exponents of type and degree seem to be anti-correlated

4. Spatial embedding

Motivation

- role of *proximity* is discussed in geography, economics, and innovation research
- faster and cheaper communication and traveling
- "shrinking world"
- to which extent reflected in company networks, i.e. collaboration networks?

map: heart-valve



example: tissue engineering

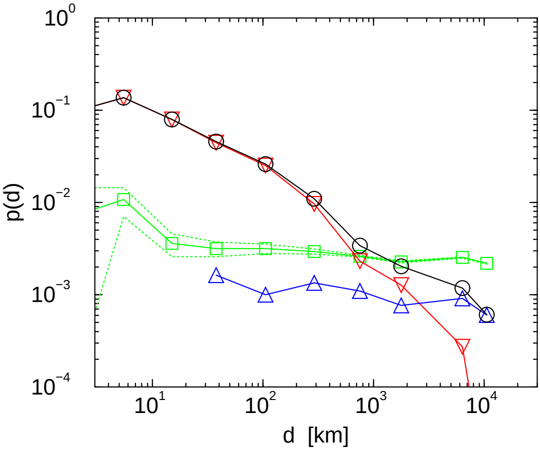
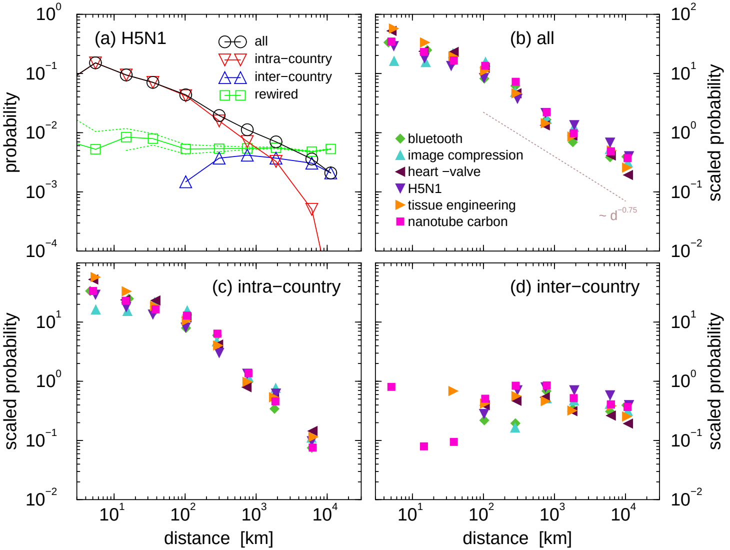


Figure: probability to have a link at distance d – not distribution of distances

collapse


(a) H5N1
all
intra−country
inter−country
rewired
probability
(b) all
bluetooth
image compression
heart −valve
H5N1
tissue engineering
nanotube carbon
~ d^−0.75
scaled probability
(c) intra−country
(d) inter−country
distance [km]
distance [km]

Thank you for your attention!

Acknowledgment:

- Organizers of the conference
- German Academic Exchange Service (DAAD)

Publications:

- Long-range degree-correlations: Rybski D. et al. EPL 2010
- Company networks: Rybski D. et al. (in prep.)
- Myth of global science collaboration: Hennemann S. et al. (submitted 2011)

http://diego.rybski.de/