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
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
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.
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?
We propose a Degree Fluctuation Analysis (kFA):
We propose a Degree Fluctuation Analysis (kFA):

Consider the shortest path between the nodes $i$ and $j$ (if it is not unique, we consider an arbitrary one).
We propose a Degree Fluctuation Analysis ($k$FA):

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 $d_{ij}$. 
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 $(k_l)$. 
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, \( C(d) \sim \langle (k_i - \langle k \rangle)(k_j - \langle k \rangle) \rangle |d\rangle \), between the degrees at distance \( d \) scales as

\[
C(d) \sim d^{-\gamma} \quad \text{for positive correlations (assortative)} \quad \text{or} \\
C(d) \sim -(d^{-\gamma}) \quad \text{for negative correlations (disassortative)}
\]

then we expect

\[
F(d) \sim d^{\alpha_k},
\]

where \( \alpha_k = -\gamma/2 \).

Fluctuation exponent differs by 1 from usual Hurst-like exponent: \( \alpha = \alpha_k + 1 \).
If the covariance, $C(d) \sim \langle (k_i - \langle k \rangle)(k_j - \langle k \rangle) \rangle |d\rangle$, between the degrees at distance $d$ scales as

- $C(d) \sim d^{-\gamma}$ for positive correlations (assortative)
- $C(d) \sim -(d^{-\gamma})$ 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)
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: $d_f = 2$
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


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

- **Fractal dimension:** $d_f = \frac{\ln(2m+x)}{\ln(3-2e)}$
- **Degree distribution:** $p(k) \sim k^{-\left(1 + \frac{\ln(2m+x)}{\ln m}\right)}$

---

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.
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 $n = 5$: $\alpha \sim e^\epsilon$, $\epsilon \approx 0.2$. 
4. Real-world networks

Figure: Degree fluctuation functions for real-world networks.
Comparison with fractal dimension

<table>
<thead>
<tr>
<th>network</th>
<th>$kFA$</th>
<th>$box\ covering$ (Song et al.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>human homology</td>
<td>$\alpha_k \simeq -0.74$</td>
<td>$d_f \simeq 2.5$</td>
</tr>
<tr>
<td>pin yeast</td>
<td>$\alpha_k \simeq -0.53$</td>
<td>$d_f \simeq 2.2$</td>
</tr>
<tr>
<td>homology</td>
<td>$\alpha_k \simeq -0.83$</td>
<td>$d_f \simeq 2.5$</td>
</tr>
<tr>
<td>metabolic</td>
<td>$\alpha_k \simeq -0.88$</td>
<td>$d_f \simeq 3.3$</td>
</tr>
</tbody>
</table>
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
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 collection (Hennemann S. et al., submitted 2011)**

Publications from isi web of science (2004-2008)

Network of institutions with at least 1 common publication

<table>
<thead>
<tr>
<th>keyword</th>
<th>nodes</th>
<th>links</th>
<th>diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>bluetooth</td>
<td>439</td>
<td>515</td>
<td>13</td>
</tr>
<tr>
<td>image-compression</td>
<td>599</td>
<td>690</td>
<td>12</td>
</tr>
<tr>
<td>heart-valve</td>
<td>835</td>
<td>1589</td>
<td>17</td>
</tr>
<tr>
<td>tissue-engineering</td>
<td>2505</td>
<td>7443</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>keyword</th>
<th>universities</th>
<th>res. inst.</th>
<th>companies</th>
<th>hospitals</th>
</tr>
</thead>
<tbody>
<tr>
<td>bluetooth</td>
<td>0.57</td>
<td>0.13</td>
<td>0.27</td>
<td>0.03</td>
</tr>
<tr>
<td>image-compression</td>
<td>0.69</td>
<td>0.13</td>
<td>0.14</td>
<td>0.04</td>
</tr>
<tr>
<td>heart-valve</td>
<td>0.42</td>
<td>0.18</td>
<td>0.07</td>
<td>0.34</td>
</tr>
<tr>
<td>tissue-engineering</td>
<td>0.44</td>
<td>0.19</td>
<td>0.17</td>
<td>0.20</td>
</tr>
</tbody>
</table>

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

- (a) bluetooth
- (b) heart-valve
- (c) image-comp.
- (d) tissue-eng.

Graphs show the distribution of fluctuation functions $F(d)$ for different sectors:

- Universities
- Research institutes
- Companies
- Hospitals

The x-axis represents the distance $d$ on a logarithmic scale, and the y-axis represents $F(d)$ also on a logarithmic scale.
Results: fluctuation functions

(a) universities

F(d)

(b) research institutes

F(d)

(c) companies

F(d)

(d) hospitals

F(d)

- bluetooth
- heart-valve
- image-compression
- tissue-engineering
Results: fluctuation exponents

\[
\alpha, \alpha^* \quad \text{universities} \quad \text{research institutes} \quad \text{companies} \quad \text{hospitals}
\]

\[
|\alpha - \alpha^*|/ (\delta \alpha + \delta \alpha^*)
\]

- Real data
- Shuffled data

- Bluetooth
- Heart valve
- Image compression
- Tissue engineering

- Universities
- Research institutes
- Companies
- Hospitals
Results: degree fluctuation functions

degree fluctuation analysis

- bluetooth
- heart valve
- image compression
- tissue engineering
Results: exponent cross-plot

\[
\begin{align*}
\text{degree correlations} & : 0, 0.2, 0.4, 0.6, 0.8, 1 \\
\text{type correlations} & : 0, 0.2, 0.4, 0.6, 0.8, 1
\end{align*}
\]
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
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?
Figure: probability to have a link at distance $d$ – not distribution of distances
(a) H5N1

(b) all

(c) intra−country

(d) inter−country

- probability
- scaled probability
- distance [km]

- H5N1
- intra−country
- inter−country
- rewired
- bluetooth
- image compression
- heart−valve
- H5N1
- tissue engineering
- nanotube carbon

$\sim d^{-0.75}$
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/