Emergence of communities in social networks

Jukka-Pekka Onnela

Department of Physics & Saïd Business School
University of Oxford

CABDyN Seminar Series
Saïd Business School, University of Oxford
19/2/2008
Emergence of communities in social networks?

- Model of large social networks with focus on how communities emerge
- Model should reproduce characteristic properties AND communities
- Start from large-scale empirical social network


Overview

1. Social networks
2. Empirical social network
3. Modelling social networks
4. Conclusion
Social networks

Social network paradigm in the social sciences: Social life consists of the flow and exchange of norms, values, ideas, and other social and cultural resources channelled through the social network.

Perspective:
- Focus on very large networks
- Focus on statistical properties
- Complex networks & statistical mechanics

Photo from http://defiant.corban.edu/gtipton/net-fun/iceberg.html
Social networks

Traditional approach:
- Data from questionnaires; $N \approx 10^2$
- Scope of social interactions wide
- Strength based on recollection

New approach:
- Electronic records of interactions; $N \approx 10^6$
- Scope of social interactions narrower
- Strength based on measurement

Constructed network is a proxy for the underlying social network
1. Social networks
2. Empirical social network
3. Modelling social networks
4. Conclusion
Constructing empirical network

Data
- One operator in a European country, 20% coverage
- Aggregated from a period of 18 weeks
- Over 7 million private mobile phone subscriptions
- Voice calls within the operator
- Require reciprocity of calls for a link
- Quantify tie strength (link weight)

Aggregate call duration $w_{ij}^D$
Total number of calls $w_{ij}^N$
About (social) network visualisation

- Snowball sampling (distance!)
- Bulk nodes & surface nodes
- Majority are surface nodes
- Neighbour visibility
## Network statistics

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>degree $k$</td>
<td>3.3</td>
<td>2.5</td>
<td>144</td>
</tr>
<tr>
<td>weight $w^N$</td>
<td>15.4</td>
<td>37.3</td>
<td>3,610</td>
</tr>
<tr>
<td>weight $w^D$</td>
<td>41 min</td>
<td>206 min</td>
<td>663 h</td>
</tr>
<tr>
<td>strength $s^N$</td>
<td>51</td>
<td>75</td>
<td>3,644</td>
</tr>
<tr>
<td>strength $s^D$</td>
<td>135 min</td>
<td>386 min</td>
<td>690 h</td>
</tr>
</tbody>
</table>

degree = # of links
Local structure

Weak ties hypothesis*: Relative overlap of two individual's friendship networks varies with the strength of their tie to one another

Define overlap $O_{ij}$ of edge $(i,j)$ as the fraction of common neighbours

Average overlap increases as a function of (cumulative) link weights

* M. Granovetter, The strength of weak ties, AJS 78, 1360 (1973)
Global structure

- Probe the global role of links of different weight and local topology
- Approach of physicists (and children): Break to learn!
- Thresholding (percolation): Remove links based on their weight
- Control parameter $f$ is the fraction of removed links
  - Initial network ($f=0$); isolated nodes ($f=1$)
Global structure

Initial connected network (f=0), small sample
⇒ All links are intact, i.e. the network is in its initial stage
Global structure

Decreasing weight thresholded network (f=0.8)
⇒ 80% of the strongest links removed, weakest 20% remain
Initial connected network ($f=0$), small sample
⇒ All links are intact, i.e. the network is in its initial stage
Global structure

Increasing weight thresholded network (f=0.8)
⇒ 80% of the weakest links removed, strongest 20% remain
Global structure

- **Qualitative** difference in the **global** role of weak and strong links
  - Phase transition when **weak** ties are removed first \( f_c(\infty) \neq 1 \)
  - No phase transition when **strong** ties are removed first \( f_c(\infty) = 1 \)
  - Suggests a point of division between weak and strong links (\( f_c \))

\[ w_c = P_{\text{cum}}^{-1}(0.80) \approx 27 \text{ min} \]

"globally connected" phase
"disconnected islands" phase

**Order parameter** \( R_{\text{LCC}} \)
- Def: fraction of nodes in LCC

**Susceptibility** \( S \)
- Def: average cluster size (excl. LCC)

\[
S = \sum_{s < s_{\text{max}}} n_s s^2 / \sum_{s < s_{\text{max}}} n_s s; \quad \tilde{S} = \sum_{s < s_{\text{max}}} n_s s^2 / N; \quad C_i = t_i / 2k_i(k_i - 1)
\]

\( f_c^w(\infty) = 0.80 \pm 0.04 \)
Summary of empirical study

- Communities have mostly strong ties within (WTH)
- Communities are interconnected mostly with weak ties (percolation)
1. Social networks
2. Empirical social network
3. Modelling social networks
4. Conclusion
Social networks appear to have some “universal features”

Can these features be reproduced with a simple microscopic model?

Network sociology: How individual microscopic interactions translate into macroscopic social systems

Statistical mechanics: How individual microscopic interactions translate into macroscopic (physical) systems
Intro to modelling

Internet & web => Simple rules work

$P(k) \sim k^{-\gamma}$
Intro to modelling

- A weighted model of social networks with focus on emergence of communities (mesoscopic structures) from microscopic rules
- Fixed number of nodes $N$
- Aim to reproduce characteristics features, no fitting to data
- Regression models in sociology
- No claim for a grand unified theory of social networks
Microscopic rules $\rightarrow$ Mesoscopic structure

- **Topology**
- **Topology & weights**

δ = 0

δ > 0
Microscopic rules in the model

- **Local attachment (LA)**

- **Global (random) attachment (GA)**
  \[ k_i = 0 \implies P(i, j) = 1; w_{ij} = w_o = 1 \]
  \[ k_i > 0 \implies P(i, j) = p_r; w_{ij} = w_o \]

- **Node deletion (ND)**
  \[ k_i > 0 \implies P(k_i = 0) = p_d \]
Microscopic rules in the model

Local attachment (LA)

(1) Weighted local search / reinforcement
\[ P(i \rightarrow j) = \frac{w_{ij}}{s_i} \]
\[ P(j \rightarrow k) = \frac{w_{jk}}{s_j - w_{ij}} \]
\[ w_{ij} \rightarrow w_{ij} + \delta \]
\[ w_{jk} \rightarrow w_{jk} + \delta \]

(2a) If (i,j,k) does not exist \rightarrow Triangle formation
\[ P(i, j, k) = p_\Delta \]
\[ w_{ik} = w_{0} = 1 \]

(2b) If (i,j,k) exists \rightarrow Triangle reinforcement
\[ w_{ik} \rightarrow w_{ik} + \delta \]
Microscopic rules in the model

- Summary of the model
  - Weighted local search for new acquaintances
  - Reinforcement of popular links & Triangle formation
  - Unweighted global search for new acquaintances

- Parameters
  - $\delta$
    - Free weight reinforcement parameter
  - $p_d = 10^{-3}$
    - Sets the time scale of the model $\langle \tau_N \rangle = p_d^{-1}$
  - $p_r = 5 \times 10^{-4}$
    - Global connections; Not sensitive
  - $p \Delta$
    - Adjusted w.r.t. $\delta$ to keep $\langle k \rangle$ constant
Model mechanisms vs. sociology

Network sociology*
- Cyclic closure
- Exponential decay
- Focal closure
- Independent of distance
- "Sample window"

Model
- Local attachment (LA)
- Global attachment (GA)
- Node deletion (ND)

Basic characteristics

(a) Fat-tailed degree distribution
(b) High clustering
(c) Assortative
(d) Small world

\[ \delta = 1 \]
\[ \delta = 0.5 \]
\[ \delta = 10^{-3} \]
\[ \delta = 0 \]

Alizations of \( N = 5 \times 10^4 \) networks. Values of \( \delta \) are 0 (□), \( 1 \times 10^{-3} \) (•), \( 1 \times 10^{-2} \) (○), 0.1 (△), 0.5 (▽), and 1 (○).
Local structure

Empirical

\[
P_{\text{cum}}(w^D_{ij}) \quad \text{and} \quad \langle O[P_c(w^N)] \rangle = \langle O[P_c(w^N)] \rangle.
\]

Model

\[
\delta = 1, \quad \delta = 0.5, \quad \delta = 10^{-3}, \quad \delta = 0
\]
Global structure

- Weak ties hypothesis (WTH)*: implies weight-topology correlations: Ties within communities are strong, ties between communities are weak
- Explore weight-topology correlation with link percolation
- Control parameter $f \in [0, 1]
- Order parameter $R_{LCC} \in [0, 1]

*M. Granovetter, “The Strength of Weak Ties”, The American Journal of Sociology 78, 1360 (1973)
Global structure

Small  \( \delta < 0.1 \)  \( \delta = 10^{-3} \)  \( \delta = 0 \)
- Network disintegrates at the same point for weak/strong link removal
- Incompatible with WTH

Large  \( \delta > 0.1 \)  \( \delta = 1 \)  \( \delta = 0.5 \)
- Network disintegrates at different points
- WTH compatible community structure

Weak go first  Strong go first
Communities by inspection

- Average number of links constant $\langle L \rangle = N \langle k \rangle / 2$
  - => All changes in structure due to reorganisation of links
- Increasing $\delta$ traps walks in communities, further enhancing trapping effect
  - => Clear communities
- Triangles accumulate weight and act as nuclei for communities
Communities by k-clique method

- Use k-clique algorithm / definition for communities*
- Focus on 4-cliques (smallest non-trivial cliques)
- Relative largest community size $R_{k=4} \in [0, 1]$
- Average community size (excl. largest) $\langle n \rangle$
- Observe clique percolation through the system for small $\delta$
- Increasing $\delta$ leads to condensation of communities

Is community size distribution stable?

- Consider community $k$ with size $N_k$
- In the large $\delta$ regime, most local random walks remain in the initial community, resulting in stable distribution

$$\frac{dN_k}{dt} = -p_d N_k + p_d N \frac{N_k}{N} = 0$$

- Community formation happens in transient state
- A triangle accumulating weight acts as a nucleus for the emerging community
- Rate of deleting nodes within the community
- Rate at which new nodes will join the community during subsequent LA steps
1. Social networks
2. Empirical social network
3. Modelling social networks
4. Conclusion
Conclusion

- Local coupling between network topology and tie strengths (WTH)
- Weak ties (PT) are qualitatively different from strong ties (no PT)
- Model: essential characteristics & local & global properties
- Need focal & cyclic closure & sufficient reinforcement of connections
- Communities result from initial structural fluctuations that become amplified by repeated application of the microscopic processes
References


See also Science 314, 914 (2006).

See http://www.physics.ox.ac.uk/users/Onnela/

THANK YOU!