Analysis and Modeling Large Social Networks

János Kertész¹,²
&
Jussi Kumpula², Jukka-Pekka Onnela²,³, Jari Saramäki², Kimmo Kaski², Gábor Szabó⁴, Albert-László Barabási⁴,⁵, David Lazer⁴

¹Budapest University of Technology and Economics, ²Helsinki University of Technology, ³University of Oxford, ⁴Harvard University, ⁵Northeastern University
Outline

**Motivation**
- Social networks; structure and socio-dynamics
- Complex networks; modular structure and weighted links; processes on networks

**Empirical analysis: Mobile phone network**
- Overall statistics
- The importance of weak links and global consequences
- Spreading phenomena

**Modeling social networks**
- Weight-topology correlations
- Network formation processes
- Comparison with empirical observations
Motivation

Research questions:
- Why large-scale social networks?
- Studying collective social phenomena requires large systems
- Examples: diffusion processes, spreading processes (epidemiology), opinion formation, evolution of language, trade etc.

- How to collect empirical observations?
  - Mobile phone records: Proxy of the social interactions
  - Weighted network construction and analysis

- How to model?
  - Using simple principles from sociology reproduce findings
  - How are weights and topology related?
Social networks

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

Nodes = people \( (N = 3 - 10^9) \)

Links = social interactions (< 150/node)
- Represent the “underlying” social network (a *knows* b)
- Represent real interactions between a and b (phone calls, emails…)
- Weight represents interaction strength or intensity

Properties: Structure -- Function -- Response

Methods: Analysis -- Modeling -- Simulation
Construction of 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
Mobile communication:
social interaction network

Aim to design models to reproduce structural properties of the network and simulate dynamical processes in it.

**Structure:**
- Cohesive groups -> high clustering: communities with dense internal and sparse external connections
- Assortativity: popular people know other popular people
- Strength of weak ties hypothesis (Granovetter)

**Functional properties:**
- Structure-affected: information spreading, opinion formation, epidemics

Onnela et al. PNAS, 104, 7332 (2007)

4.6 $10^6$ nodes
7.0 $10^6$ links
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>
Social network

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</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>3610</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>3644</td>
</tr>
<tr>
<td>Strength $s^D$</td>
<td>135 min</td>
<td>386 min</td>
<td>690 h</td>
</tr>
</tbody>
</table>

Clustering coefficient

$C(k_i) = \frac{2t_i}{k_i(k_i - 1)}$

Degree-degree correlations

$K_{nn,i}^w = \frac{1}{s_i} \sum_{j \in N(i)} w_{ij} k_j$

Assortativity: high degree nodes connect to other high degree nodes

Popular people know other popular people
Social network - empirical structure

Strength of weak ties hypothesis (Granovetter):
Tie strength between two people increases with the overlap of their friendship circles

Relative neighborhood overlap

\[ O_{ij} = \frac{n_{ij}}{(k_i - 1) + (k_j - 1) - n_{ij}} \]

Cumulative weight

\[ P_{\text{cum}}(w') = \sum_{w \leq w'} P(w) \]

- Hypothesis verified -

Call network

Same but randomly reshuffled weights

(a) \( O_{ij} = 0 \)

(b) \( O_{ij} = 1/3 \)

(c) \( O_{ij} = 1 \)
**Order parameter** $R_{LCC}$
- Fraction of nodes in LCC
- LCC is robust, collapses when $f \geq 0.80$

**Susceptibility** $S$
- Average cluster size (excl. LCC)
- Divergence? $\Rightarrow$ Global role of links?

**Average shortest path length** $\langle l \rangle$
- Number of links along shortest path
- Removing weak links leads to longer paths

**Average clustering coefficient** $\langle C \rangle$
- Fraction of interconnected neighbors
- Removing strong links decreases $\langle C \rangle$
- WL removal 'invisible' locally; compare $R_{LCC}$

Red: remove weak first; Black: remove strong first
Diffusion of information

Knowledge of information diffusion based on unweighted networks

Use the present network to study diffusion on a weighted network:
Does the local relationship between topology and tie strength have an effect?

Spreading simulation: infect one node with new information

1. Empirical: \( p_{ij} \propto w_{ij} \)
2. Reference: \( p_{ij} \propto \langle w \rangle \)

Spreading significantly faster on the reference (average weight) network

Information gets trapped in communities in the real network
Communities with strong links connected by weak links

Has impact on the global structure

What is the interplay btw weight and community formation?
Complex network properties

- **Structured as modules or communities** - groups of nodes with more internal than external links:
  - Meso-scale structure can play a definite functional role
  - How do the communities emerge?

- **Weighted links** - interaction between a pair of nodes characterized by its existence and the varying strength assigned to it:
  - E.g. traffic, metabolic or correlation based networks
  - Affect the properties or function of the networks, e.g., disease spreading, synchronisation dynamics of oscillators and motif statistics weights
  - Influence the formation of topology and communities
Towards complex network modelling

- Several methods for community detection, but few models to produce communities from microscopics
- **Key question in sociology: Emergence of mesoscale communities from microscale mechanisms**
- **HERE:** A weighted model of social networks to produce communities from microscopics
- **Weights generated dynamically and shape the developing topology:** weights ↔ topology interplay
**Network formation processes**

- **Question:** How the microscale mechanisms translate to forming mesoscale communities and macroscale system

- **Social networks satisfy the weak tie hypothesis:** weak links keep the network connected whereas strong links are associated with communities (Granovetter)

- **Network sociology**: Two fundamental mechanisms for tie formation -> network evolution:
  - **Cyclic closure** form ties with one's network neighbors - "friends of friends"; its probability decreases exponentially as a function of the geodesic distance
  - **Focal closure** form ties independently of the geodesic distance through shared activities (hobbies etc.)

- **Simple scenario:** New ties are created preferably through strong ties, every interaction making them even stronger

Weighted social network model

Modelling how the people get acquaintances with local and global search mechanisms*:

- **Fixed size network of N nodes**
  - Internal structural changes faster than changes in the size of the network

- **Network subject to following dynamics:**
  - Local weighted search for new acquaintances and reinforcement of popular links
  - Global search by creation of random links
  - Random removal of nodes

Microscopic rules in the model

**Local attachment (LA)**

1. **Weighted local search / reinforcement**
   - Choose neighbor $j$ of $i$ with probability $(w_{ij}/s_i)$
   - Choose neighbor $k$ of $j$ with probability $(w_{jk}/(s_j - w_{ij}))$
   - Reinforce: $w_{ij} \rightarrow w_{ij} + \delta$
   - Reinforce: $w_{jk} \rightarrow w_{jk} + \delta$

2a) **If** $(i,j,k)$ **does not exist:**
   - With probability $p_\Delta$ create link $w_{ik} = w_0 = 1$
   - $\Rightarrow$ Triangle formation $\Rightarrow$ Triadic closure

2b) **If** $(i,j,k)$ **exists:**
   - Reinforce: $w_{ik} \rightarrow w_{ik} + \delta$
   - $\Rightarrow$ Triangle reinforcement
Microscopic rules in the model

Local attachment (LA)
- Weighted search & reinforcement
- Triangle formation with prob. \( p_{\Delta} \)
- Triangle reinforcement

Global attachment (GA)
- If \( k_i = 0 \) create random link: \( w_{ij} = w_0 \)
- If \( k_i > 0 \) create with prob. \( p_r \) random link: \( w_{ij} = w_0 \)

Node deletion (ND)
- With prob. \( p_d \) node \( i \) is deleted: \( k_i = 0 \)

LA: Search favors friends
GA: Search beyond neighbors
ND: All links of a node deleted

Parameters:
- Initial weight: \( w_0 = 1 \)
- Weight increment: \( \delta \in [0,1] \)

(2a)
(2b)
Microscopic mechanisms in sociology

Network sociology
- Cyclic closure
  - Exponential decay for growing geodesic distance
- Focal closure
  - Distance independent
- “Sample window”

Network model
- Local attachment (LA)
  - Special case of cyclic closure: Triadic closure
- Global attachment (GA)
- Node deletion (ND)
Microscopic rules in the model

Summary of the model
- Weighted local search for new acquaintances
- Reinforcement of existing (popular) links
- Unweighted global search for new acquaintances
- Node removal, exp.link & weight lifetimes: $<\tau>=2 <\tau_w>=(p_d)^{-1}$

Model parameters
- $\delta$ Free weight reinforcement parameter
- $p_d = 10^{-3}$ Sets the time scale of the model $<\tau_N>=1/p_d$ (average node lifetime of 1000 time steps)
- $p_r = 5\times10^{-4}$ Global connections; results not sensitive for it (one random link per node during 1000 time steps)
- $p_\Delta$ Adjusted in relation to $\delta$ to keep $<k>$ constant (structure changes due to only link re-organisations)
Social network model

Samples of $N = 10^5$ network for variable weigh-increase $\delta$

Tie strength: weak $\rightarrow$ intermediate $\rightarrow$ strong tie

$\delta = 0$

No communities

$\delta = 0.1$

Communities start nucleating

$\delta = 0.5$

Communities forming

$\delta = 1$

Communities with dense & strong internal and sparse & weak external connections (cf. phone network)
Microscopic rules $\rightarrow$ Mesoscopic structure

<table>
<thead>
<tr>
<th>Microscopic</th>
<th>Macroscopic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topology</td>
<td>Topology &amp; weights</td>
</tr>
</tbody>
</table>

- $\delta = 0$

- $\delta > 0$ (small)

- $\delta > 0$ (large)
Communities by k-clique method

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

Community structure

- Is community size distribution stable?
- If most local random walks remain in the initial community ($\delta$ large), a simple argument shows that community size distribution is stationary:

$$\frac{dN_s}{dt} = -p_d N_s + p_d \frac{N_s}{N} = 0$$

- Community formation happens in transient state
- Triangles acts as nuclei for emerging community as a results of weight accumulation
Weight-topology correlation

**Weak ties hypothesis (WTH)***: The stronger the tie between nodes $i$ and $j$, the greater the overlap of their friendship circles:

- WTH implies weight-topology correlations: Ties within communities strong, ties between communities weak

**Explore weight-topology correlation with link percolation:**

- Control parameter - fraction of links removed: $f \in [0,1]$,
- Order parameter: $R_{LCC} \in [0,1]$,
- Normalized susceptibility: $\hat{s} = \sum n_s s^2 / N$

*M. Granovetter, “The Strength of Weak Ties”, The American Journal of Sociology 78, 1360 (1973)
Weight-topology correlation

Small $\delta < 0.1$
- Network disintegrates at the same point for weak/strong link removal
- Incompatible with WTH

Large $\delta > 0.1$
- Network disintegrates at different points
- WTH compatible community structure

Link percolation: Control parameter: $f \in [0,1]$, Order parameter $R_{LCC} \in [0,1]$, Normalized susceptibility: $\hat{s} = \sum n_s s^2 / N$
Link percolation analysis

**Model network**

- **Ascending link removal**
- **Descending link removal**

**Phone network**

- **Ascending & Descending**

Phase transition for ascending tie removal (weaker first)
As a model of social networks

(a) Skewed degree distribution
(b) High clustering
(c) Assortative
(d) Small world
(e) WTH compliant
Network characteristics

Model network

- Weight distribution
- Neighborhood overlap

Mobile phone network

Weight distribution

\[ P(w) = a(w + w_0)^{-1.9} e^{-w/w_c} \]

Neighborhood overlap

\[ \langle \theta \rangle_w \]
Conclusion

- Weak ties maintain network’s structural integrity; Strong ties maintain local communities; Intermediate ties mostly responsible for first-time infections.

- How can one efficiently search for information in a social network? “Go out of your community!”

- Social networks seem better suited to local processing than global transmission of information.

- Are there simple rules or mechanisms that lead to observed properties?

- Efficient modeling possible.
Model couples interaction strengths and network structure in a simple way.

Communities emerge after nucleating from structural fluctuation but only if link weight reinforcement is strong enough.

Focal closure & cyclic closure are not sufficient by themselves.

Model not only complies with the Weak Tie Hypothesis (weight-topology correlation), but suggests a plausible mechanism for it.

This mechanism may be applicable to other complex networks in modelling community formation?
References


www.phy.bme.hu/~kertesz
Marc Granovetter, Connections, 1990:

>>In the history of public speaking, there have been many famous denials. One sunny day in 1880, Karl Marx declared: "I am not a Marxist". On a less auspicious occasion in 1973, Richard Nixon insisted "I am not a crook". Neither Marx' nor Nixon's audience gave much credence to their denials, and you too may respond with disbelief when I tell you that "I am not a networker".<<

......

>>Instead, the slogan of the day will be "We are all networkers now".<<