Computational Sociology: Studies of *in vivo* / *in situ* Social Networks

Kimmo Kaski
Dept of Biomedical Engineering and Computational Science
Aalto University School of Science, FINLAND
&
CABDyN Complexity Center, Said Business School, University of Oxford, UK
1. Complexity of Social Systems
2. Analysis of Human Communication Networks
3. Modelling Social Networks
4. More Data Science Examples on Sociology:
   • Demographic data based Egocentric Social Networks
   • Modelling Opinion Formation in a co-evolving Society
Forms of Human Complexity

Brain & Cognitive system

Emergent properties of the system’s Structure & Function & Response

Data

Analysis & Modelling & Simulation

Is complexity in number?
FRUIT FLY: 13600 genes
C. ELEGANS: 19500 genes
HOMO SAPIENS: 23300 genes
ARABIDOPSIS (mustard): 27000 genes

Communication system:
Many non-identical elements linked with diverse interactions

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

**Dunbar**: “Egocentric” social networks consists of layers.

Total number of active human relations does not exceed 150 => **Dunbar number**

Evolutionary explanation: cognitive limit to the number of individuals with whom any person can maintain stable relationships (Social Brain Hypothesis).

- Support clique: (3-5 people)
- Sympathy group: (12-20 people)
- Band: (30-50 people)
- Clan: (150 people)
- Megaband: (500 people)
- Tribe: (1500 people)
In less than a generation deep changes in human behavior due to ICT coming to the hands of people:

- Availability
- Mobility
- Working
- Information gathering and learning
- Shopping and leisure
- Contacting habits and networking
- Privacy concept
- Social and public activity
- Games and gamification

--> Mobile & locality-independent “Twitter society”
Modern communication leaves detailed information about who with whom, when and where...

- Phone (mobile and fixed line)
- SMS, MMS
- Skype
- Email...

Activities leave behind electronic records or traces of:
- Commercial activities (eBay, point cards, credit cards, etc)
- Open collaborative environments (Wikipedia, gnu, etc)
- E-communities (Facebook, MySpace, etc)
- E-games (Roleplaying, Where is George, etc)...

In today’s Techno-Social Society:
**ICT -> Data -> Big Data -> Data Science**
What matters?

The way how elements/basic constituents are connected

To study we need to:

• Identify the skeleton of the system: a network?
• Learn about the topology (micro-, meso- and macro-scale structure)
• Uncover the relation between properties of the elements and the topology (e.g., strength of ties)
• Relate the structure to functions
• Describe dynamic processes and the influencing factors including structure
Scales of Social Networks

**Structural:**
- Friendships/Kin
- Groups/Communities
- Society

**Dynamical:**
- Social interaction
  - Events (e.g. calls)
- Dynamics of groups
  - Dynamics in networks
    - (e.g. group formation)
    - (e.g. rumor spreading)

Laws of sociality not known (yet?):
**How does microscopic translate to macroscopic?**
Social Science: Laws not Known

Properties:
- Structure
- Function
- Response

Methods:
- Analysis
- Modelling
- Simulation

- Electronic databases
- Computer speed
- Science of complex systems
- Social sciences

Computational social science
Empirical Approach to Social Complexity
Traditional & ICT-records based approach

Traditional approach:
- Data from questionnaires; $N \approx 10^2 - 10^3 \ldots$
- Scope of social interactions is wide
- **Strength based on individual recollection**

ICT-records’ (digital footprints’) approach:
- Electronic records of interactions; $N \approx 10^6 \ldots$
- Scope of social interactions is new but narrow
- **Strength based on measurement**
• Mobile phone operator data over a period of 2 years
• Millions of private subscriptions
• Voice calls / sms ‘ within the operator
• Lately also demographic data
• Require reciprocity of calls for a link
• Quantify link weights

Constructing a network from data
Proxy for the underlying Social Network

Sample from $4.6 \times 10^6$ nodes & $7.0 \times 10^6$ links

Formation of community structure

The strength of weak ties (M. Granovetter, 1973):

Hypothesis about the local scale (micro-) structure of the society:

1. “The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie.”

2. “The stronger the tie between A and B, the larger the proportion of individuals S to whom both are tied.”

Consequences on global scale (macro-) structure:
Society consists of strongly connected / “wired” communities linked by weak ties. The latter hold the society together.

M.S. Granovetter "The Strength of Weak Ties", American Journal of Sociology 78, 1360 (1973)
Structure and Tie Strength: Local Structure – Strength of Weak Ties

Relative neighbourhood overlap
\(k_i = \text{degree} \& n_{ij} = \# \text{ of triangles with edge } ij\)

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

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

Verifies Granovetter’s hypothesis (up to 95%)

GLOBAL STRUCTURE
Percolation / Threshold Analysis

Order parameter $R_{LCC}$
- Fraction of nodes in LCC

Susceptibility $S = \sum ns^2 / N$
- Average cluster size (not LCC)

Critical fraction $f_c$
- $f_c = 0.80; w_c = P^{-1}(0.80) = 27$ min

Role of weak and strong links different

Descending threshold: 20% weakest
Ascending threshold: 20% strongest remain

Dynamics of Spreading

Spreading simulation (SI): infect one node with new information

1. Empirical: $p_{ij} = a \cdot w_{ij}$

2. Reference: $p_{ij} = a \cdot w = constant (\approx 1)$

Spreading significantly faster on the reference network
Information gets trapped within network communities

Even ICT-aided human society not most effective in information spreading
Simulation: ‘Rumor’ Spreading in a Real Network

At network diffusion level time scale:

The rumor spreads through in 4 days
Can Social Systems be Modelled?

- Can modelling illuminate sociological questions? (Believe so)
- How parsimonious can these models be? (Quite)
- How can these models be validated? (Comparing)

Models are like maps - useful when contain details of interest and ignore others (Boyd & Richerson: Mathematical Models of Social …)

Lewis Carroll (Sylvie and Bruno concluded): “What do you consider the largest map that would be really useful?” “About six inches to the mile”. “Only six inches!” exclaimed Mein Herr. “We very soon got six yards to the mile. Then we tried hundred yards to the mile. And then came the grandest idea of all! We actually made a map of the country, on the scale of a mile to the mile! "Have you used it much?" I enquired. "It has never been spread out, yet," said Mein Herr. "The farmers objected: they said it would cover the whole country and shut out the sun-light! So now we use the country itself, as its own map, and I assure you it does nearly as well.”
“Truth is ever to be found in the simplicity, and not in the multiplicity and confusion of things.”

-- Sir Isaac Newton
“Everything should be made as simple as possible, but not simpler.”

-- Albert Einstein
“All models are wrong but some models are useful.”

-- George E.P. Box, statistician
Modelling how the people get acquaintances with local and global search mechanisms*:

- **Fixed size network of N nodes (individuals)**
  - 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

Two basic mechanisms for social tie formation*

- **Cyclic closure** forms ties with one's network neighbours - "friends of friends"
- **Focal closure** forms ties independently of the geodesic distance through shared activities (hobbies etc.)

Mechanisms of Friendship formation

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

**Network model**
- Local attachment (LA)
  - Special case of cyclic closure: Triadic closure
- Global attachment (GA)
- Node deletion (ND)
- Link reinforcement

Formation of a Social Network

Model (based on sociology): weighted local search (cyclic closure), global search (focal closure), node deletion & link reinforcement $\delta$

- $\delta = 0$: 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)

Similar Local & Global Structure

Overlap vs. tie strength

- \( \Delta=0 \)
- \( \Delta=0.01 \)
- \( \Delta=0.1 \)
- \( \Delta=0.5 \)
- \( \Delta=1 \)

Ascending link removal
- \( \delta < 0.1: \) WTH -
- \( \delta > 0.1: \) WTH +

Descending link removal

Order param.: \( R_{LCC} \)

Suscept.: \( \hat{s} = \frac{\Sigma n_i s_i^2}{N} \)

Clustering coeff.

Reproduces different roles of weak vs. strong ties

- "phase chance" for ascending link removal

Weak Ties Hypothesis behavior reproduced!
Model shows properties of real networks:

- Network statistics & community structure turn out to be similar
- Local structure: weak ties hypothesis verified
- Global structure: role of weak and strong ties different

Egocentric (in vivo) Networks: Sex Differences in Intimate Relationships

Age and gender correlations in friendships


**Homophily** – a strong tendency for individuals to associate with others whom they perceive as being similar to themselves in some way.

**Assortative mixing** by degree in social networks.

**Dissortative mixing** by genders in a network of sexual contacts.

Are there gender and age specific properties in making human relations?
Humans can handle roughly three times as many social contacts as apes.

**Social Brain Hypothesis**

**Group size vs. Neocortex ratio**

**Dunbar's Law**

Group Size $N = 5 \times (R^3)$

where $R = \text{relationship strength} \times \text{N relationships}$

**Social Brain Theory**

We study the innermost circle
The same as earlier, i.e. one mobile phone operator in a European country but more recent and detailed dataset:

- ~ 8 million service subscribers
- ~ 1.95 billion calls
- ~ 489 million text messages
- Covers 7 months period (2009)
- In addition demographic data:
  - Gender of subscriber
  - Age of subscriber
  - Postcode of subscriber

Communication events with their #, duration & timings through a single+ channel (calls+SMS) reflect a high degree of ego-alter sociality

-> Phone network serves as a proxy for social network
Carrying out initial data filtering we arrive at 3.2 million subscribers:

- 1.8 million males;
- 1.4 million females.

Gender variable $g = \begin{cases} +1 \text{ for males} \\ -1 \text{ for females} \end{cases}$

Balance: $\langle g \rangle = 0$

Here: $\langle g \rangle = 0.13$
Men and women between 18 and 45 have best friends of the opposite sex and the 2nd best friends are generally of the same sex.
Women are more focused on opposite sex relationships than men.
As people age their attention shifts from the spouse to the children.

Women are more active in maintaining family relationships.

The mother-daughter link is particularly strong.
Distribution of best friends by age

As people age their attention shifts from the spouse to the children

Women are more active in maintaining family relationships

The mother-daughter link is particularly strong.
SAMPLE OF EGO – ALTER NETWORK
Conclusions

- Special status of the best friend;
- Women are more focused on opposite-sex relationships than men are during the reproductively active period.
- As they age, women’s attention shifts from their spouse to their children, but in particular to their daughters.
- Women in particular switch individuals around in their preference rankings much more than men do. Men tend to keep a steadier pattern over a longer period.
- The data provide strong evidence for the importance of female matrilineal relationships in human social organization.