CABDyN Seminar
CABDyN Complexity Center, Said Business School, University of Oxford
Keble College, Oxford, 11th February, 2014

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

Kimmo Kaski

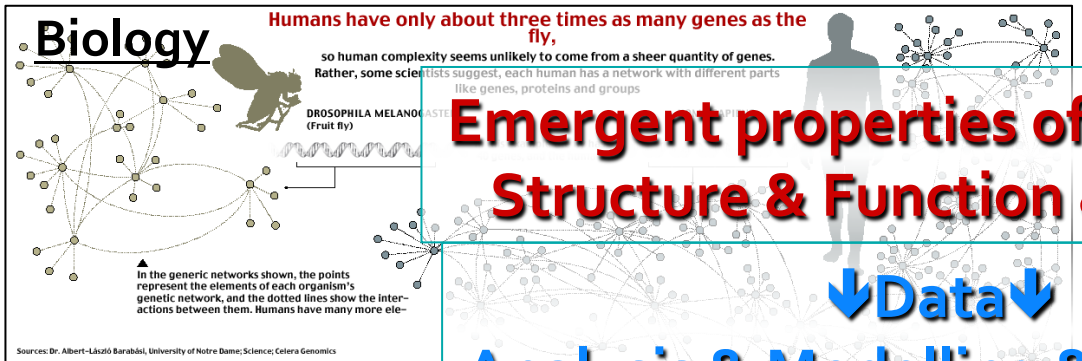
Dept of Biomedical Engineering and Computational Science
Aalto University School of Science, FINLAND
&
CABDyN Complexity Center, Said Business School,
University of Oxford, UK

Outline

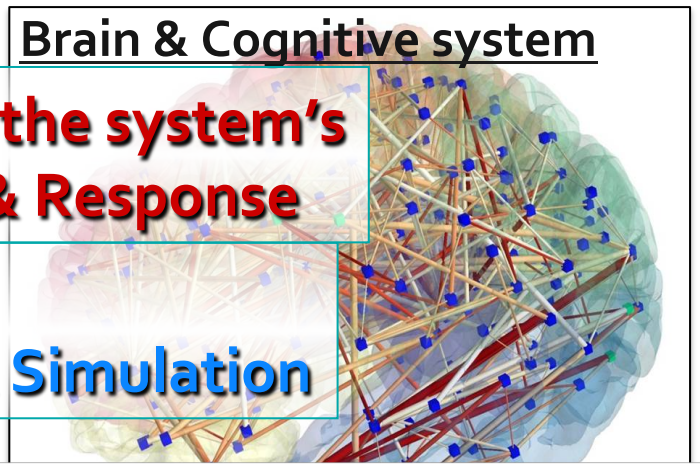
1. Complexity of Social Systems
2. Analysis of Human Communication Networks
3. Modelling Social Networks
4. More Data Science Examples on Sociology:
   - Demographic data based Egocentric Social Networks
   - Modelling Opinion Formation in a co-evolving Society

Forms of Human Complexity

Biology



Brain & Cognitive system



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

↓Data↓

Analysis & Modelling & Simulation

Is complexity in number?

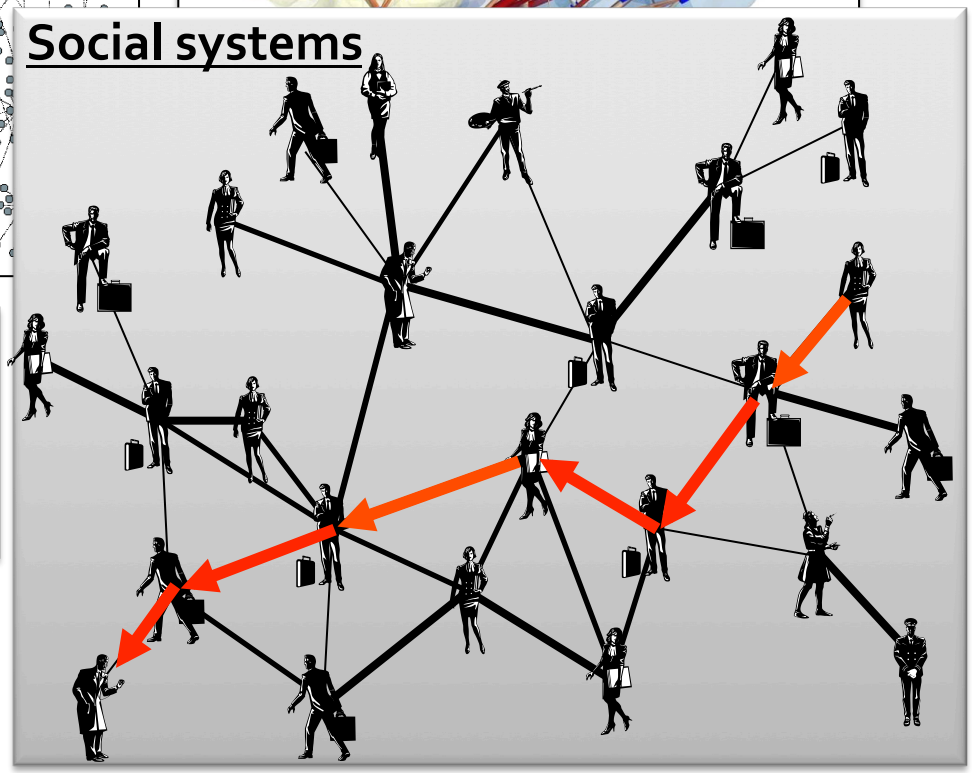
FRUIT FLY : 13600 genes

C. ELEGANS: 19500 genes

HOMO SAPIENS: 23300 genes

ARABIDOPSIS (mustard): 27000 genes

Social systems



Communication system:

Many non-identical elements linked with diverse interactions

↓

NETWORK

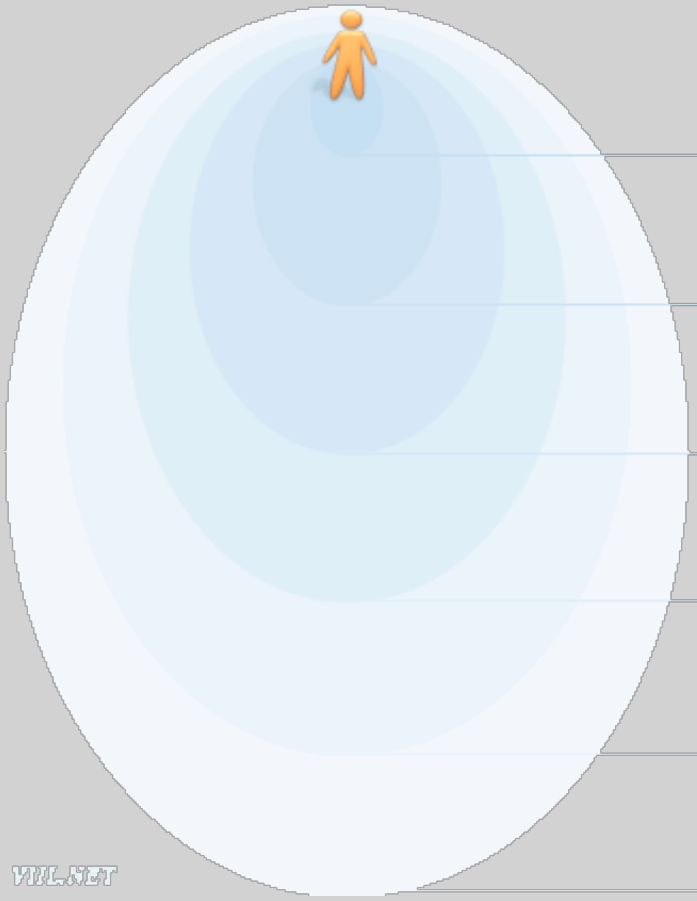
Paradigm of Sociality
Structure & Dynamics & their Interplay

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)



ICT Changes Social Complexity

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”

ICT: CHANGES OF METHODOLOGY

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

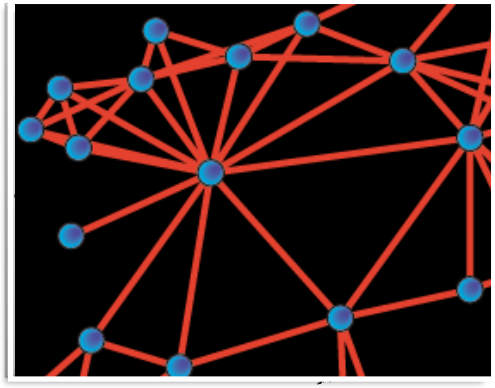
To study we need to:

- 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

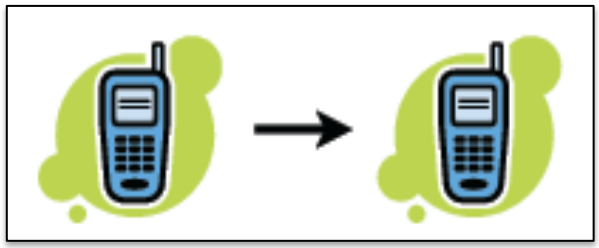
Structural:

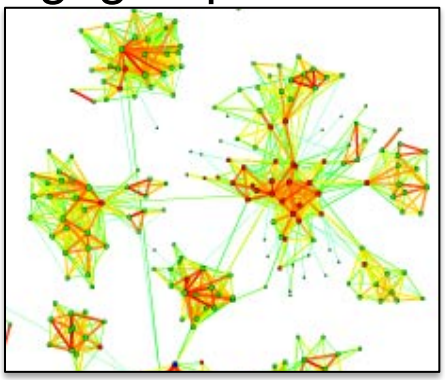
Friendships/Kin – Groups/Communities – Society - ...

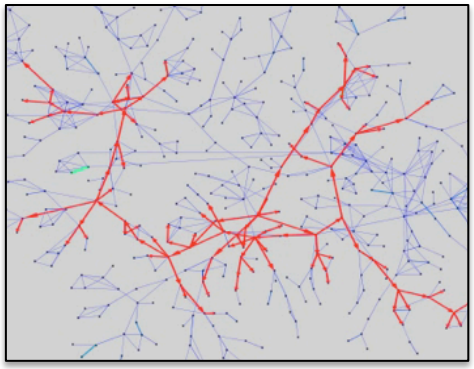


Dynamical:

social interaction – Dynamics of groups – Dynamics in networks
events (e.g. calls) (e.g. group formation) (e.g. rumor spreading)

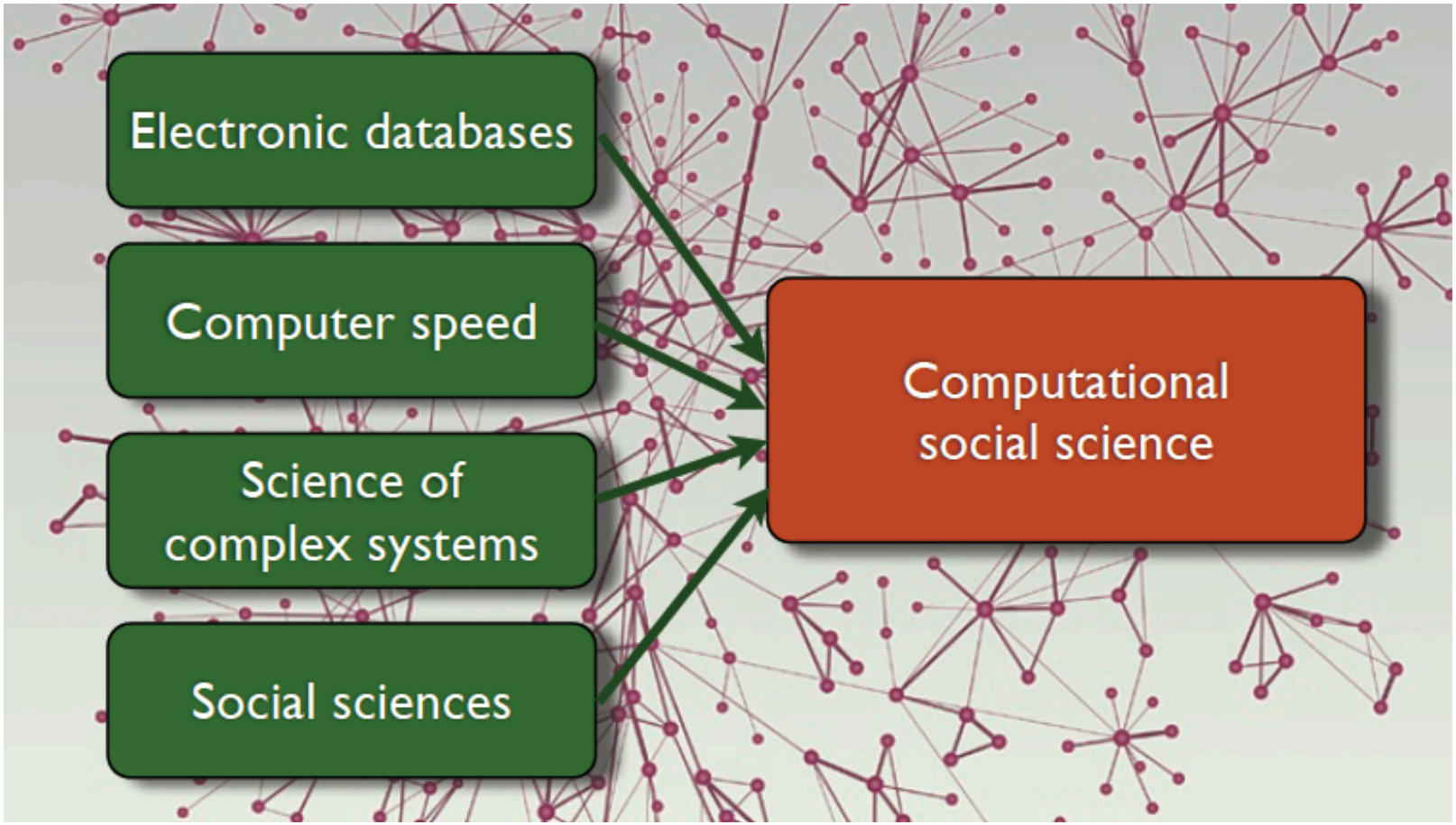






Laws of sociality not known (yet?):
How does microscopic translate to macroscopic?

Social Science: Laws not Known



Properties: Structure -- Function -- Response

Methods: Analysis -- Modelling -- Simulation

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

Traditional approach:

- Data from questionnaires; N ≈ 10^2 - 10^3 …
- Scope of social interactions is wide
- Strength based on individual recollection

How to measure?

I know him/her
We are on first name basis
We are friends
We are good friends
We are very good friends

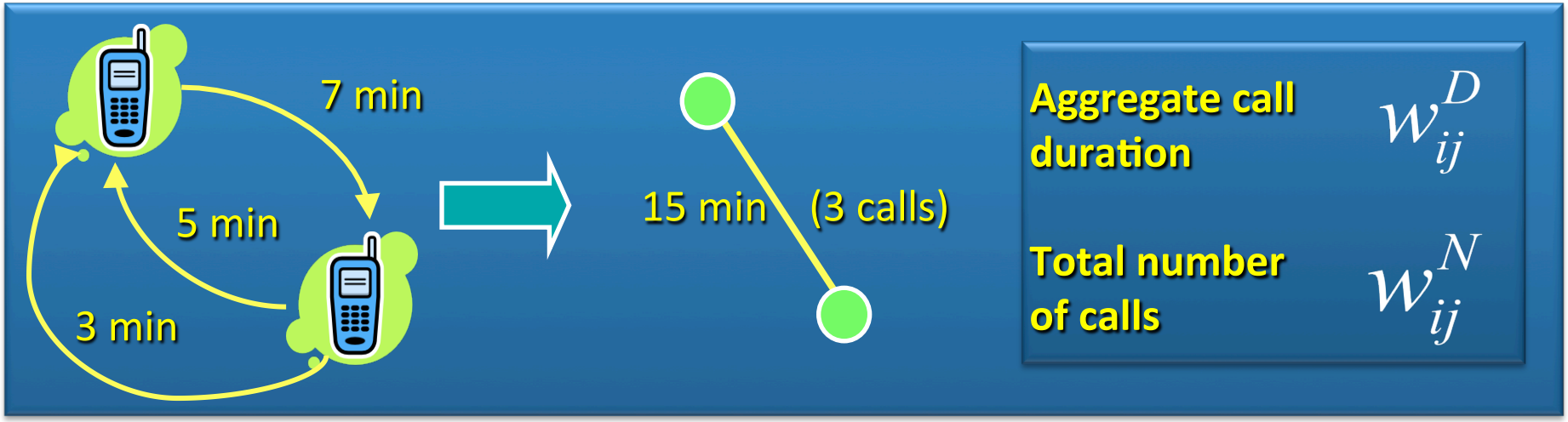
Scale?
Subjectivity?

ICT-records' (digital footprints') approach:

- Electronic records of interactions; N ≈ 10^6…
- Scope of social interactions is new but narrow
- Strength based on measurement

Constructing a network from data
Proxy for the underlying Social Network

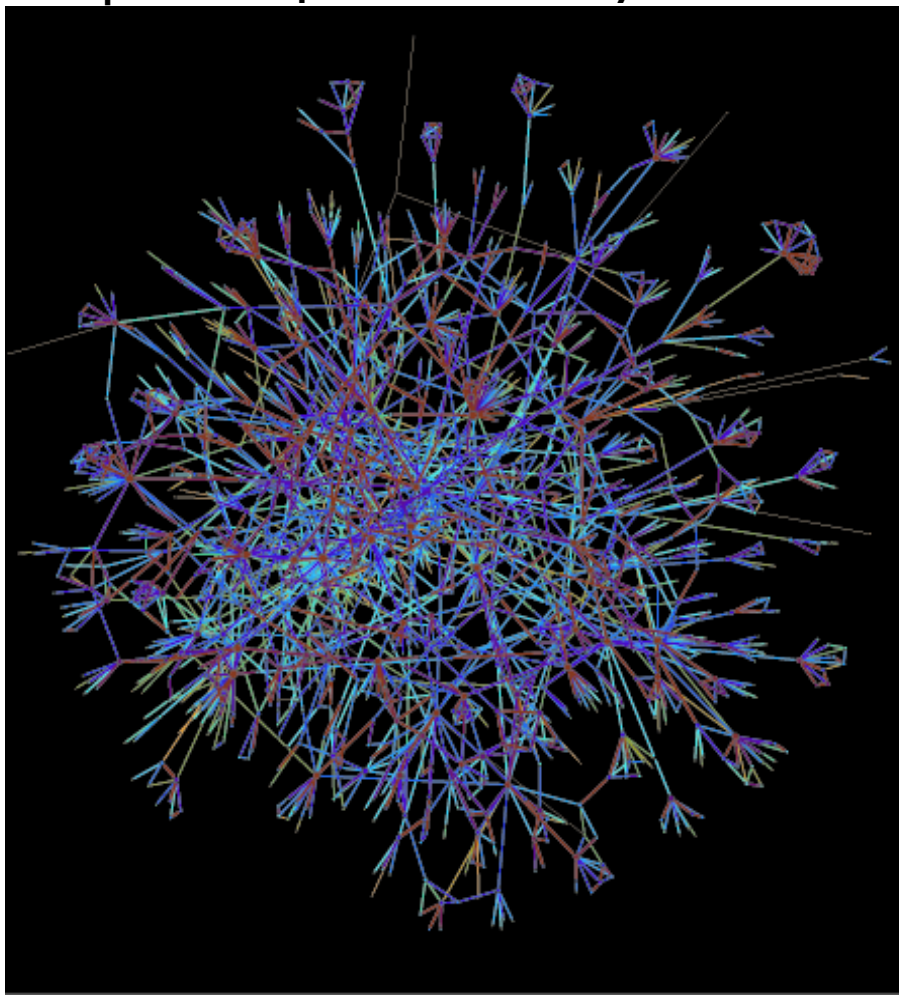
- Mobile phone operator data over a period of 2 years
- Millions of private subscriptions
- Voice calls / sms ' within the operator
- Lately also demographic data
- Require reciprocity of calls for a link
- Quantify link weights



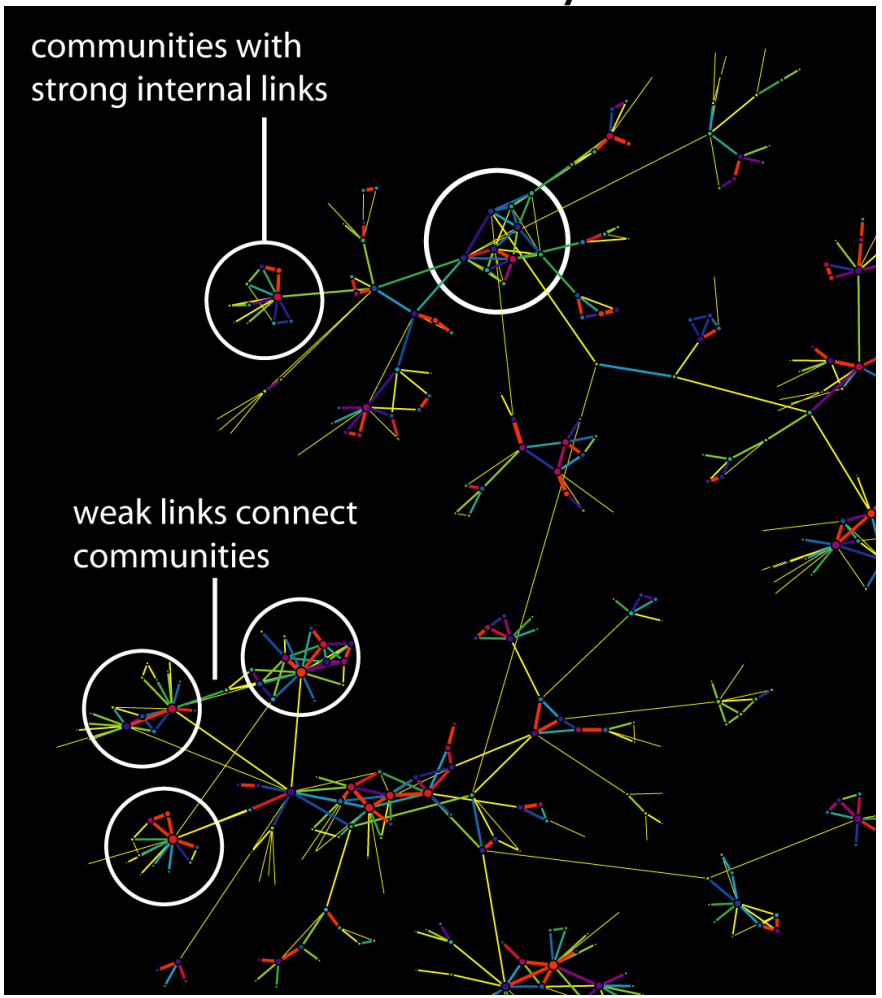
J.P. Onnela, J. Saramäki, J. Hyvönen, G. Szabo, D. Lazer, J. Kertesz, A.-L. Barabasi, K. Kaski, PNAS, 104, 7332, 2007

Constructing a network from data
Proxy for the underlying Social Network

Sample from 4.6 10^6 nodes & 7.0 10^6 links



Formation of community structure



J.P. Onnela, J. Saramäki, J. Hyvönen, G. Szabo, D. Lazer, J. Kertesz, A.-L. Barabasi, K. Kaski, PNAS, 104, 7332, 2007

Structure and Tie Strength:
Interplay Local and Global Scales

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

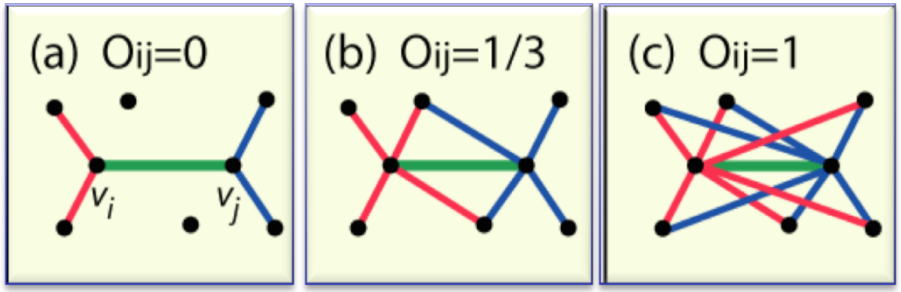
Relative neighbourhood overlap

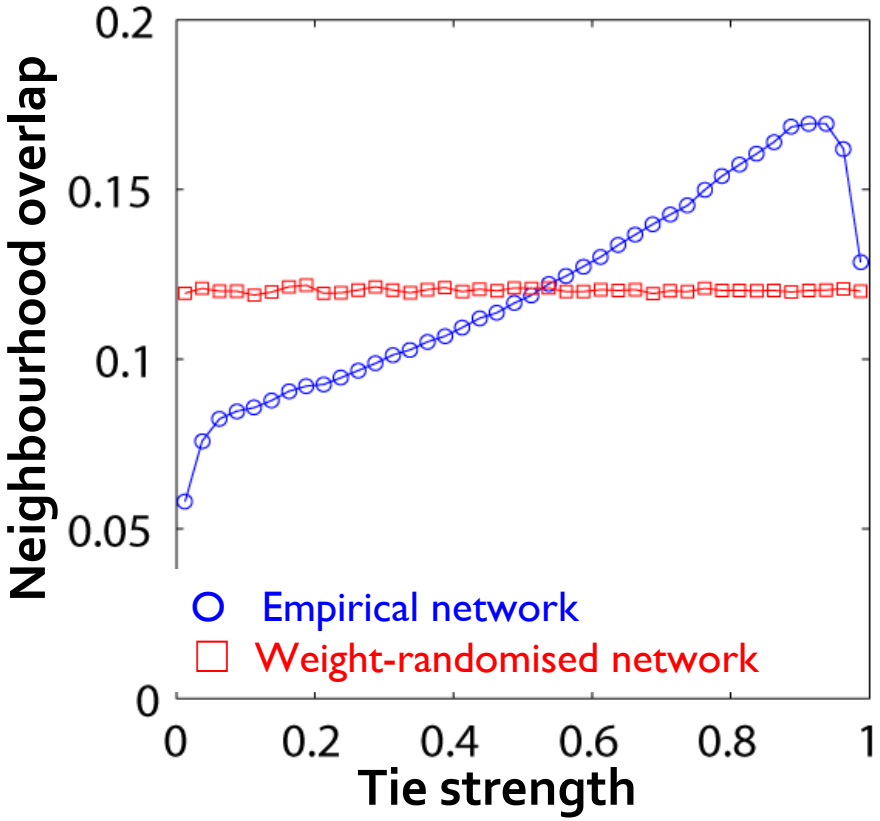
(k_i =*degree* & n_{ij} = *# of triangles with edge ij*)

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

Cumulative weight

$$P_{\text{cum}}(w') = \sum_{w \le w'} P(w)$$



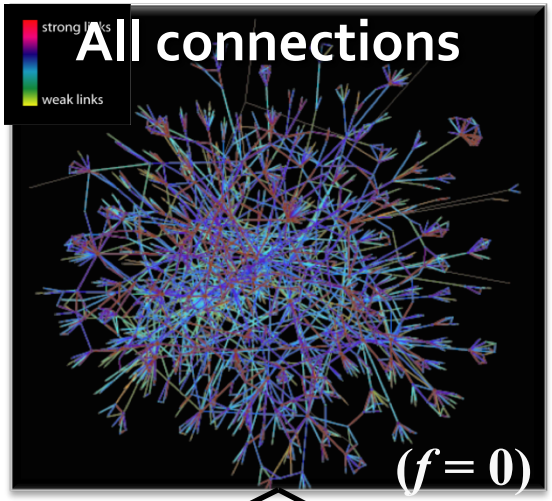


Verifies Granovetter's hypothesis (up to 95%)

M. Granovetter, The Strength of Weak Ties, T*he American Journal of Sociology* 78, 1360, 1973.

Global Structure
Percolation / Threshold Analysis

Proc. Natl. Acad. Sci. 104, 7332, 2007

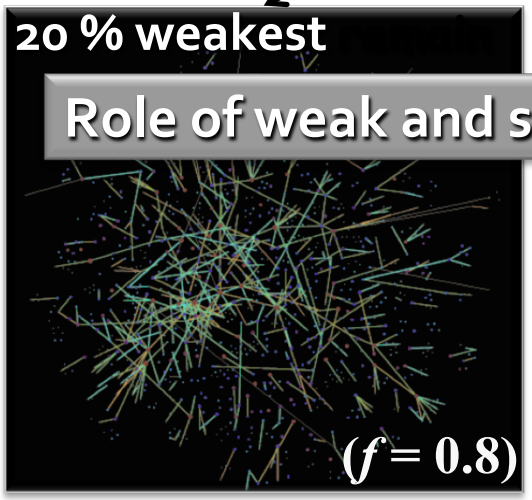


Order parameter R_{LCC}
- Fraction of nodes in LCC

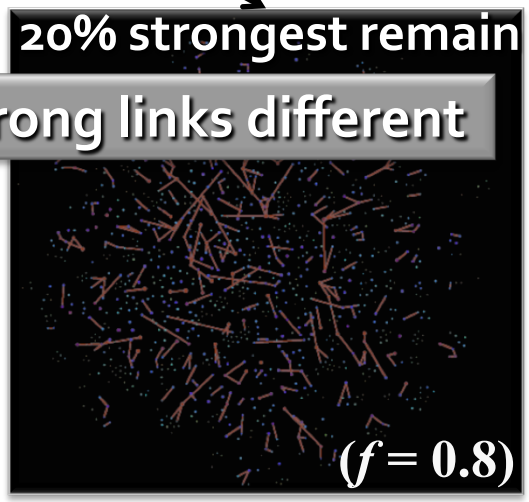
Susceptibility $S = \sum ns^2 / N$
- Average cluster size (not LCC)

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

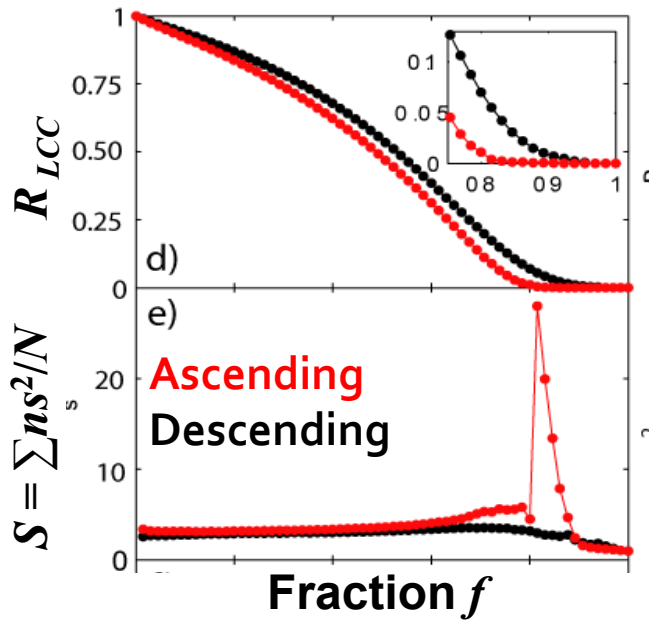
Descending threshold:



Ascending threshold:



Role of weak and strong links different



Dynamics of Spreading

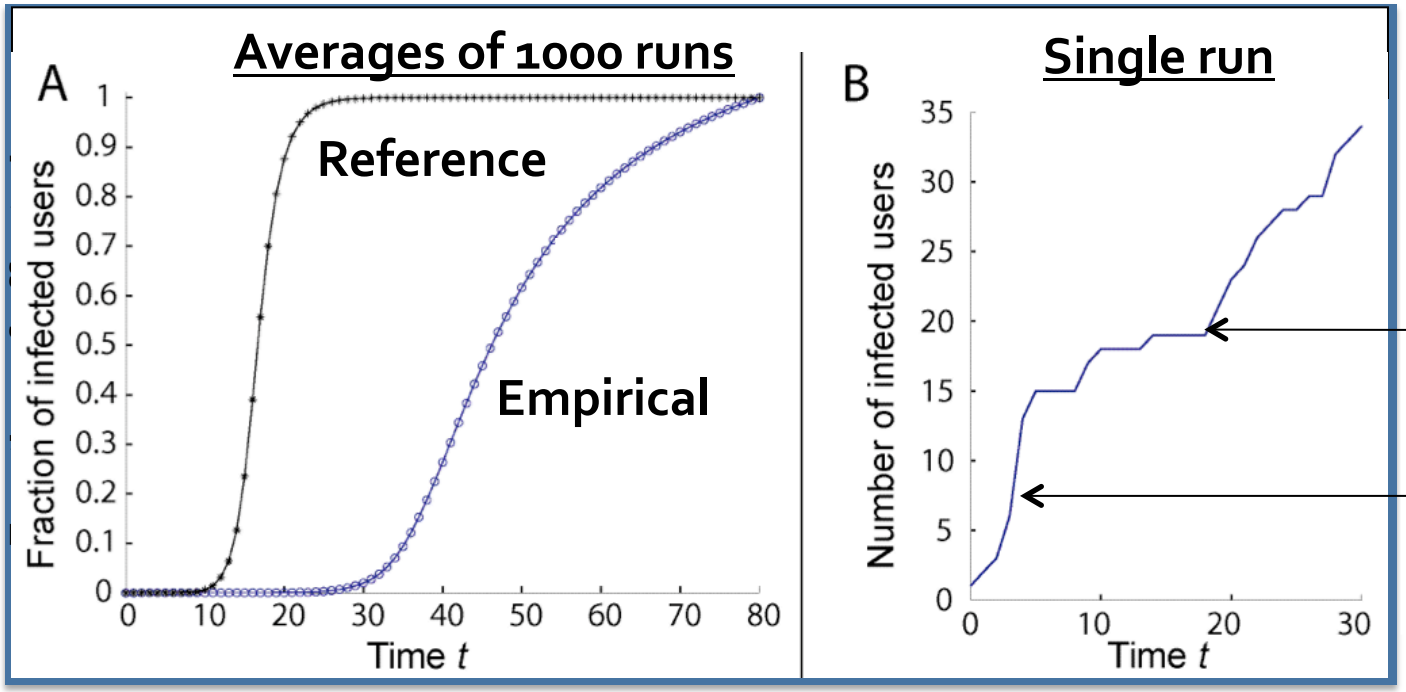
Spreading simulation (SI): infect one node with new information

(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

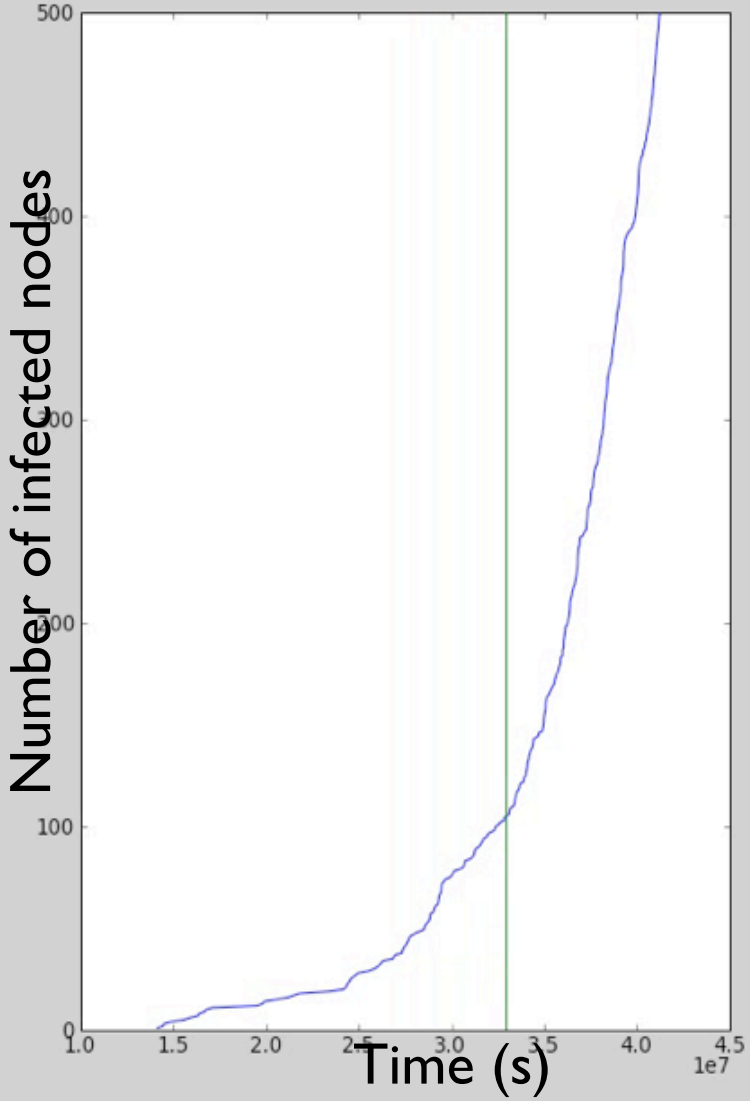
Information gets trapped within network communities

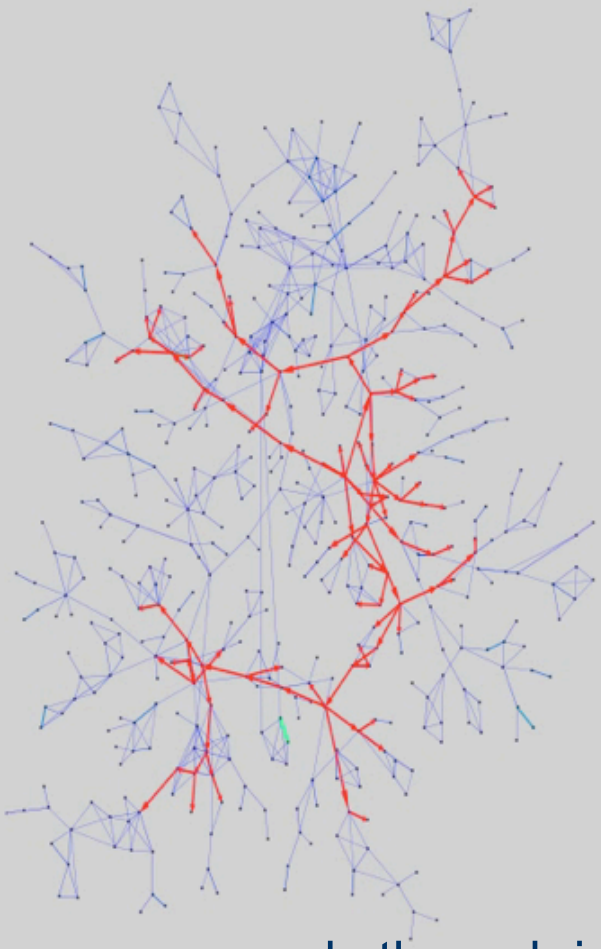


Even ICT-aided human society not most effective in information spreading

Simulation:
'Rumor' Spreading in a Real Network

At network diffusion level time scale:





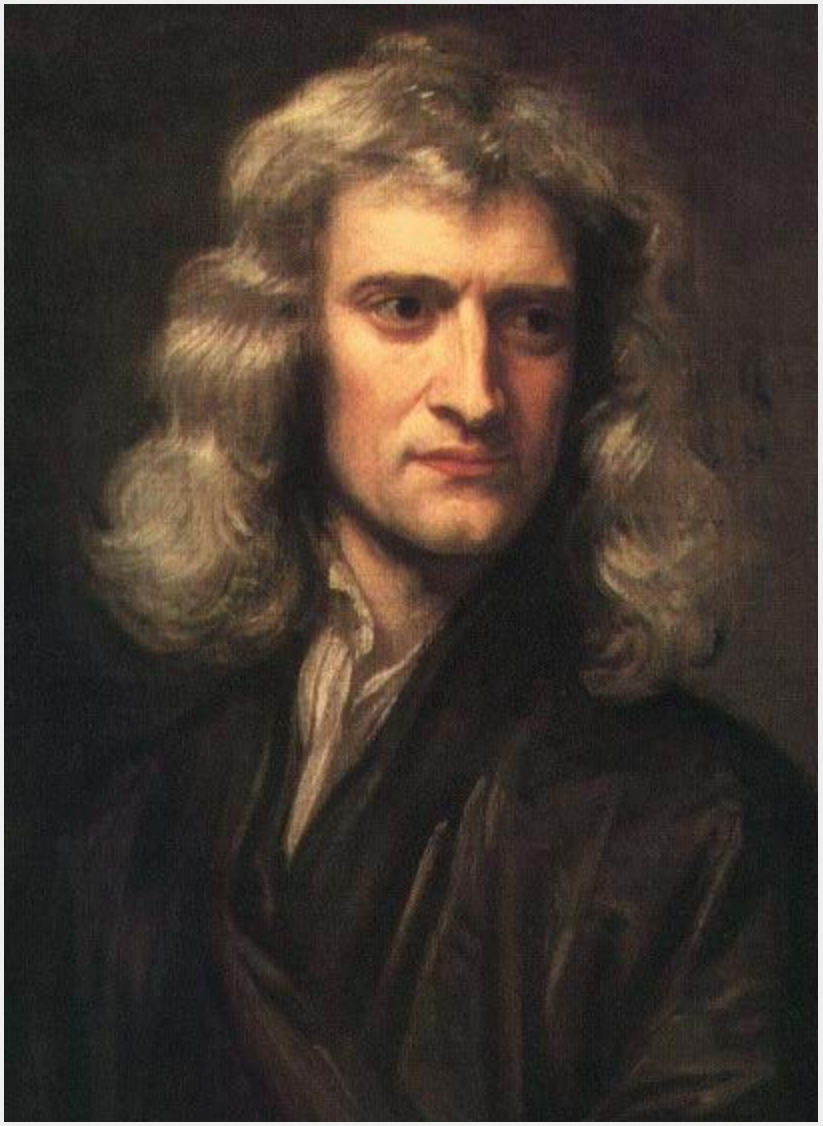
The rumor spreads through in 4 days

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

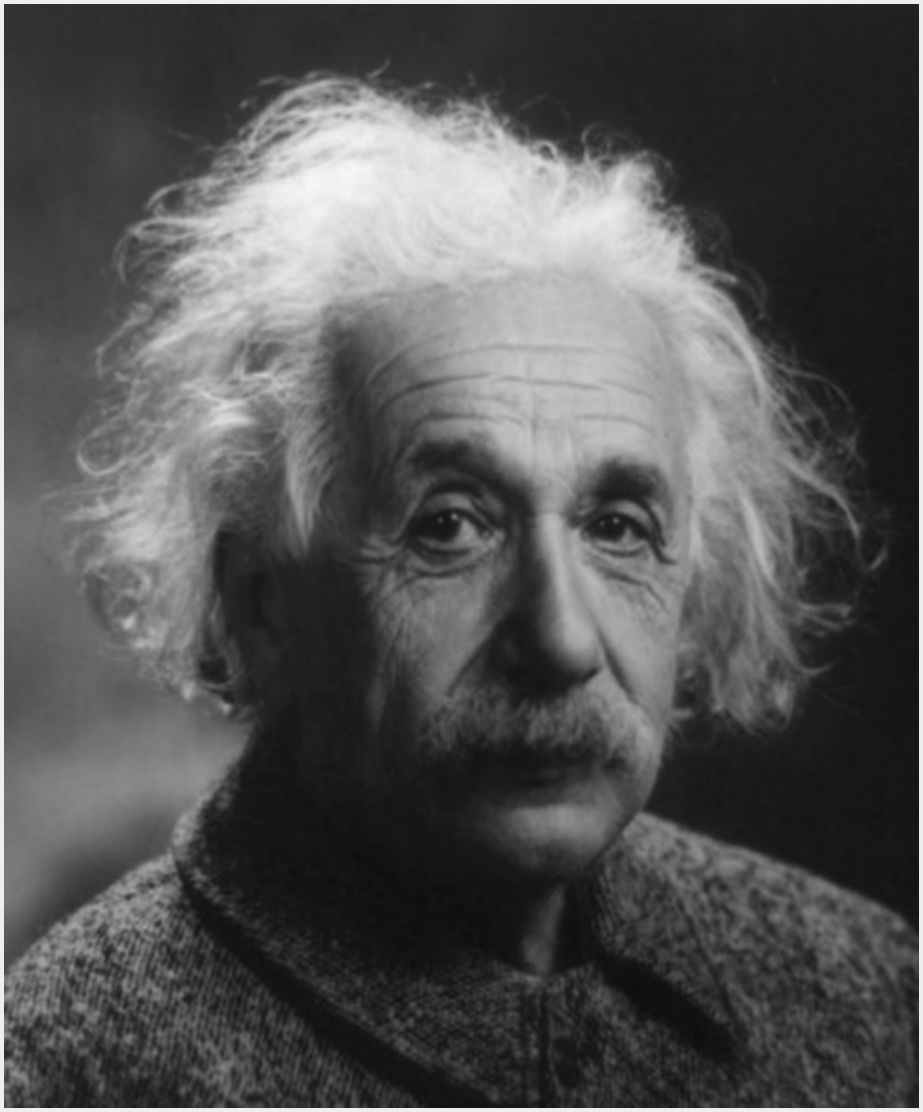
Models are like maps - useful when contain details of interest and ignore others (Boyd & Richerson: Mathematical Models of Social …)

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

Weighted Social Network Model
Agent based approach to Network formation

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

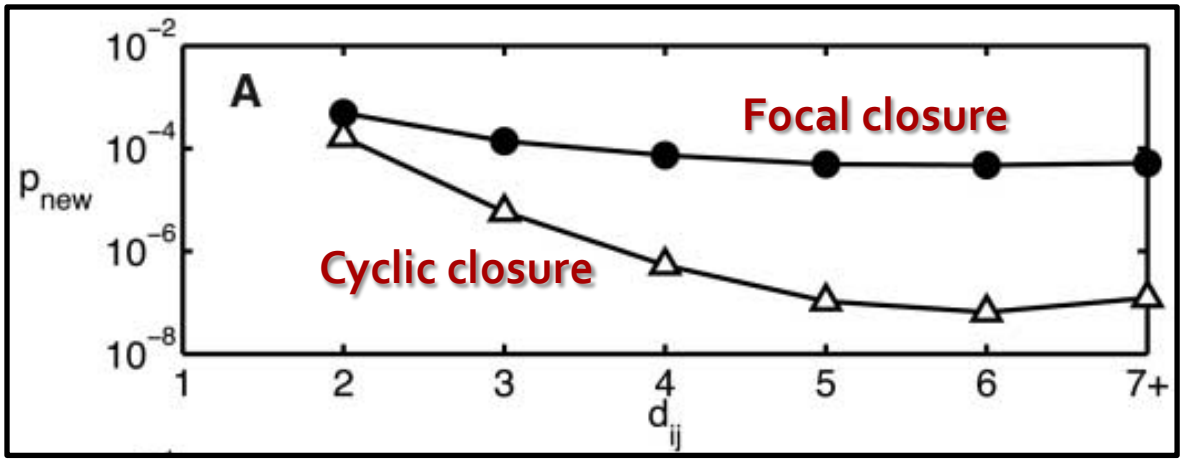
- **Fixed size network of N nodes (individuals)**
  - Internal structural changes faster than changes in the size of the network
- **Network subject to following dynamics:**
  - Local weighted search for new acquaintances and reinforcement of popular links
  - Global search by creation of random links
  - Random removal of nodes

*Unweighted search proposed by Marsili et al, PNAS (2004) & Davidsen et al. PRL (2002)

Social Network Formation
According to Network Sociology

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.)



Generative network model needs to account separately for triadic closure, focal closure, and compounding effect of both biases together.

*Kossinets & Watts, "Empirical Analysis of an Evolving Social Network", Science (2006)

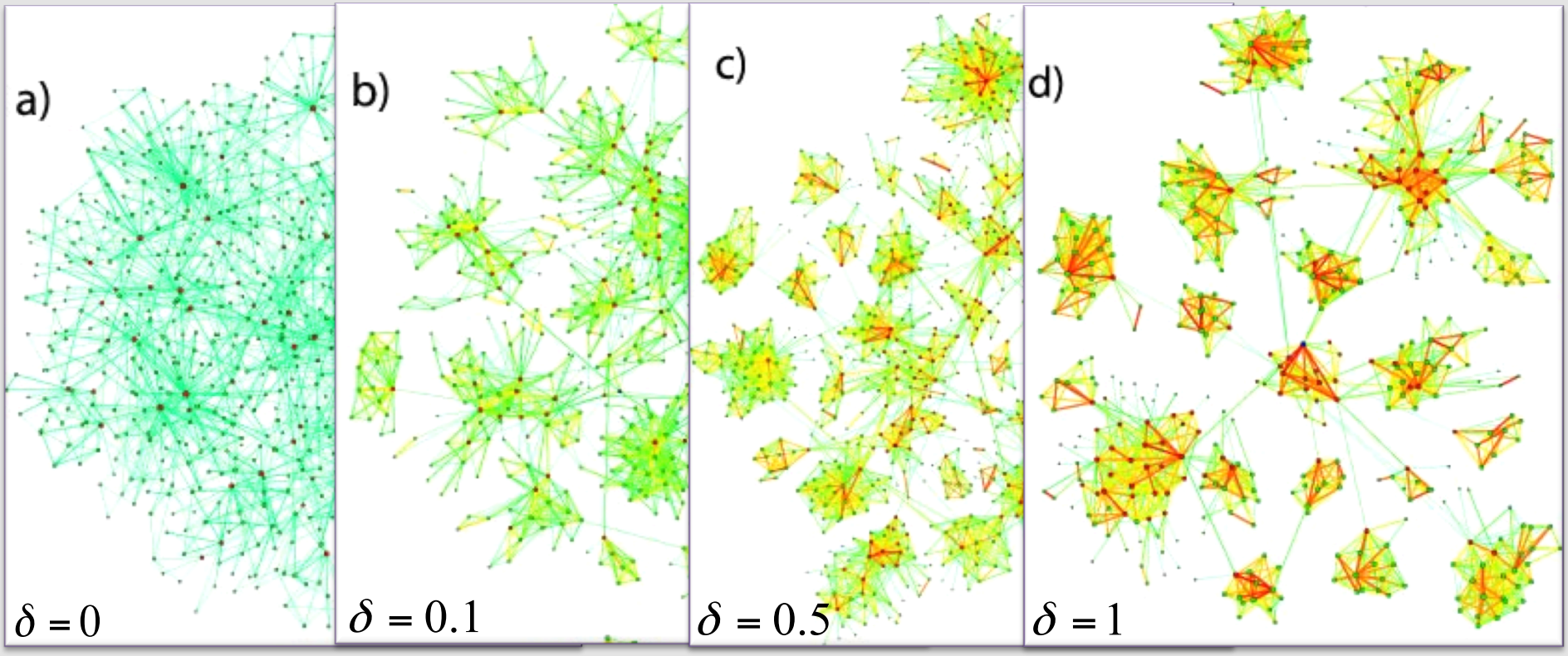
Mechanisms of Frendship formation

Network sociology	Network model
• Cyclic closure – Exponential decay for growing geodesic distance	• Local attachment (LA) – Special case of cyclic closure: Triadic closure
• Focal closure – Distance independent	• Global attachment (GA)
• "Sample window fixed"	• Node deletion (ND)
• -	• Link reinforcement

* M. Kossinets et al., "Empirical Analysis of an Evolving Social Network", Science 311, 88 (2006)

Formation of a Social Network

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



No communities

Communities start nucleating

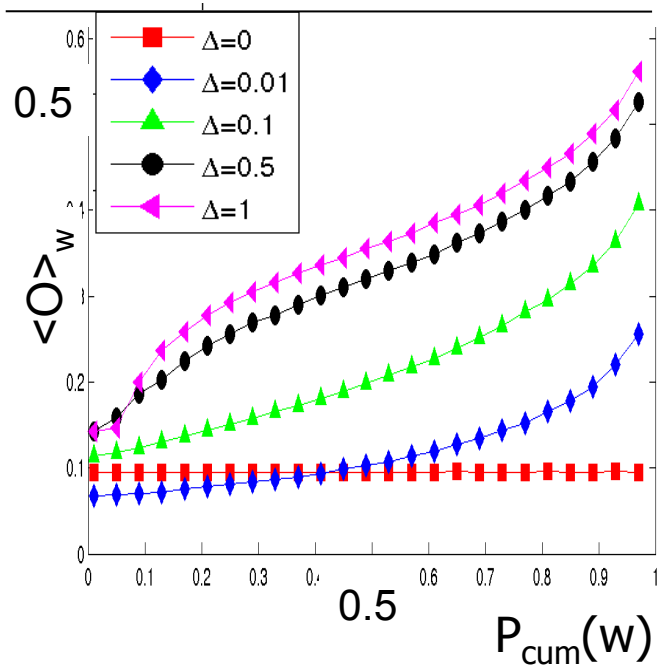
Communities forming

Communities with dense & strong internal and sparse & weak external connections (cf. phone network)

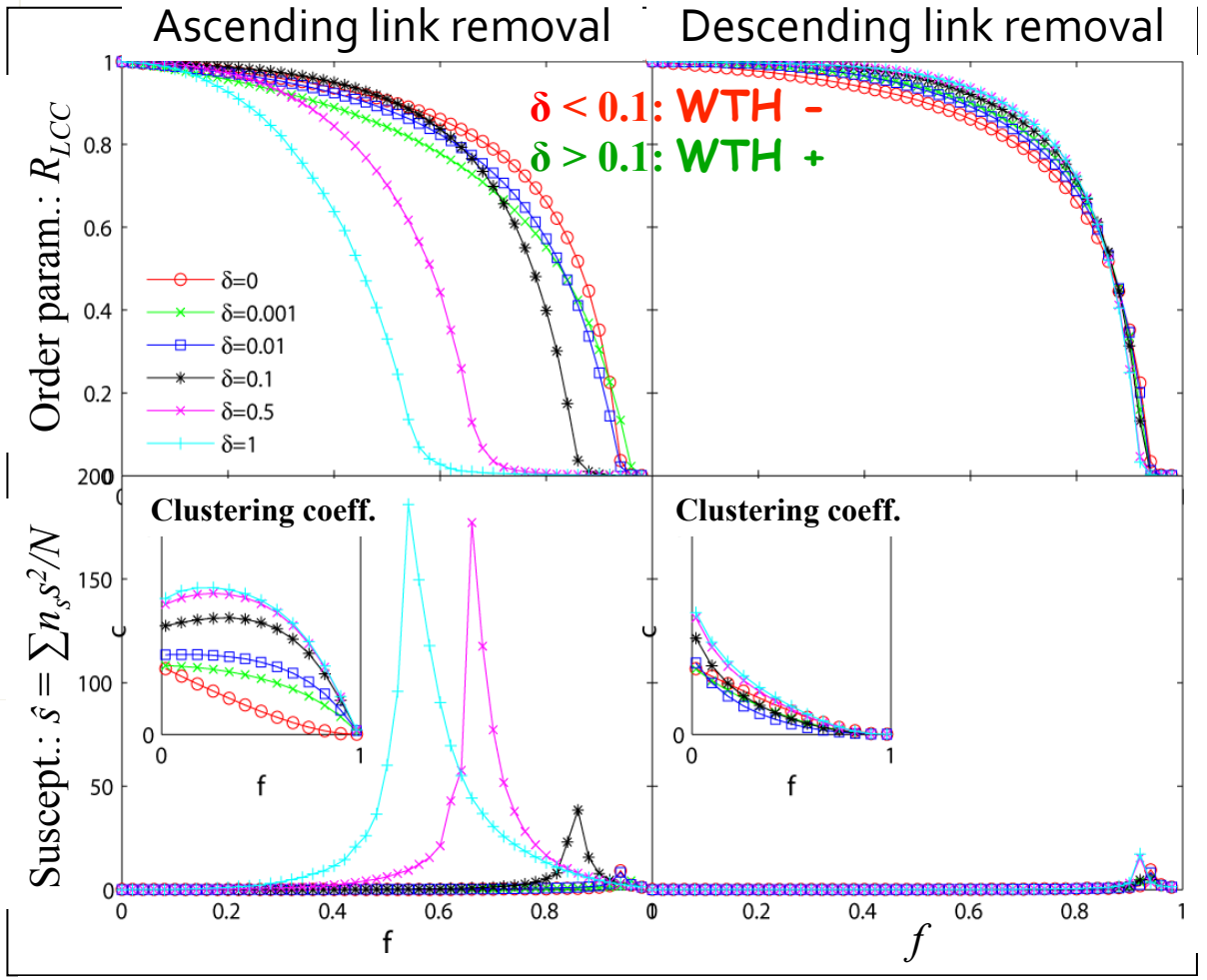
J. Kumpula, J.-P. Onnela, J. Saramaki, K.Kaski, J. Kertesz, PRL 99 (2007) 228701.

Similar Local & Global Structure

Overlap vs. tie strength



Weak Ties Hypothesis behavior reproduced!



Reproduces different roles of weak vs. strong ties -> "phase chance" for ascending link removal

Summary: Model vs. Real Network

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

J. Kumpula, J.-P. Onnela, J. Saramaki, K.Kaski, J. Kertesz, PRL 99 (2007) 228701.

Egocentric (*in Vivo*) Networks:
Sex Differences in Intimate Relationships

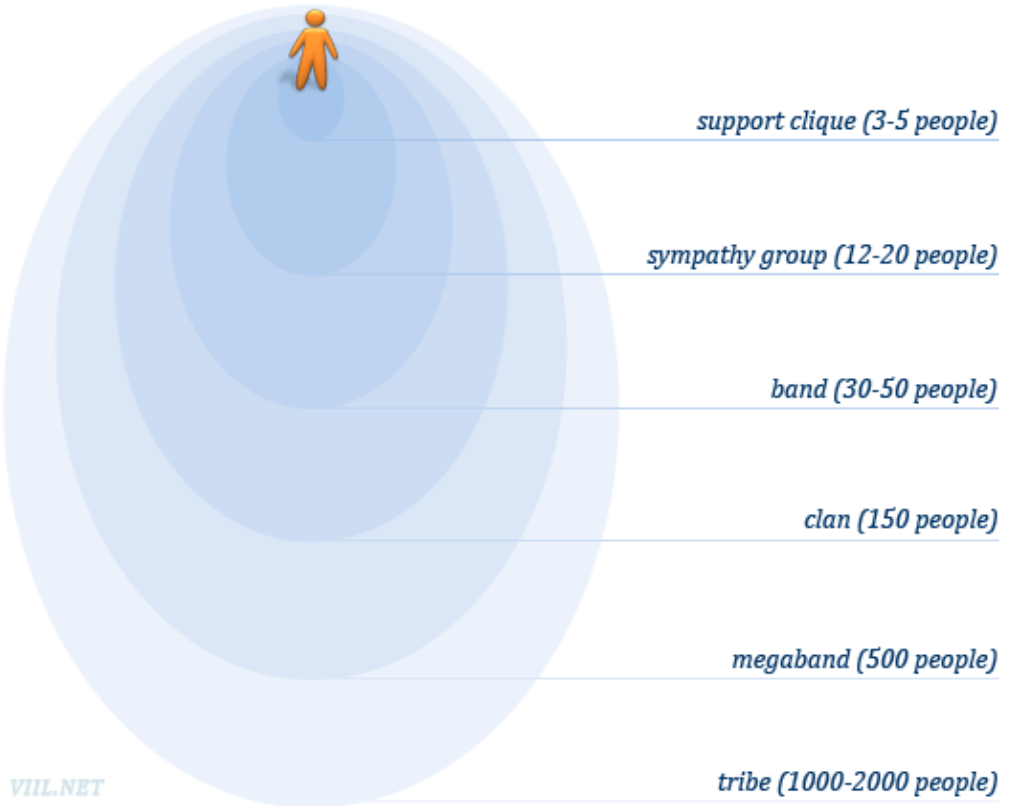
Age and gender correlations in friendships

V. Palchykov, K. Kaski, J. Kertész, A.-L. Barabási and R.I.M. Dunbar,
Nature Scientific Reports ***2**, 370 (2012)*

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.

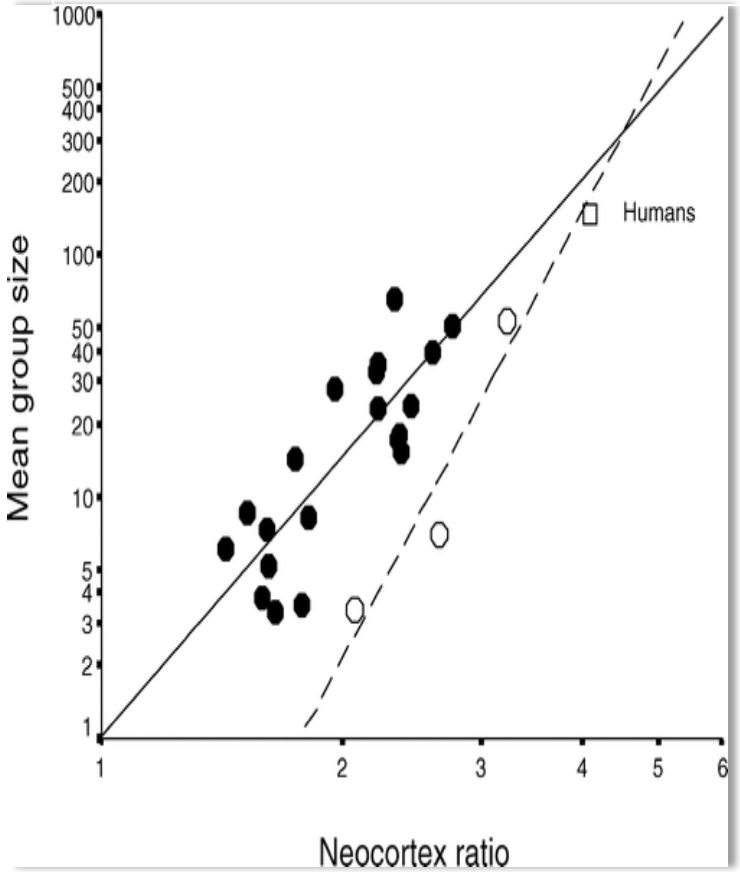


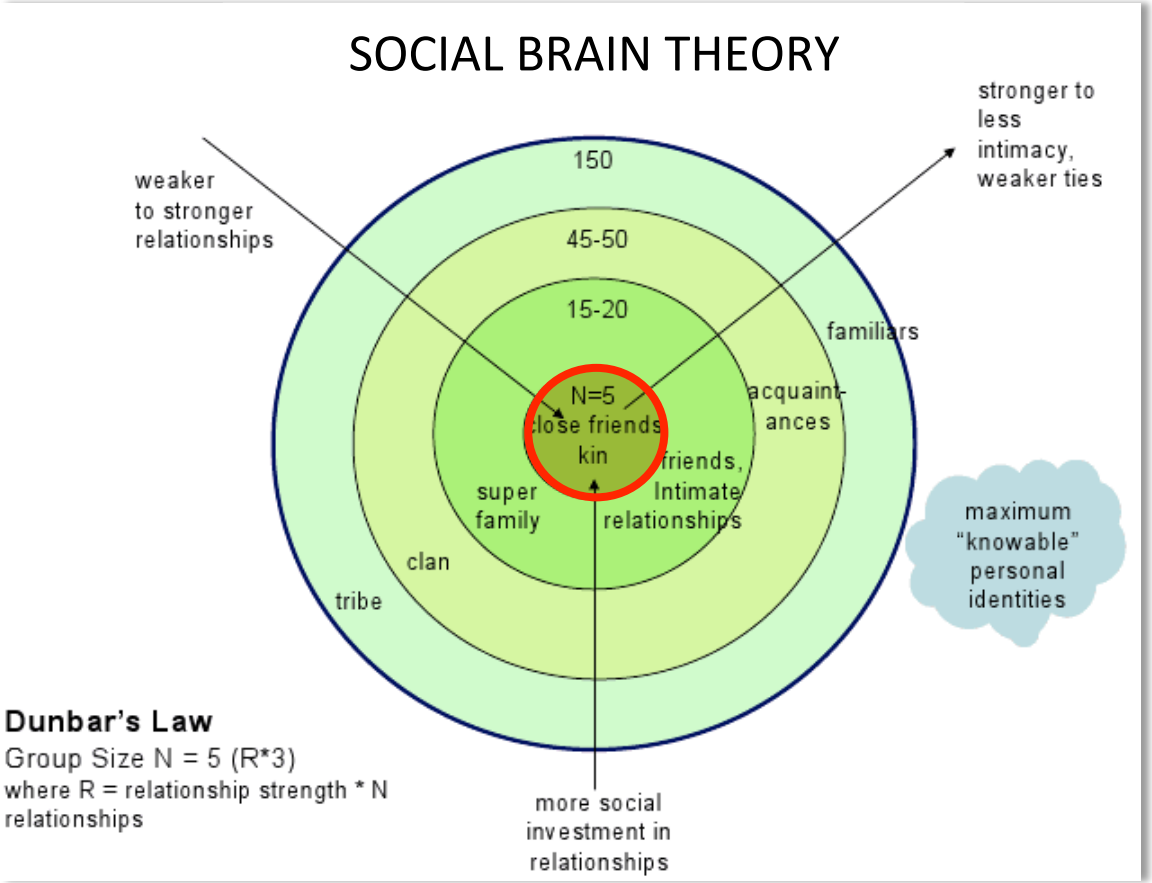
Are there gender and age specific properties in making human relations?

Social Brain Hypothesis

Humans can handle roughly three times as many social contacts as apes

Group size vs. Neocortex ratio

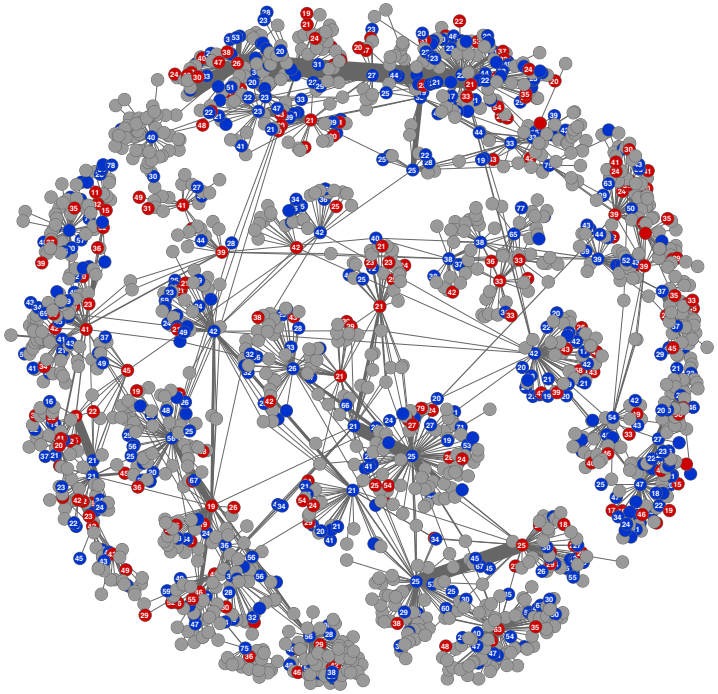




We study the innermost circle

Dataset with Demographic Data

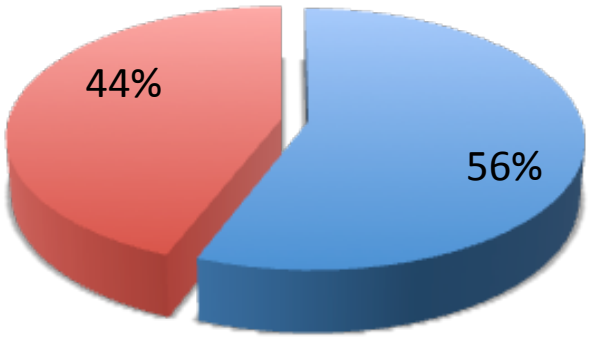
- The same as earlier, i.e. one mobile phone operator in a European country but more recent and detailed dataset:
- ~ 8 million service subscribers
- ~ 1.95 billion calls
- ~ 489 million text messages
- Covers 7 months period (2009)
- In addition demographic data:
  - Gender of subscriber
  - Age of subscriber
  - Postcode of subscriber



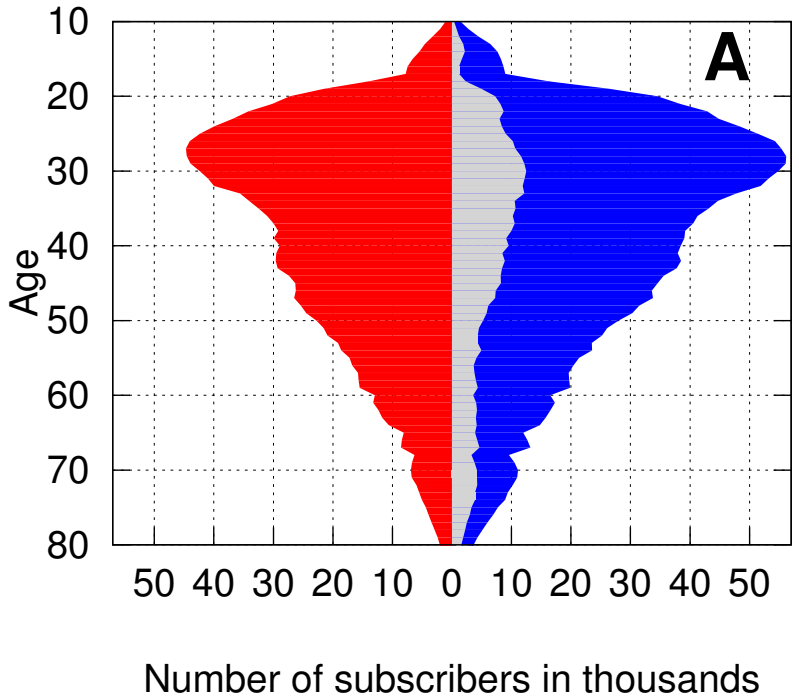
Communication events with their #, duration & timings through a single+ channel (calls+SMS) reflect a high degree of ego-alter sociality -> Phone network serves as a proxy for social network

Demographics & Gender Balance

Carrying out initial data filtering we arrive at 3.2 million subscribers:

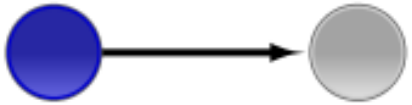


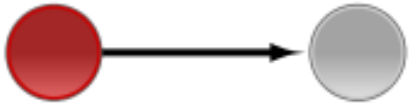
- 1.8 million males;
- 1.4 million females.



$$\text{Gender variable } g = \begin{cases} +1 \text{ for males} & \text{Balance: } \langle g \rangle = 0 \\ -1 \text{ for females} & \text{Here: } \langle g \rangle = 0.13 \end{cases}$$

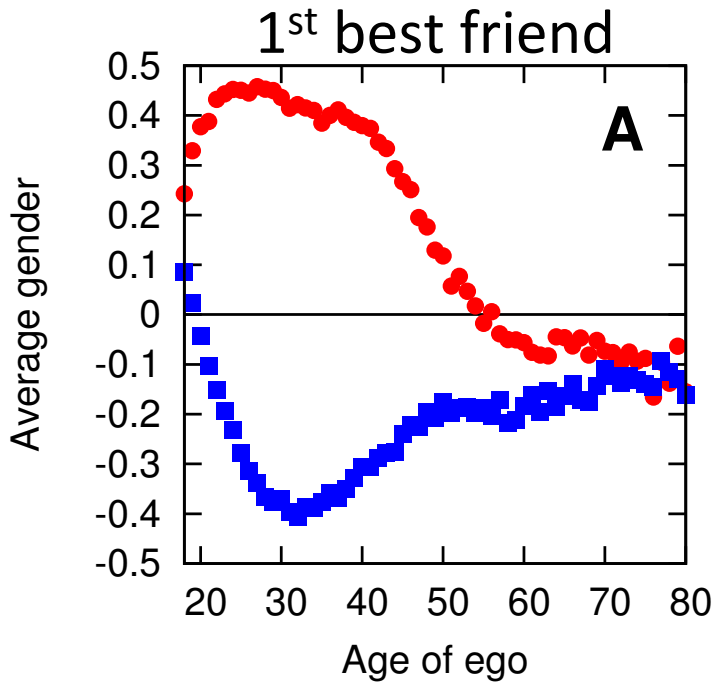
Gender Correlations:
1st & 2nd best friend

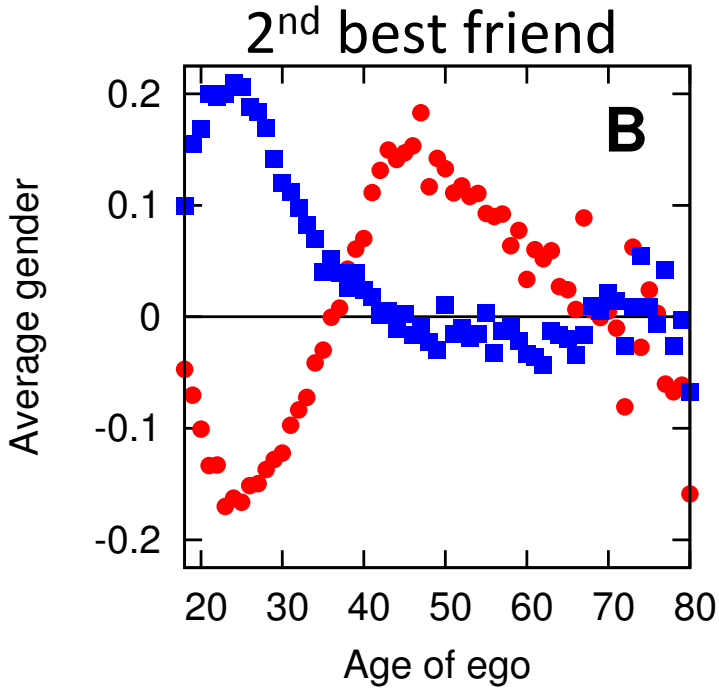
$g_i = +1$  f_i

$g_i = -1$  f_i

Average gender of the best friend:

$$\langle f \rangle = \frac{1}{N} \sum_i f_i$$





- 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

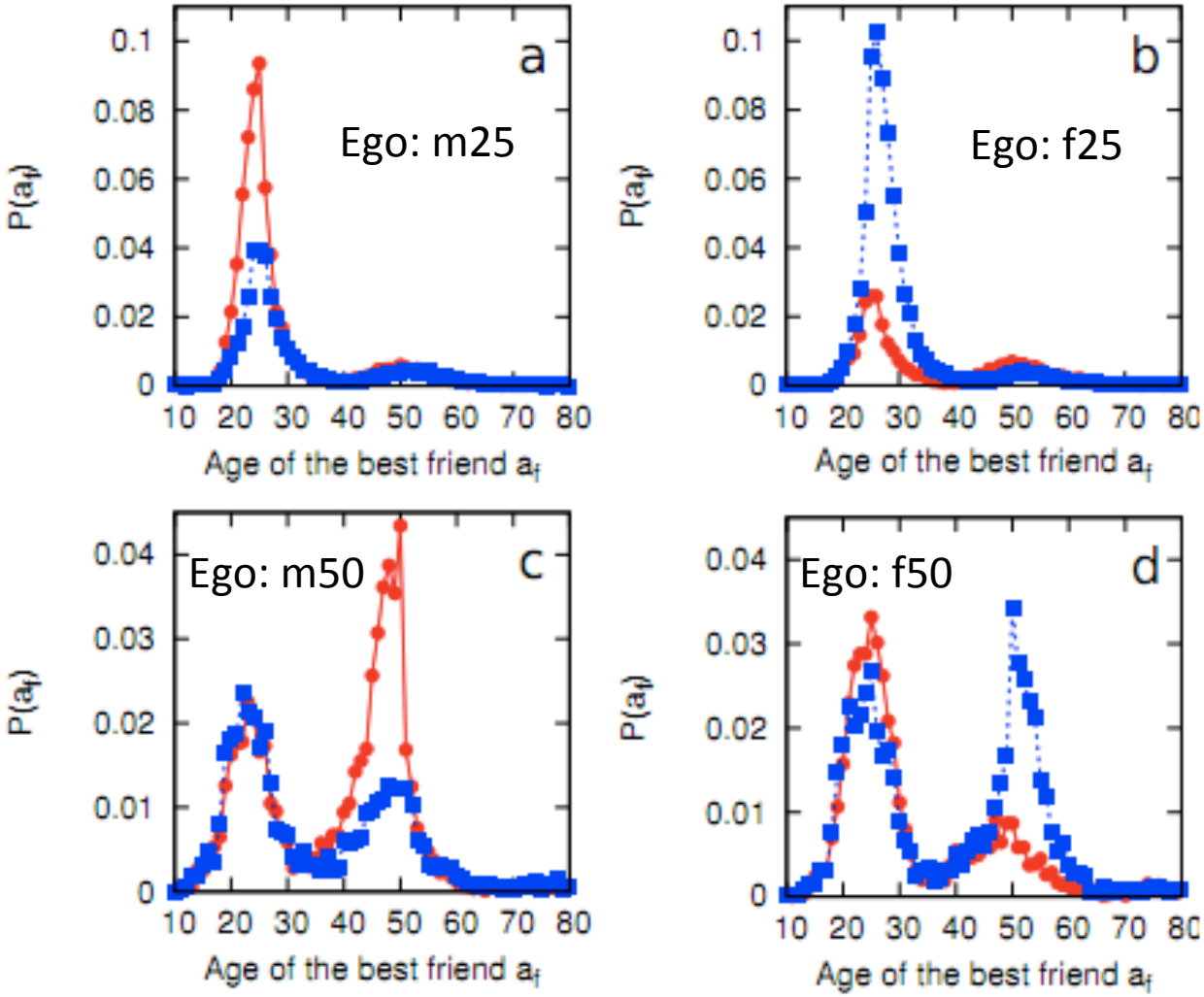
Age distributions of the best friend

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.



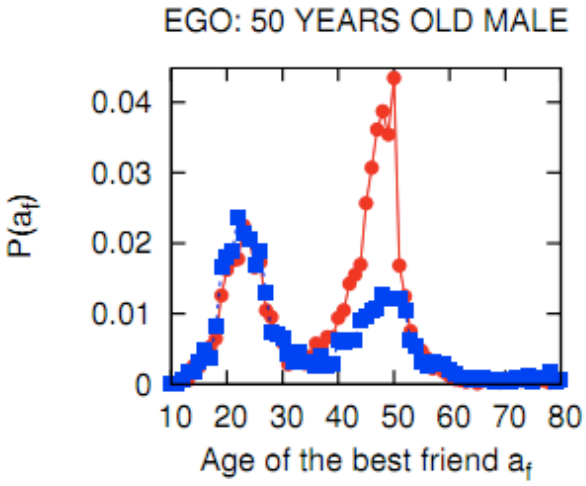
Age distributions of the best friend

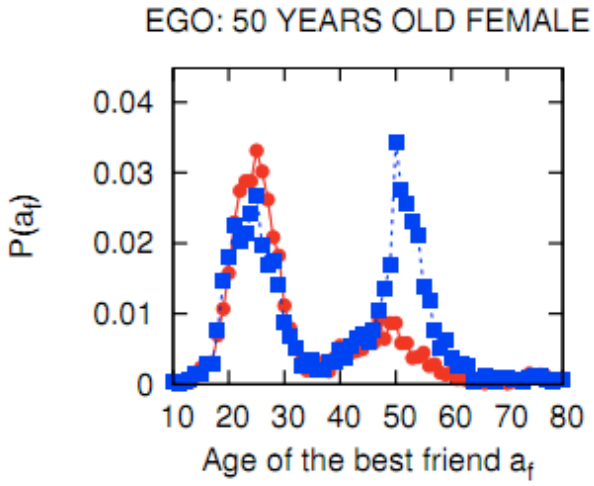
Distribution of best friends by age

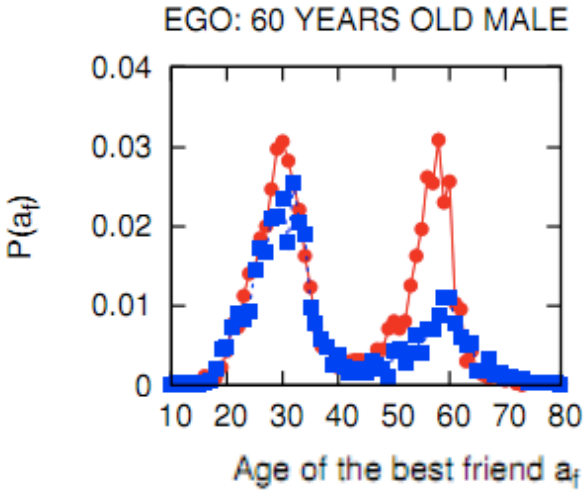
As people age their attention shifts from the spouse to the children

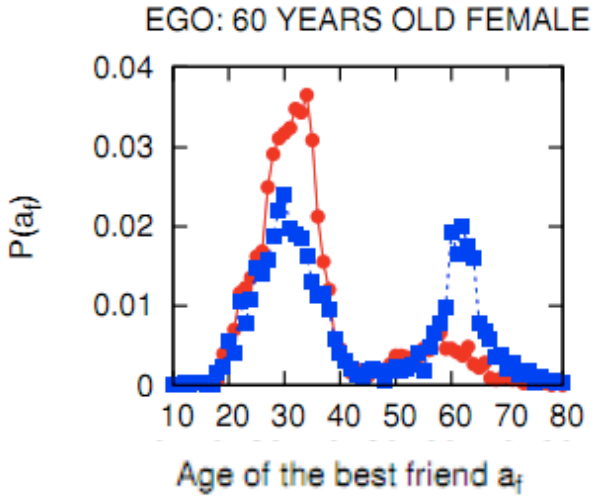
Women are more active in maintaining family relationships

The mother-daughter link is particularly strong.

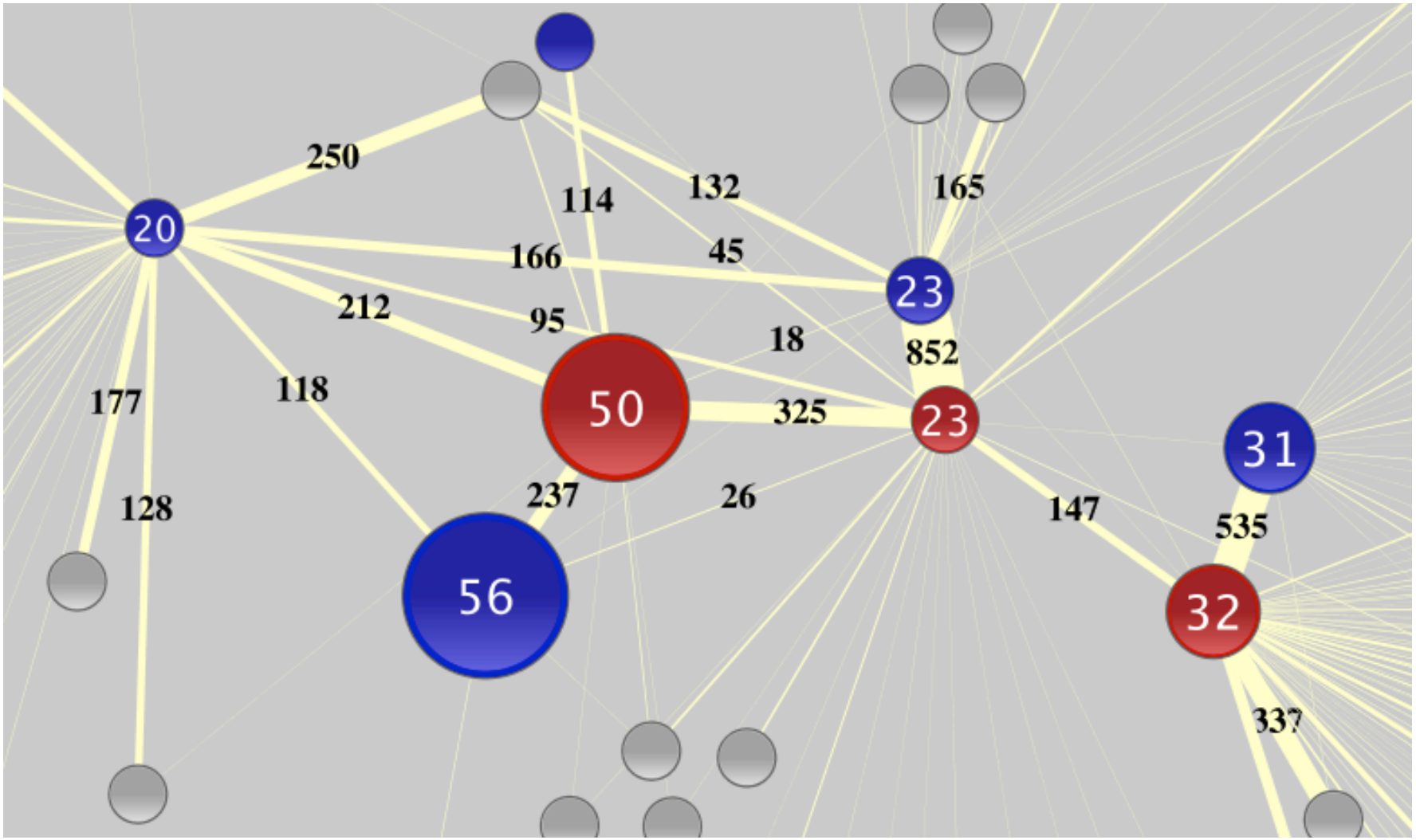








Sample of Ego – Alter Network



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.

V. Palchykov, K. Kaski, J. Kertész, A.-L. Barabási, R.I.M. Dunbar. Sci. Rep. ***2****, 370 (2012).*