Emberi viselkedés info-kommunikációs adatokból

Kertész János
CEU és BME

ELTE: ATOMOKTÓL A CSILLAGOKIG

COMPLEX SYSTEMS

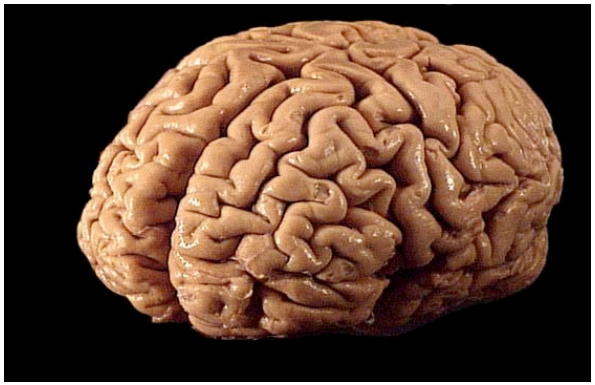
Complex systems: Many interacting components, feedback, nonlinearity, cooperativity (self)adaptation, emergent phenomena

THE WHOLE IS MORE THAN THE MERE SUM OF THE PARTS

(Complex ≠ complicated)

Watch:
complicated

linear chain of logic

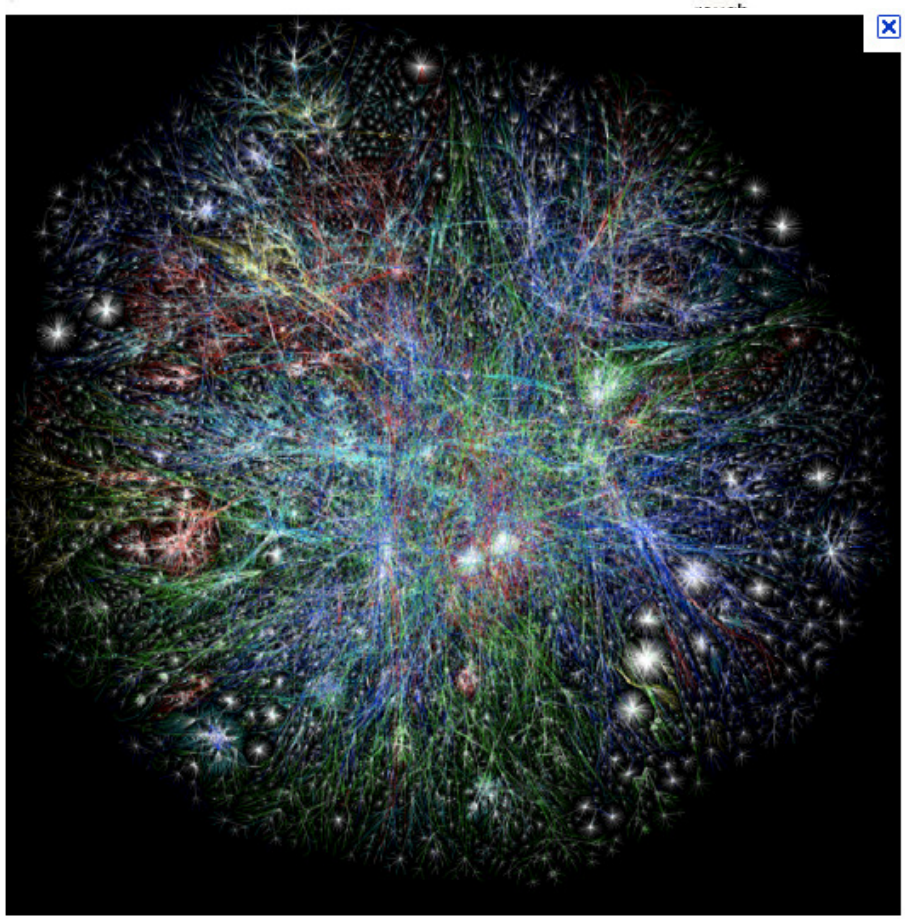


Brain:
complex

cells→ … thoughts, emotions

COMPLEX SYSTEMS

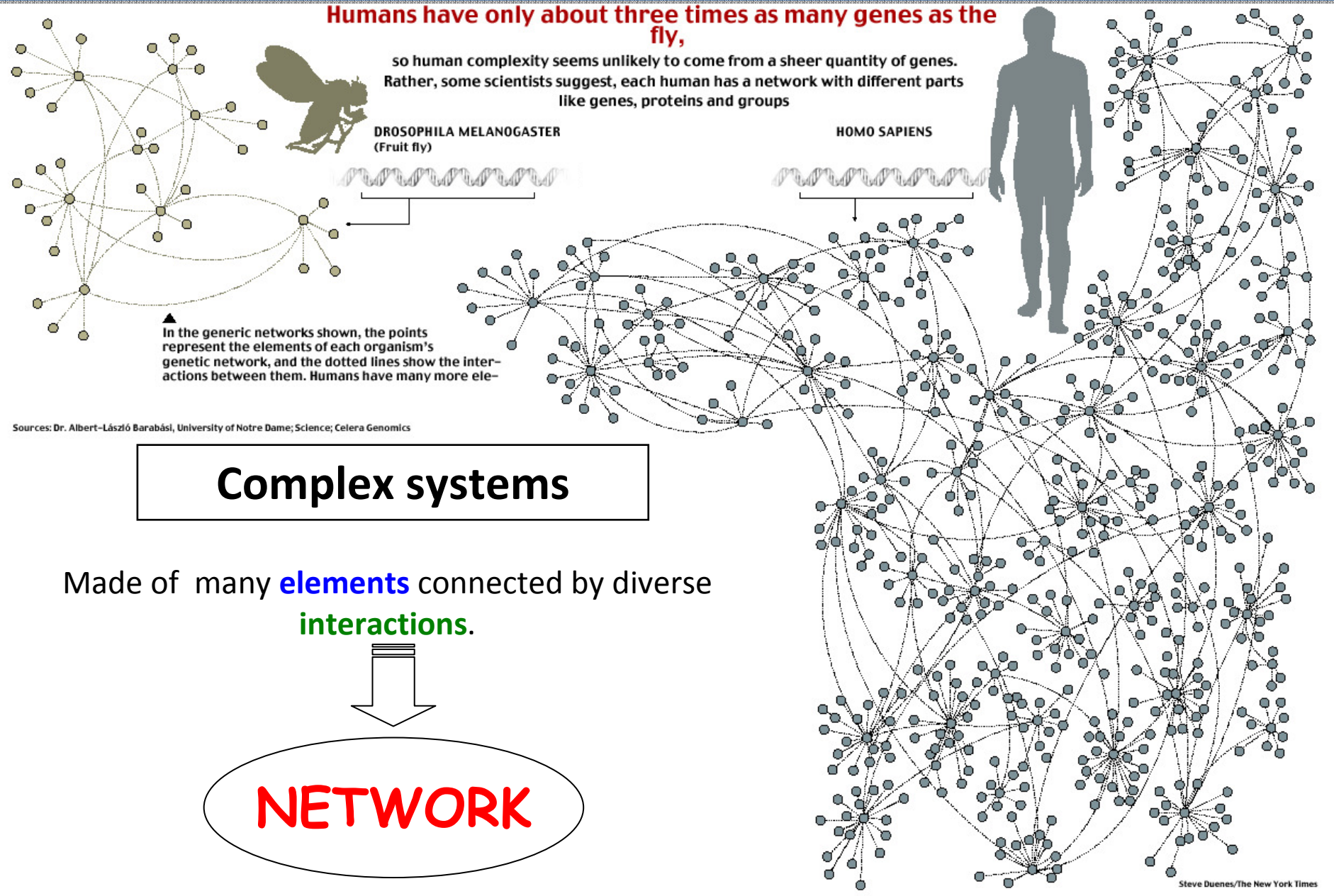
Examples:



The Internet

The economy

COMPLEX SYSTEMS



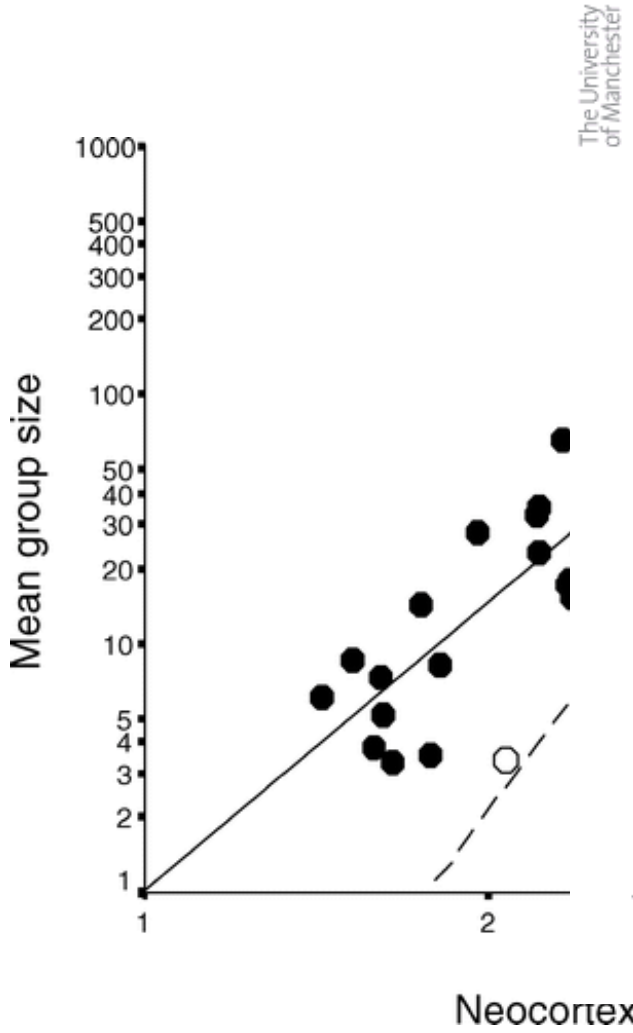
Complex systems

Made of many **elements** connected by diverse **interactions**.

NETWORK

COMPLEX SYSTEMS

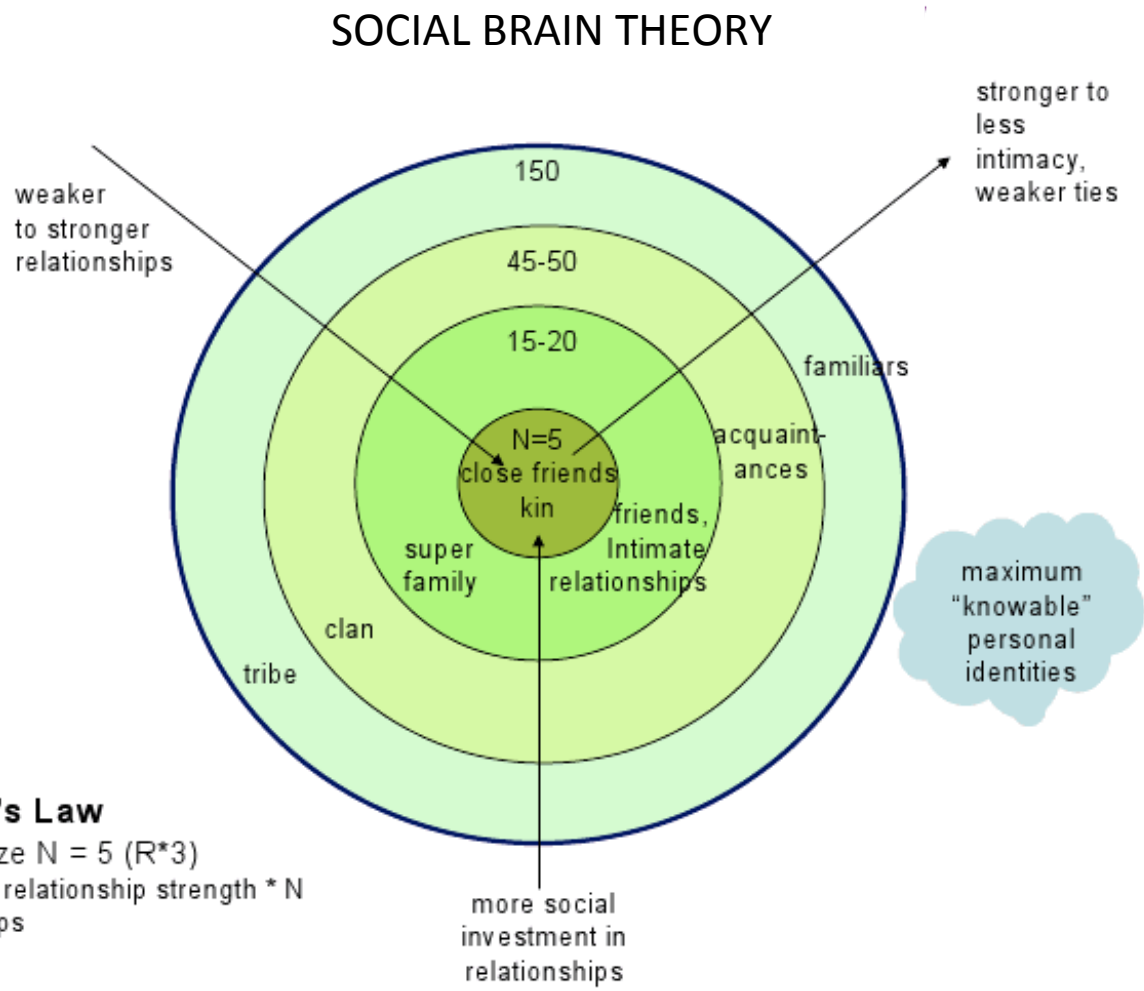
Society: We can handle roughly three times as many social contacts as apes...



Dunbar's Law

Group Size N = 5 (R*3)

where R = relationship strength * N relationships



R. Dunbar: Annual Review of Anthropology, 32, 163 (2003)

COMPLEX SYSTEMS

What matters?
The way how elements are connected.

Road map for studying complex systems:

- Identify the skeleton of the system: the 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 network to functions
- Describe dynamic processes and the influencing factors including network structure

Data needed

ICT: CHANGES OF METHODOLOGY

Within less than one generation deep changes in human behavior due to development in ICT:

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

Changes in the whole society: ”Facebook generation”
ICT in the hands of people has got history shaping factor

Egypt's Facebook Revolution: Wael Ghonim Thanks The Social Network

Catharine Smith | First Posted: 02/11/11 03:36 PM | Updated: 02/14/11 04:59 PM

: CHANGES OF METHODOLOGY

Google

Speciális keresés
Beállítások
Nyelvi eszközök

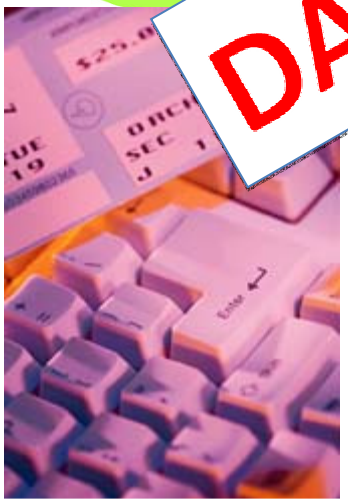
Magyarországról

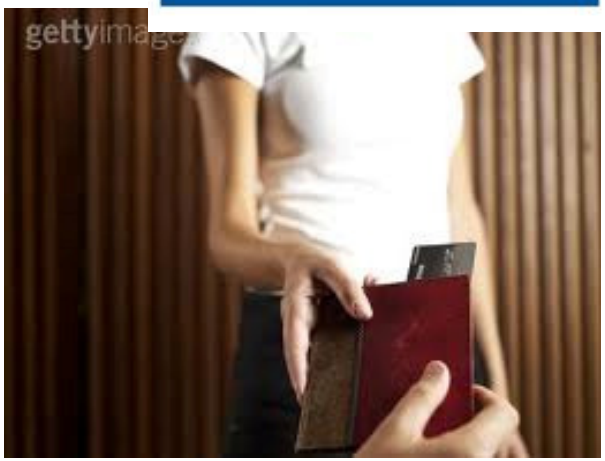
a Google-ról - Google.com in English

DATA DATA DATA DATA

twitter

facebook





ICT: CHANGES OF METHODOLOGY

A new way to study the complex system of the society:

Computational Social Science

Until now, social science has struggled to obtain tools that do more than scratch the surface of some of its questions. These range from identifying the driving forces behind violence, to the factors influencing how ideas, attitudes and prejudices spread through human populations. The available tools have largely remained in a time warp, consisting of analyses of national censuses, small-scale surveys, or lone researchers with a notebook observing interactions within small groups.

Being able to automatically and remotely obtain massive amounts of continuous data opens up unprecedented opportunities for social scientists to study organizations and entire communities or populations.

Nature |Vol 449|11 October 2007

ICT: CHANGES OF METHODOLOGY

Communications leave detailed information about who with whom, when and where…

- phone (mobile and fixed line)
- sms, mms
- Skype
- email

In a broader sense all kinds of activities can be used, which leave electronic records, including

- commercial activities (eBay, point collecting cards, credit cards, etc)
- open collaborative environments (Wikipedia, gnu, etc)
- E-communities (Facebook, MySpace, etc)
- E-games (Roleplaying, Where is George, etc)

ICT: CHANGES OF METHODOLOGY

„A BÚS FÉRFI PANASZAI”-BÓL

Beírtak engem mindenféle Könyvbe
és minden módon számon tartanak.
Porzó-szagú, sötét hivatalokban
énrólam is szól egy agg-szürke lap.
Ó, fogcsikorgatás. Ó, megalázás,
hogy rab vagyok és nem vagyok szabad.
nem az enyém már a kezem, a lábam,
és a fejem, az is csak egy adat.
Jobb volna élni messze sivatagban,
vagy lenn rohadni zsíros föld alatt,
mivel beírtak mindenféle Könyvbe
és minden módon számon tartanak.

Kosztolányi Dezső, 1924

LAMENTS OF A SORROWFUL MAN

They've entered me in books of every kind,
I'm registered and checked in every way.
I'm kept in musty, ink-stained offices,
in folders that are growing grizzly-grey.
Oh, gnashing of teeth, oh, humiliation,
that I am captive till my dying day,
that they dispose of me from top to toe,
that I am just a record, filed away.
I'd much prefer to live in the Sahara
or rot beneath a mound of heavy clay,
for I am kept in books of every kind,
and registered and checked in every way

Translation by Peter Zollman

BUT A GOLD MINE FOR RESEARCH!

STRUCTURE AND TIE STRENGTH IN MOBILE COMMUNICATION NETWORK

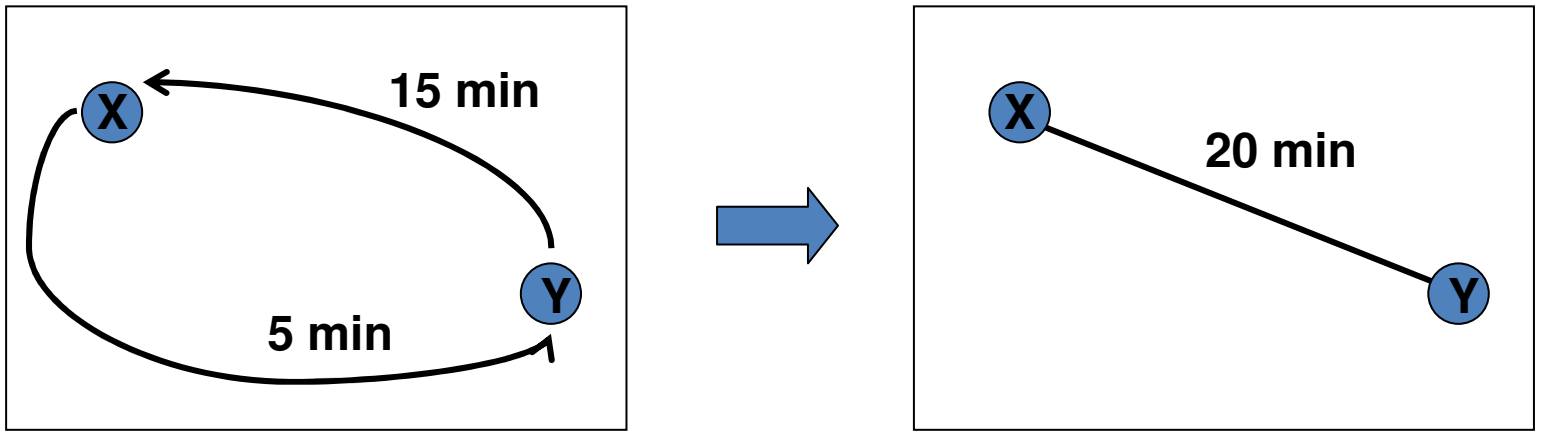
- Mobile phones play a unique role in today's communication
- Allmost 100% coverage in the adult population
- Communication network as a proxy for the social interactions

Over 7 million private mobile phone subscriptions

Focus: voice calls within the home operator

Data aggregated from a period of 18 weeks, anonymized users

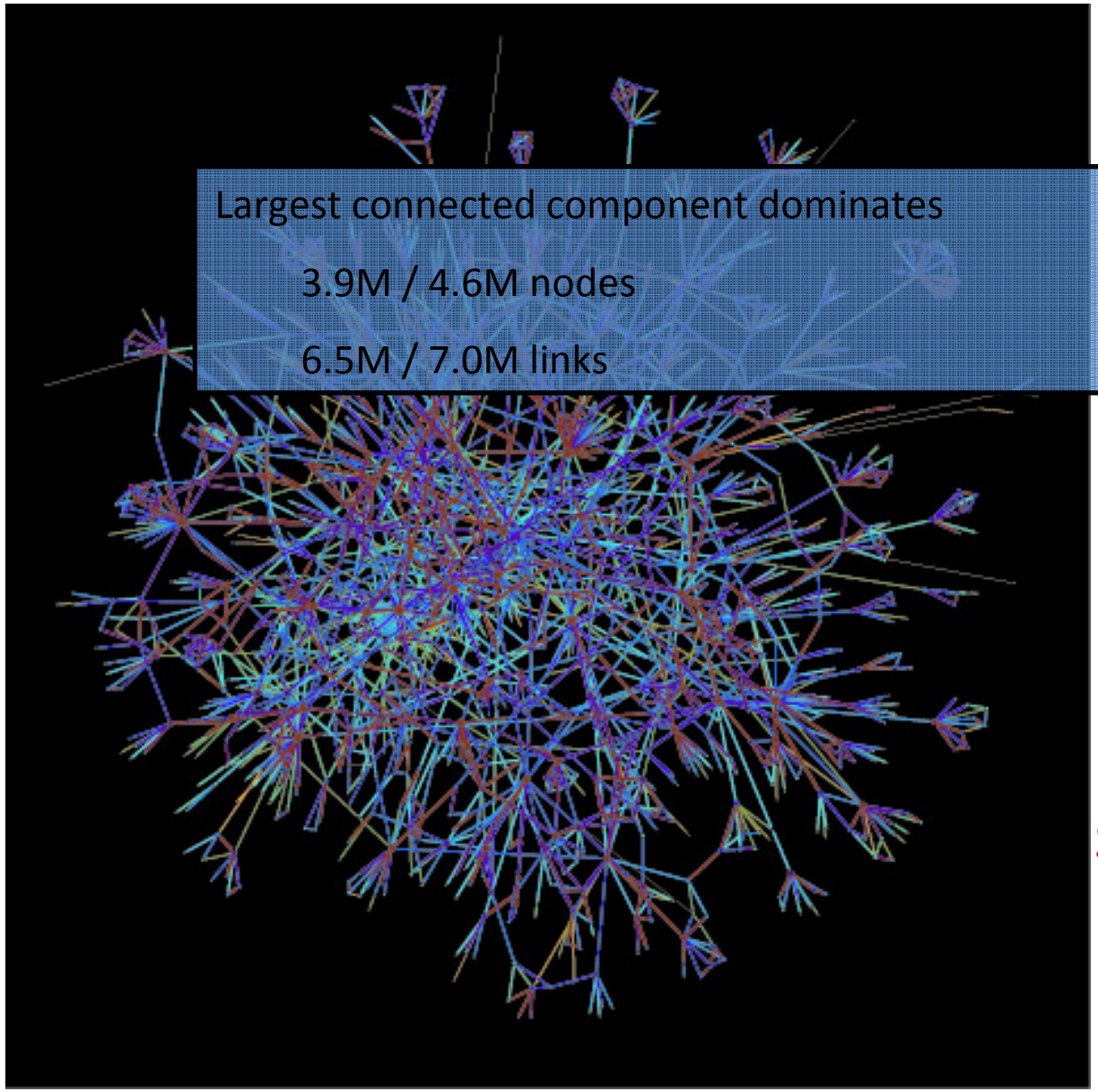
Require reciprocity (X→Y AND Y→X) for a link



J.-P. Onnela, et al. PNAS 104, 7332-7336 (2007)
J.-P. Onnela, et al. New J. Phys. 9, 179 (2007)

Weighted undirected graph

STRUCTURE AND TIE STRENGTH IN MOBILE COMMUNICATION NETWORK



Largest connected component dominates

3.9M / 4.6M nodes

6.5M / 7.0M links

Small world

STRUCTURE AND TIE STRENGTH IN MOBILE COMMUNICATION NETWORK

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

Hypothesis about the small 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 large (macro-) scale:

Society consists of strongly 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 IN MOBILE COMMUNICATION NETWORK

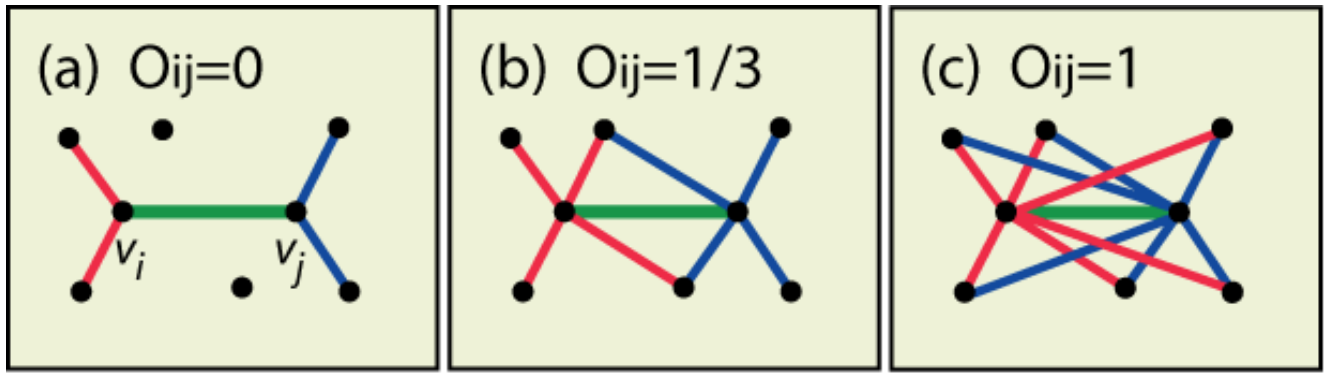
OVERLAP

- **Definition: relative neighborhood overlap (topological)**

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

where the number of triangles around edge (i,j) is n_{ij}

- Illustration of the concept:



STRUCTURE AND TIE STRENGTH IN MOBILE COMMUNICATION NETWORK

- Let $<O>_w$ denote O_{ij} averaged over a bin of *w*-values

- Use cumulative link weight distribution:
  (the fraction of links with weights less than w')

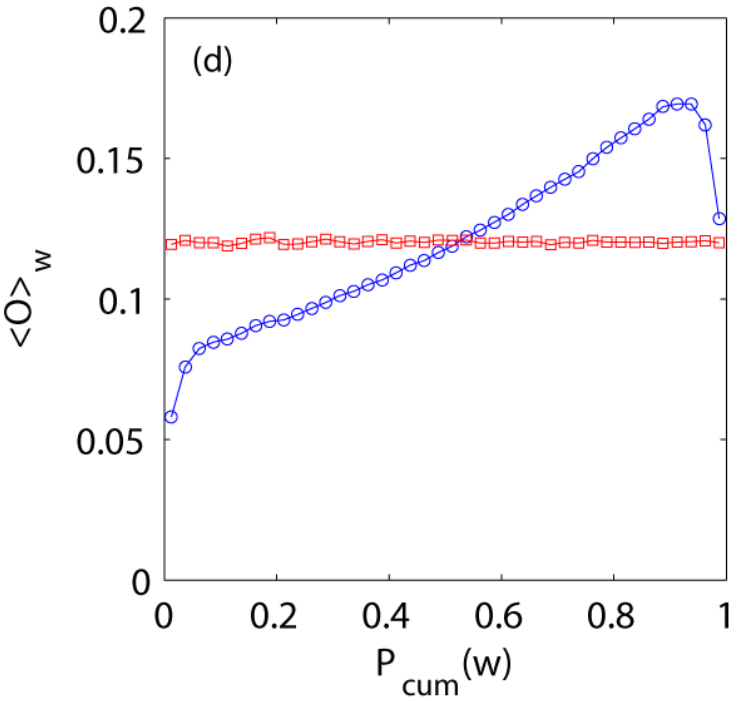
$$P_{\text{cum}}(w') = \sum_{w \le w'} P(w)$$

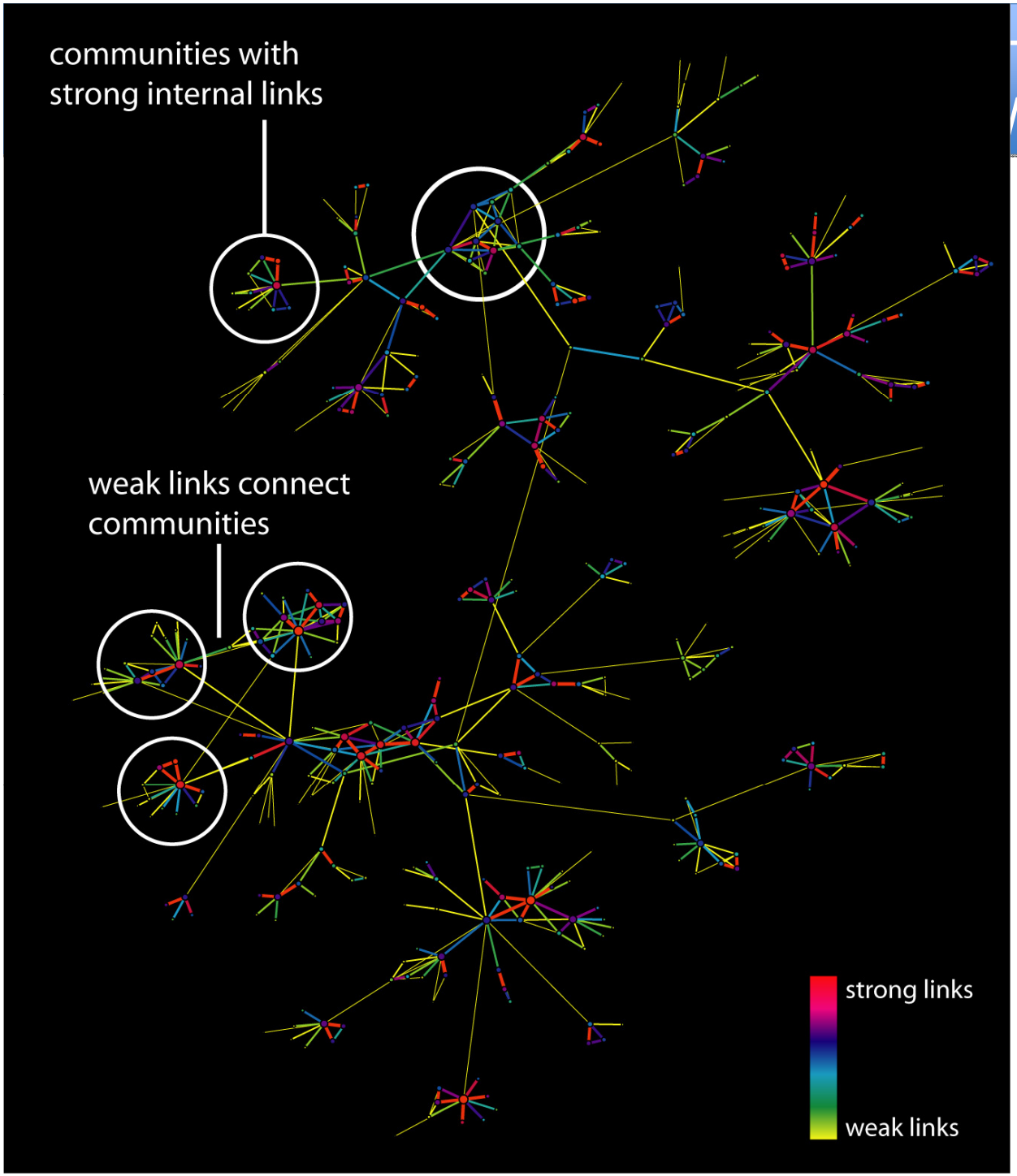
- Relative neighbourhood overlap increases as a function of link weight

⇒Verifies Granovetter's hypothesis (~95%)
(Exception: Top 5% of weights)

Blue curve: empirical network

Red curve: weight randomised network





H IN MOBILE
WORK

Granovetterian structure of the society:
Strongly wired communities are connected by weak links

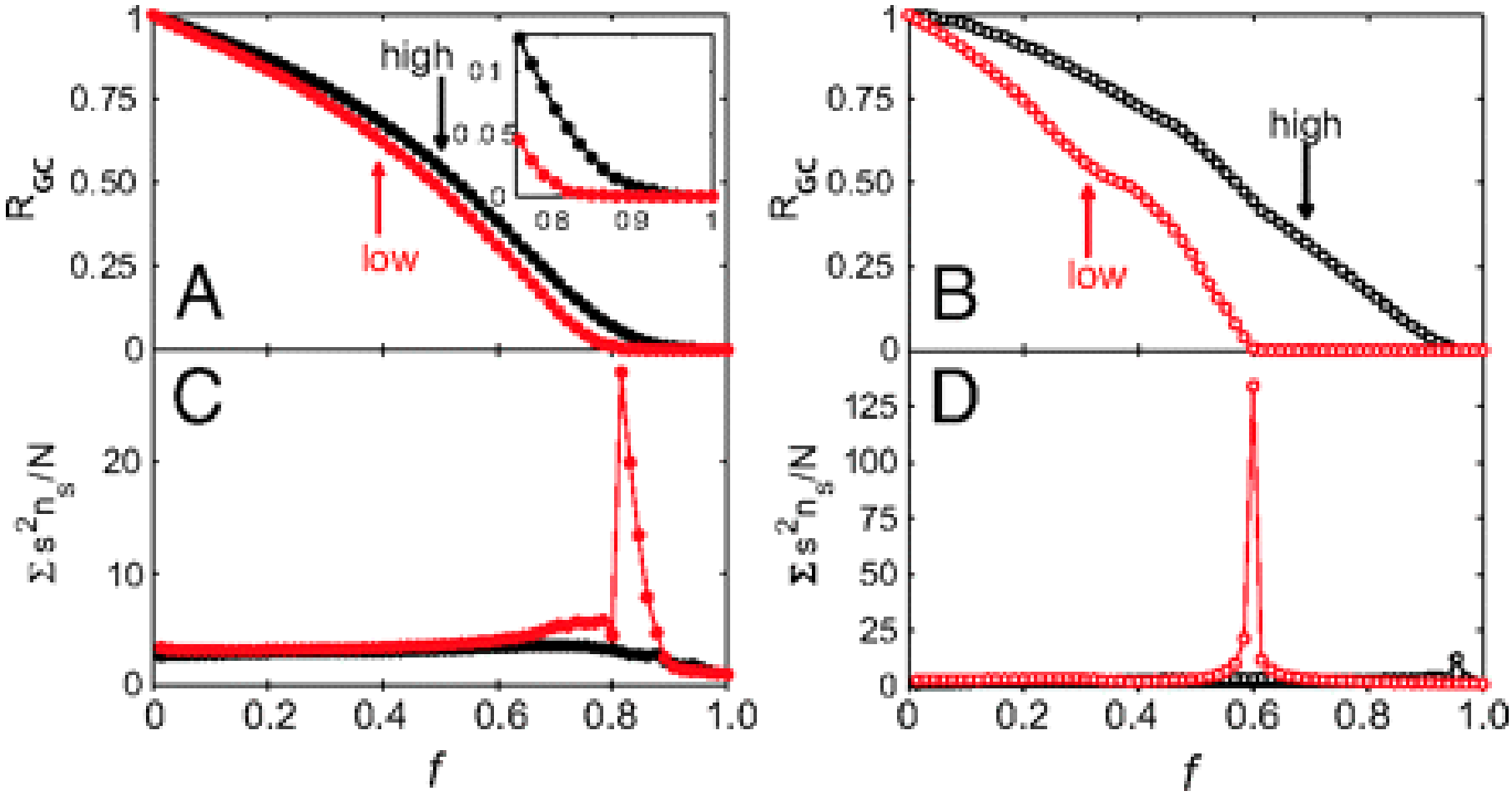
Local property with global consequences

STRUCTURE AND TIE STRENGTH IN MOBILE COMMUNICATION NETWORK

Percolation aspects

Weights

Overlap



The local relationship between weights and topology has global consequences

DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK

Spreading phenomena in networks

- epidemics (bio- and computer)
- rumors, information, opinion
- innovations
- etc.

Nodes of a network can be:
- Susceptible
- Infected
- Recovered (immune)

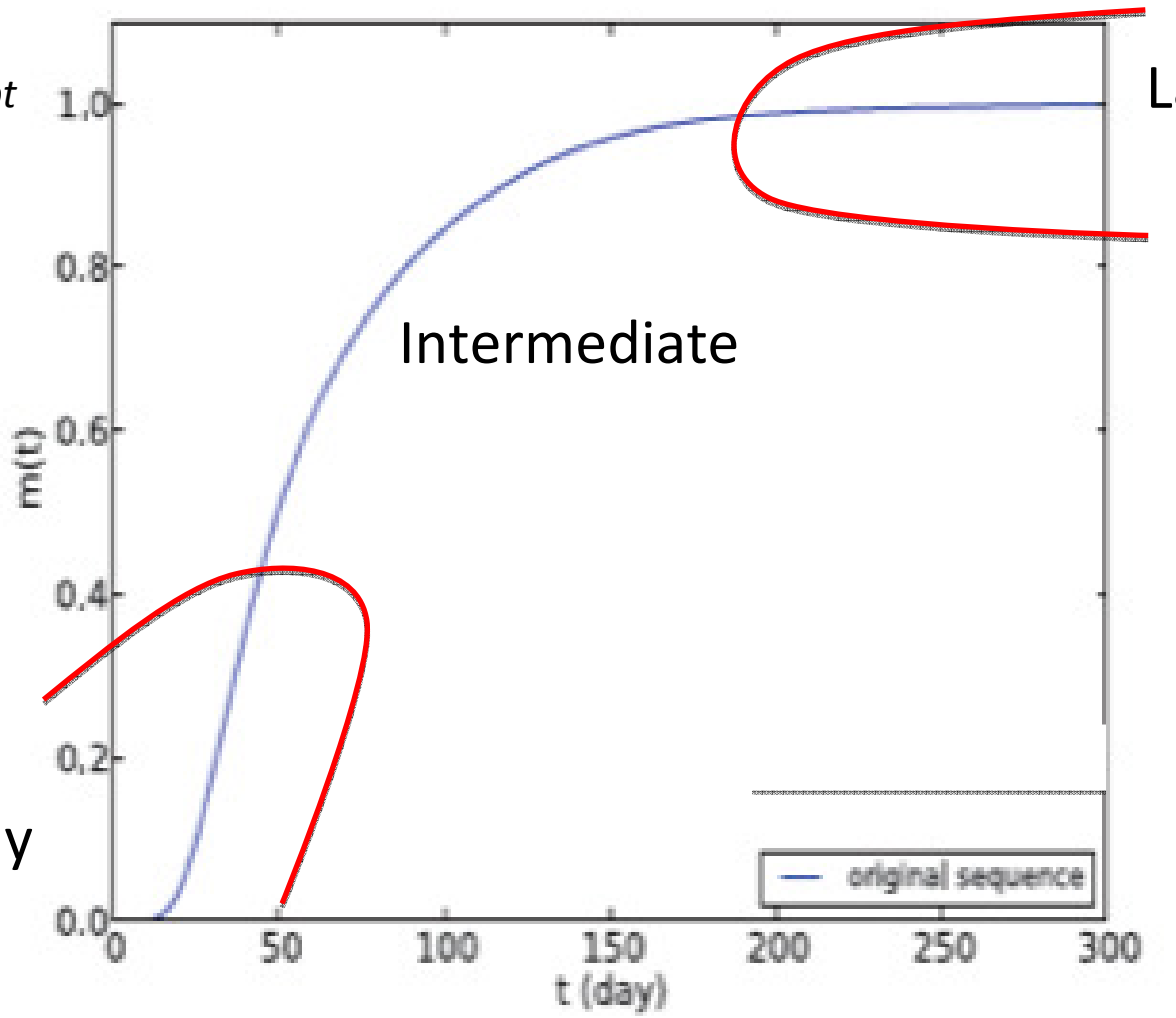
Corresponding models: SI, SIR, SIS...

Important: **speed of spreading**

DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK

Spreading curve (SI)

$m(t)=N_{inf}/N_{tot}$



DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK

Spreading in the society

Small world property; “Six Degrees of Separation”

Not only social nw-s: Internet, genetic transcription, etc.
In many networks the average distance btw two arbitrary nodes is small (grows at most log with system size).

Distance: length of shortest path btw two nodes

Small world: fast spreading?
There are short, efficient paths. Are they used?

DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK

Needed information:

- **Structure** of the society: Network at the societal level
- Local **transmission dynamics**: Detailed description, how information (rumor, opinions etc) is transmitted

Impossible to know – we use the mobile phone network

We have data about

- who called whom, voice, SMS, MMS
- when
- how long they talked

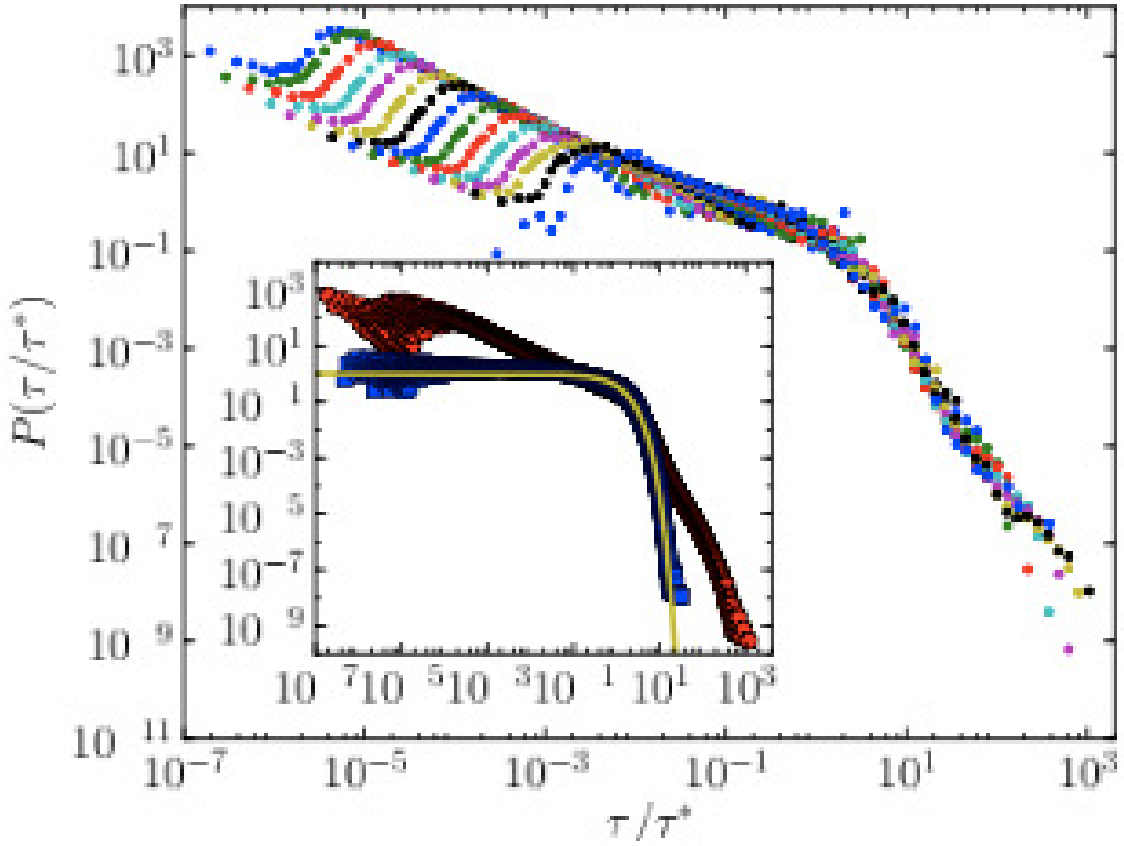
(+ metadata – gender, age, postal code
+ mostly used tower,…)

306 million mobile call records of 4.9 million individuals during 4 months with 1s resolution

Movie

M.Karsai et al. Phys. Rev. E83, 025102 (2011)

DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK



Scaled inter-event time distr.

Binned according to weights (here: number of calls)

Calls are non-Poissonian

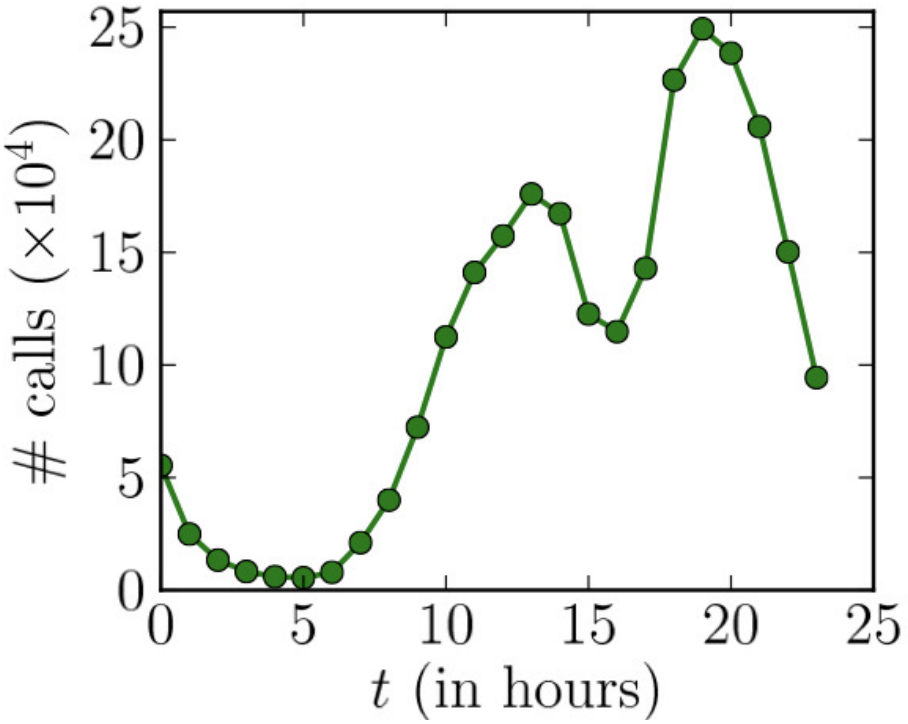
Inset: time shuffled

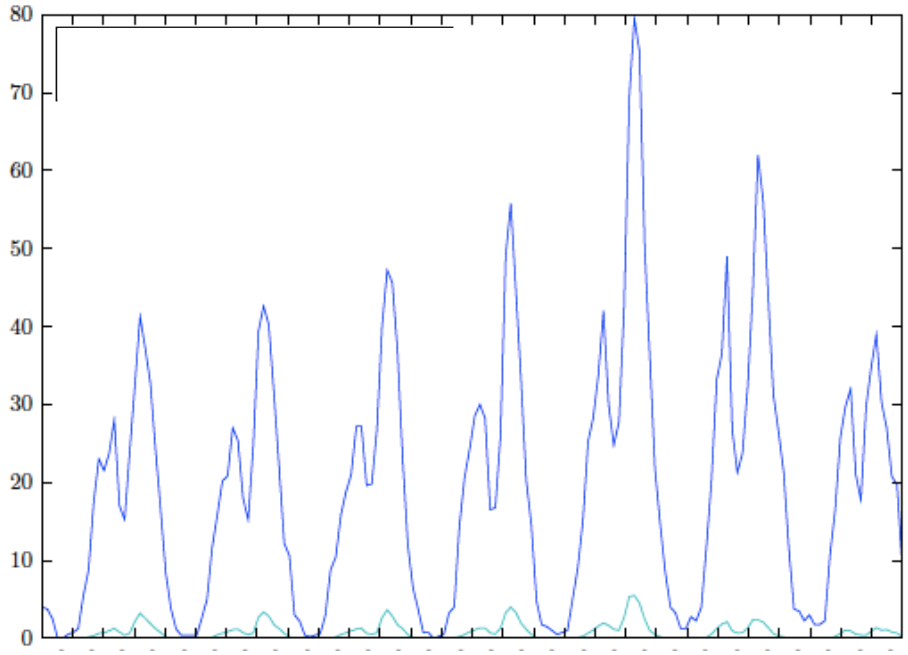
DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK

Correlations influence spreading speed

-Topology (community structure)

- Weight-topology (Granovetterian structure)
- Daily, weekly patterns
- Bursty dynamics
- Link-link dynamic correlations



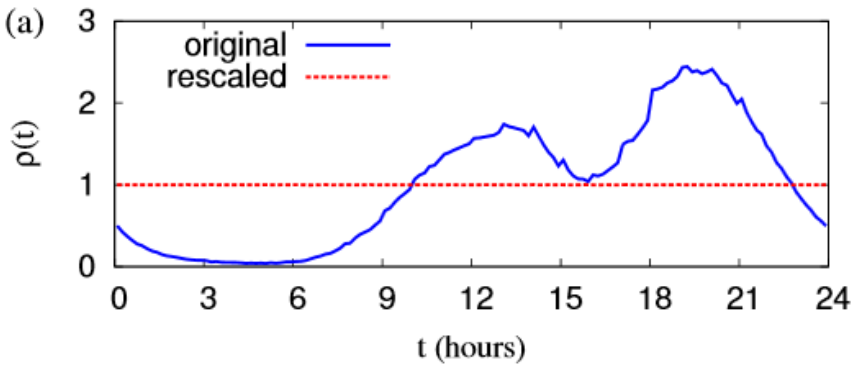


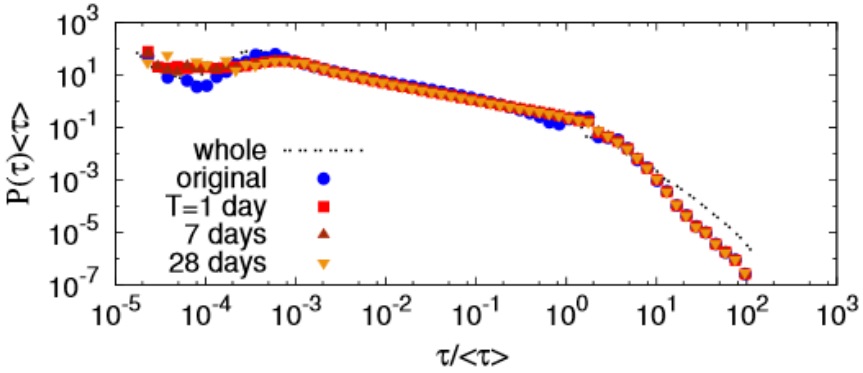
days

DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK

Deseasoning the data by time rescaling $t^*(t) = \sum_{0 \le t' < t} \rho_{\Lambda,T}(t')$

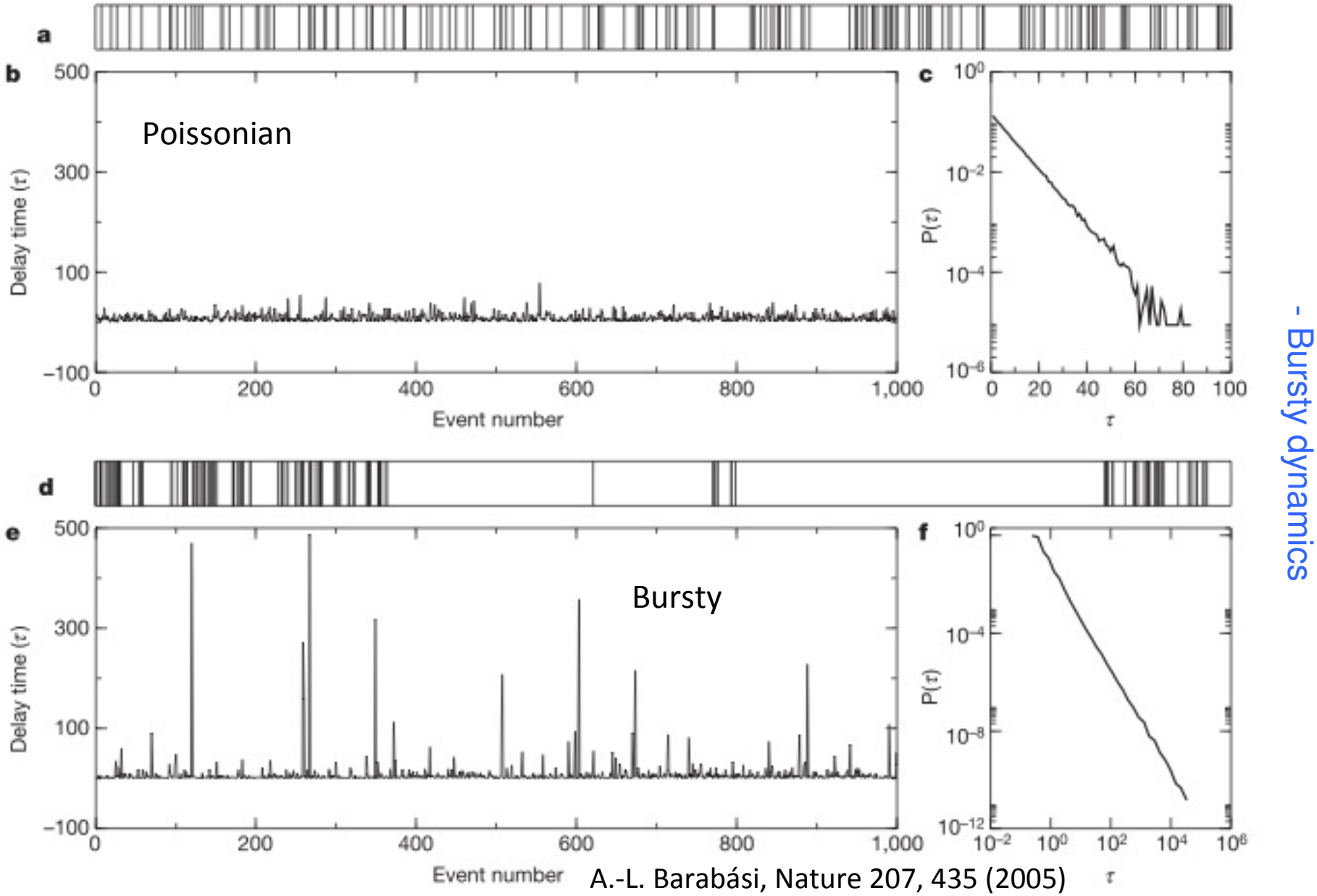
with $\rho_{\Lambda,T}(t) = \frac{T}{s_\Lambda} \sum_{k=0}^{\lfloor T_f/T \rfloor} n_\Lambda(t + kT), \; s_\Lambda = \sum_{t=0}^{T_f} n_\Lambda(t)$





Thus the non-Poissonian character is not due to the circadian pattern.

DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK

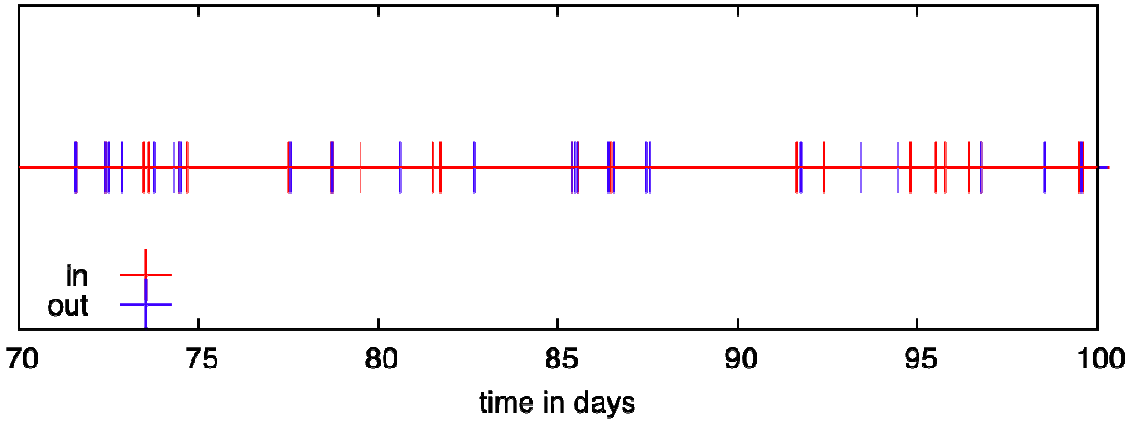


A.-L. Barabási, Nature 207, 435 (2005)

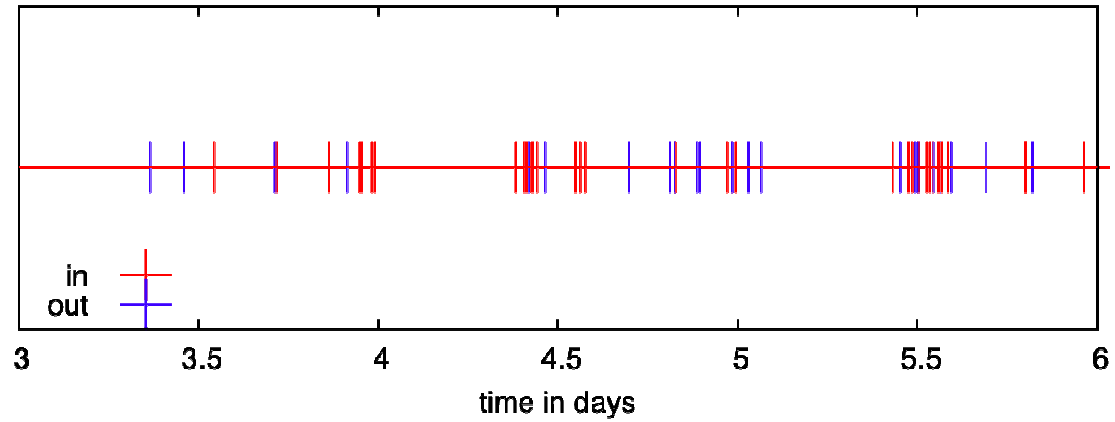
- Bursty dynamics

DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK

Bursty call patterns for individual users



Average user



Busy user

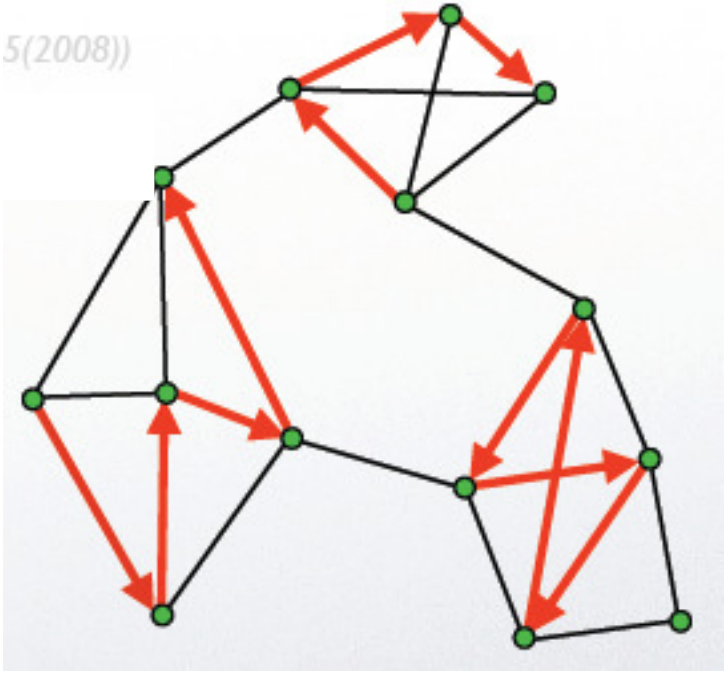
Note the different scales

DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK

- Link-link dynamic correlations

Triggered calls, cascades, etc.

Temporal motifs



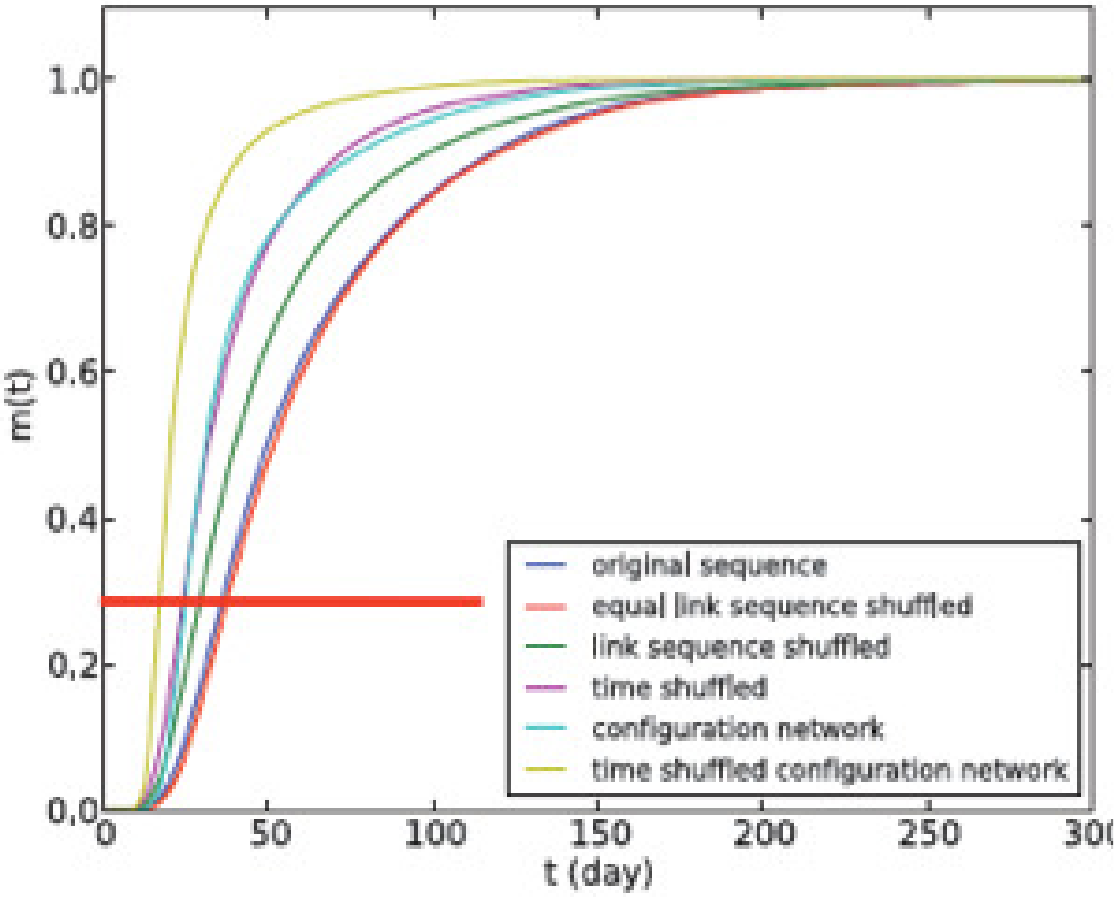
Experiment: ”Infect” a random node, the empirical call data and assume that ”infection” is transmitted by each call.
How to identify the effect of the different correlations on spreading?
Introduce different null models by appropriate shuffling of the data.

DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK

Results:

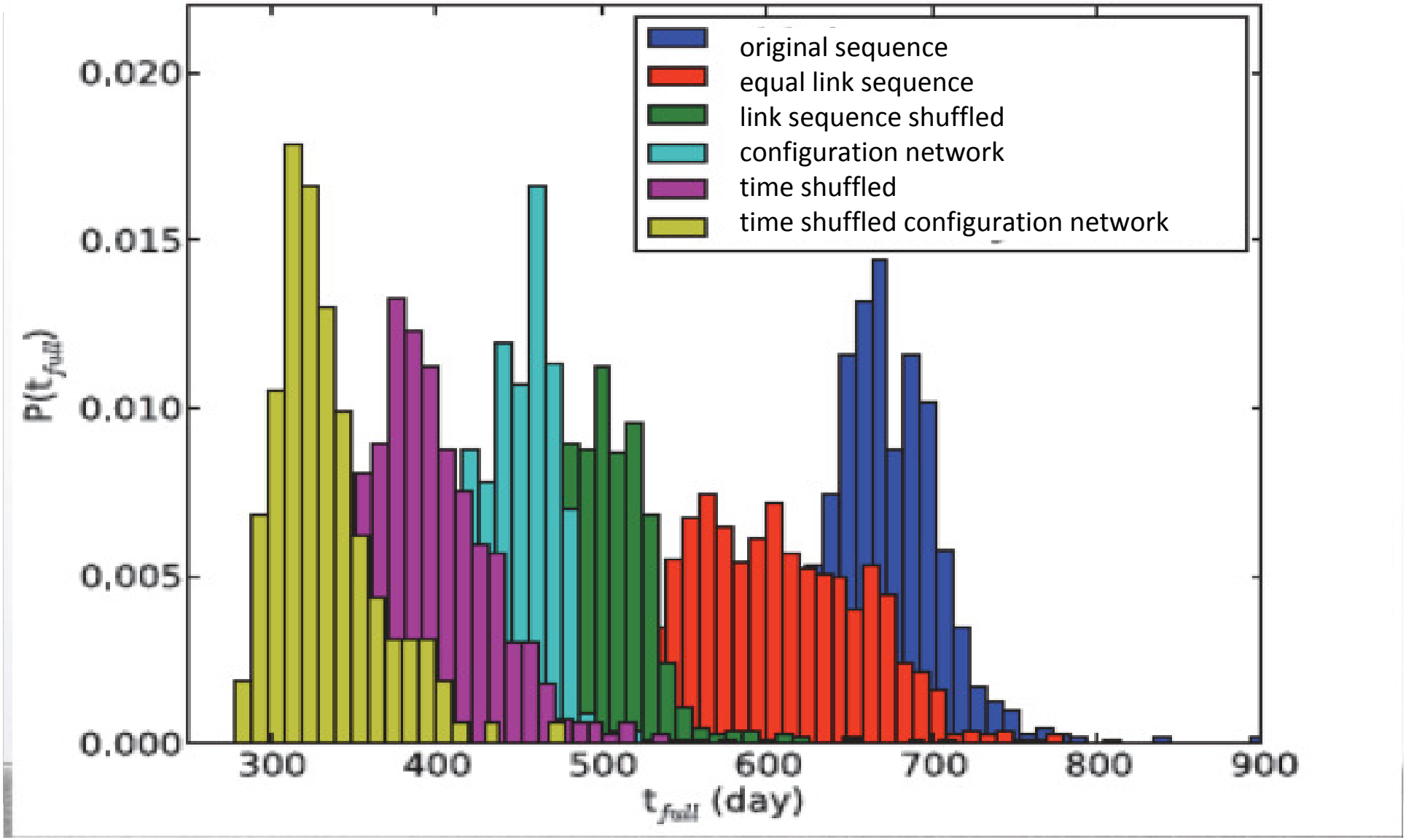
Strong slowing down due to
- topology (communities)
- link-topology correlations
- burstiness

Minor effect:
- circadian etc. patterns
- temporal motifs



DYNAMICS OF SPREADING IN MOBILE COMMUNICATION NETWORK

Long time behavior (time needed to get m=1)

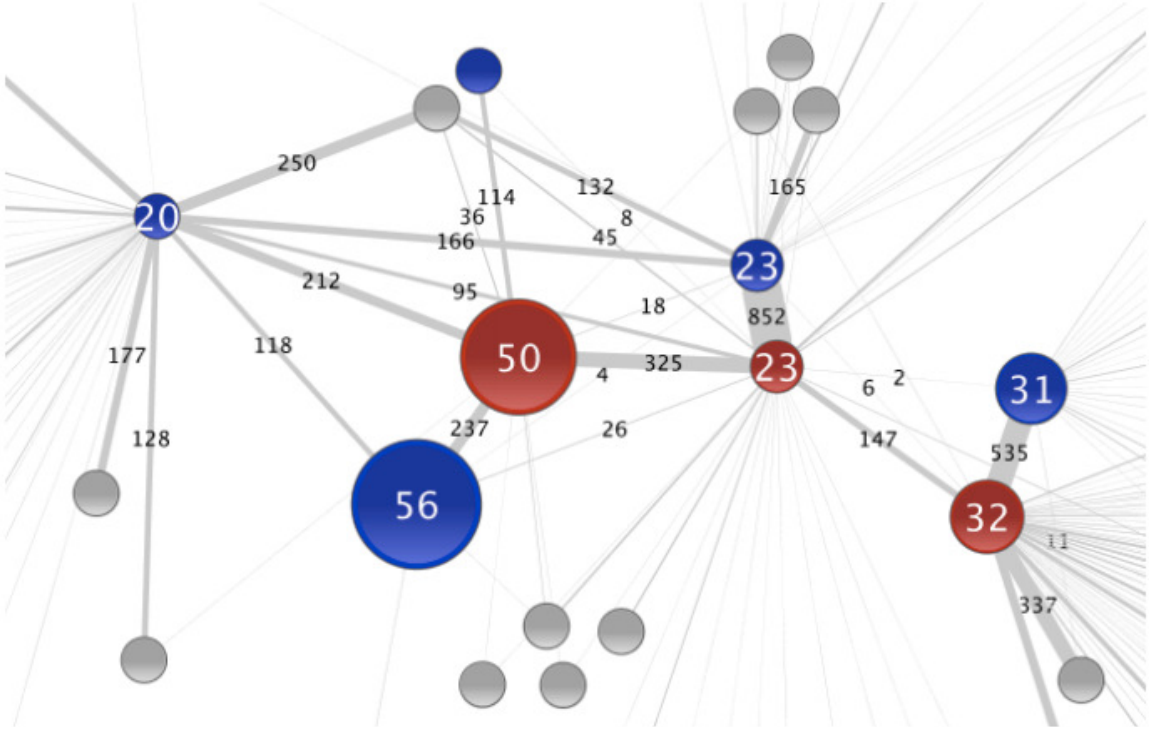


EGOCENTRIC NETWORKS: SEX DIFFERENCES IN INTIMATE RELATIONSHIPS

Data available on age and gender, most frequent locations (33.2 M users, 6.8 M within provider; 1.95 billion calls, 0.489 SMS.

Egocentric networks: All connections of a central site (ego)

Question: Are there gender and age specific properties in making connections?



We define a gender variable

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

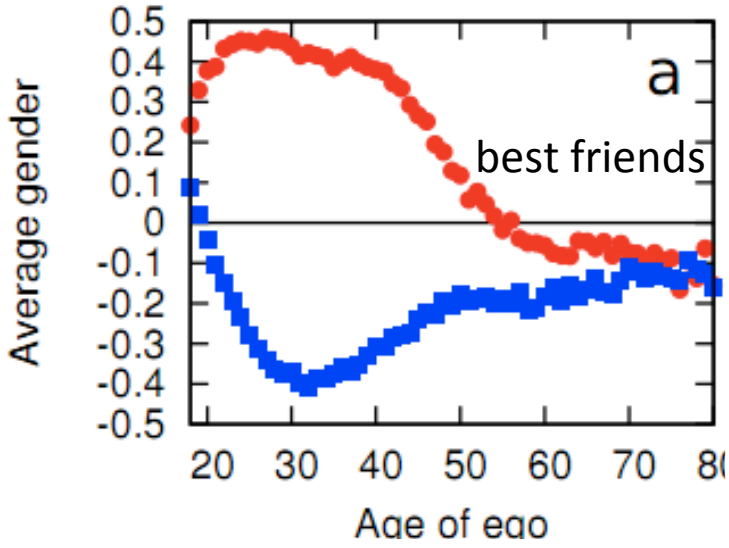
<g> = 0 means balance here

<g> = 0.13

EGOCENTRIC NETWORKS: SEX DIFFERENCES IN INTIMATE RELATIONSHIPS

Ranking of contacts according to frequency of calls:
”best friends”, ”second best friends” etc.

Ego: ■male, ●female



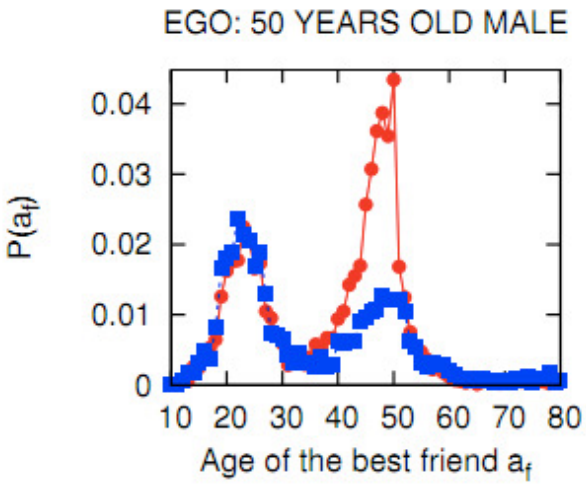
- Between the ages of 18 and 45, men and women have best friends of the opposite sex. Second best friends are generally of the same sex at this age
- Women are more focused on opposite sex relationships than men are.

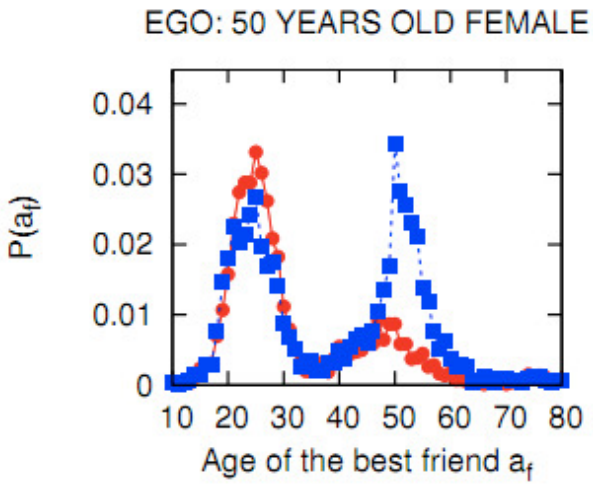
V. Palchykov et al. 2011, 2013

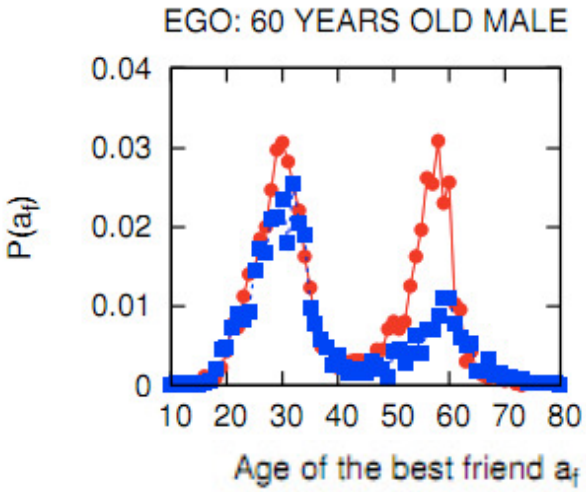
EGOCENTRIC NETWORKS: SEX DIFFERENCES IN INTIMATE RELATIONSHIPS

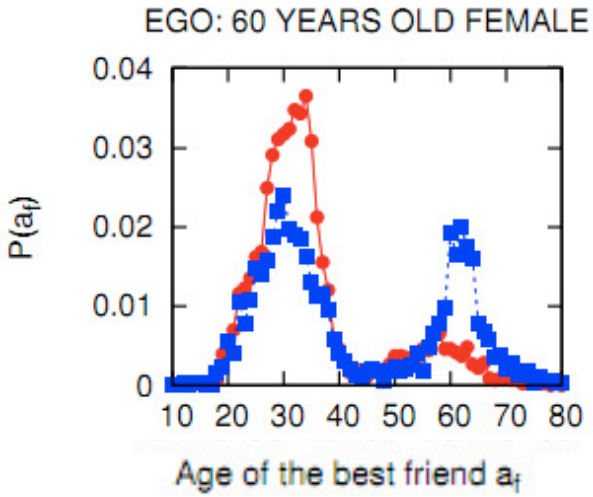
Distribution of best friends by age

- 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.









ÖSSZEFOGLALÁS

- Az információ-kommunikációs technológia hatalmas adatözönt zúdít ránk, amit hatékonyan lehet a társadalmi jelenségek tanulmányozásánál használni.
- A mobiltelefonálási adatok a társadalom szerkezetének, az emberek mozgásának leképezéséhez kiválóan felhasználhatók, mivel a lefedettség közel 100%-os és a készülékeket állandóan magunkkal hordjuk.
- A társadalom szerkezetét jól írja le a Granovetter-hipothézis A gyenge kötések fontossága
- A társadalom kis-világ jellege ellenére a terjedés a szociális hálózaton nem gyors a kapcsolatok inhomogén, villanásos jellege és a granovetteri szerkezet miatt.
- Az adatok elemzése bevilágít a kapcsolatok kortól és nemtől való függésébe.