Temporal networks of human (face-to-face) interactions

Alain Barrat

Centre de Physique Théorique, Marseille, France

ISI Foundation, Turin, Italy

http://www.cpt.univ-mrs.fr/~barrat
http://www.sociopatterns.org
alain.barrat@cpt.univ-mrs.fr

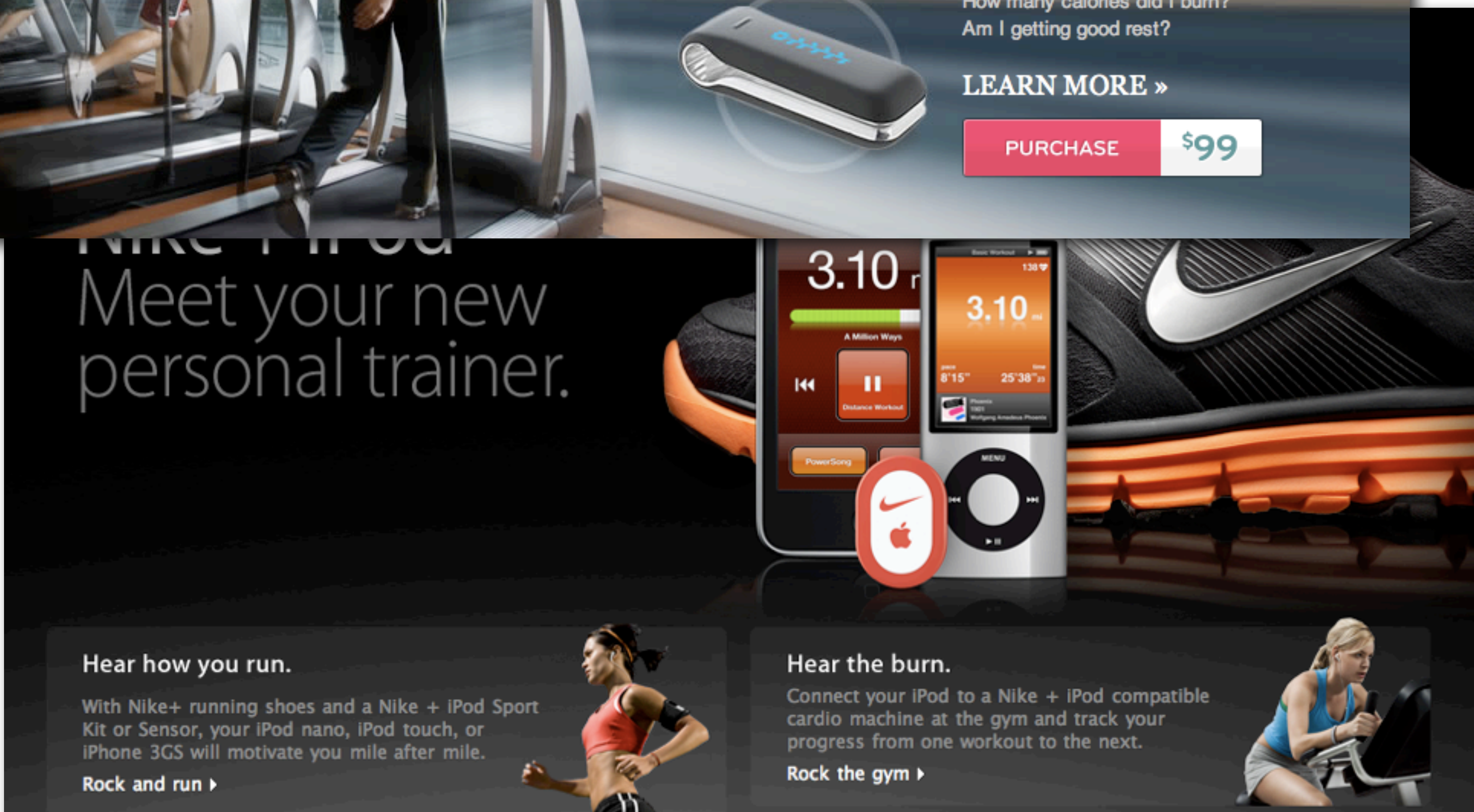
- Introduction
- SocioPatterns
  - infrastructure and resulting data
  - data main features, comparing datasets
- Representing data, finding structures
  - aggregation procedures, temporal motifs, NTF, detecting timescales, temporal cores, backbones
- Models
  - Models of temporal networks
  - Models of f2f interactions
  - Null models for temporal networks
- Processes on temporal networks
- Using data
  - studying human behaviour
  - data-driven simulations of epidemic spreading:
    - how much details?
    - interventions
- Using incomplete data or different proxies

New technologies

↓

digital *traces* of human behavior


fitbit.com
fitbit automatically tracks your
fitness & sleep
Did I get enough exercise today?
How many calories did I burn?
Am I getting good rest?
LEARN MORE »
PURCHASE
$99


Meet your new personal trainer.
3.10
A Million Ways
Distance Workout
PowerSong
MENU
Hear how you run.
With Nike+ running shoes and a Nike + iPod Sport Kit or Sensor, your iPod nano, iPod touch, or iPhone 3GS will motivate you mile after mile.
Rock and run
Hear the burn.
Connect your iPod to a Nike + iPod compatible cardio machine at the gym and track your progress from one workout to the next.
Rock the gym


Ciro
Find Friends
Add Things
Apps
Help
Settings
foursquare
Find places, people, tags
SEARCH
ME
HISTORY
STATS
FRIENDS
Currently in Maratea, Basilicata
?
Welcome to foursquare!
We're all about helping you find new ways to explore the city.
We'll help you meet up with your friends and let you earn points and unlock badges for discovering new places, doing new things and meeting new people.
[Learn more!]
If you own a bar or restaurant, foursquare can help
Available on the iPhone App Store
It's on android
BlackBerry App World
Newly Crowned Mayors near Marate
Stefan C. @ Villa Del Mare Centro Congressi Maratea
Giuseppe S. @ Pepp1land
Raffaella C. @ Scario
Gianpaolo D. @ Circolo
Paolo M. @ L'uorto
Starbucks - Folsom
Check in
People
Details
Tips
Your check-in score:
Total: 6 pts

New technologies

↓

digital *traces* of human behavior

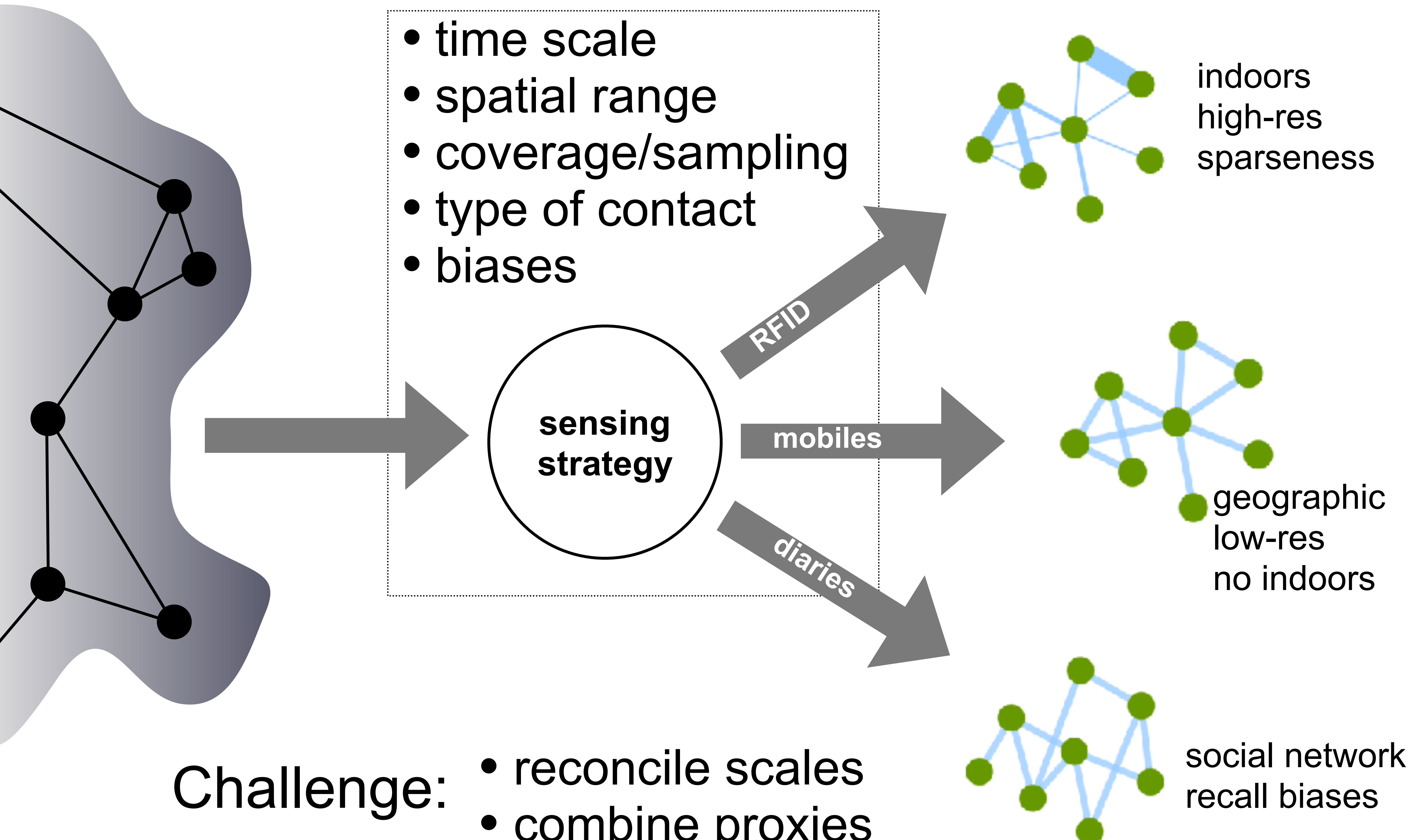
↓

proxies of human behaviour and interactions
in different contexts,
often with temporal resolution

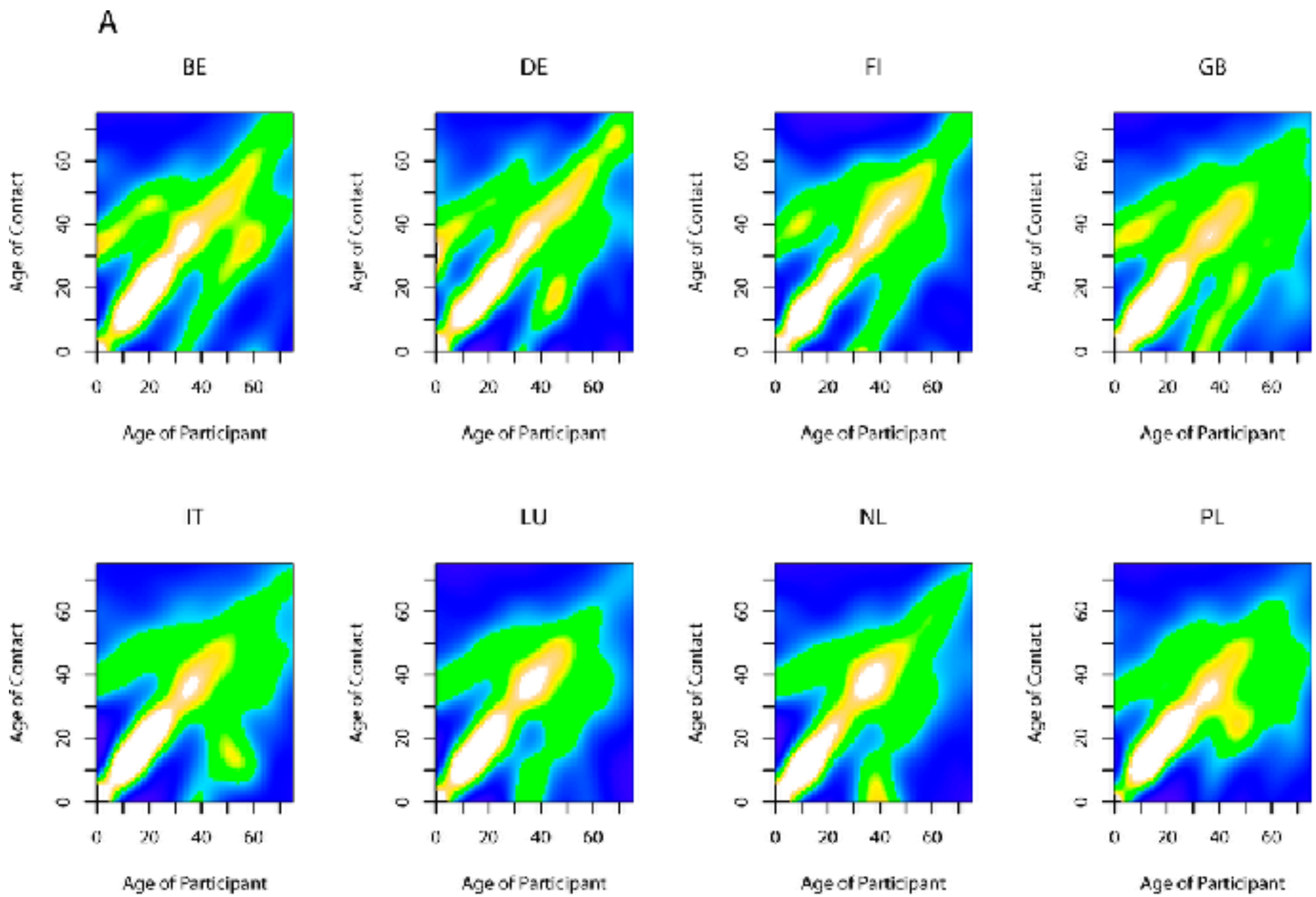
Measuring different types of interactions

- phone calls
- messaging
- online interactions
- co-presence, proximity
- face-to-face contacts

Proxies of contact networks



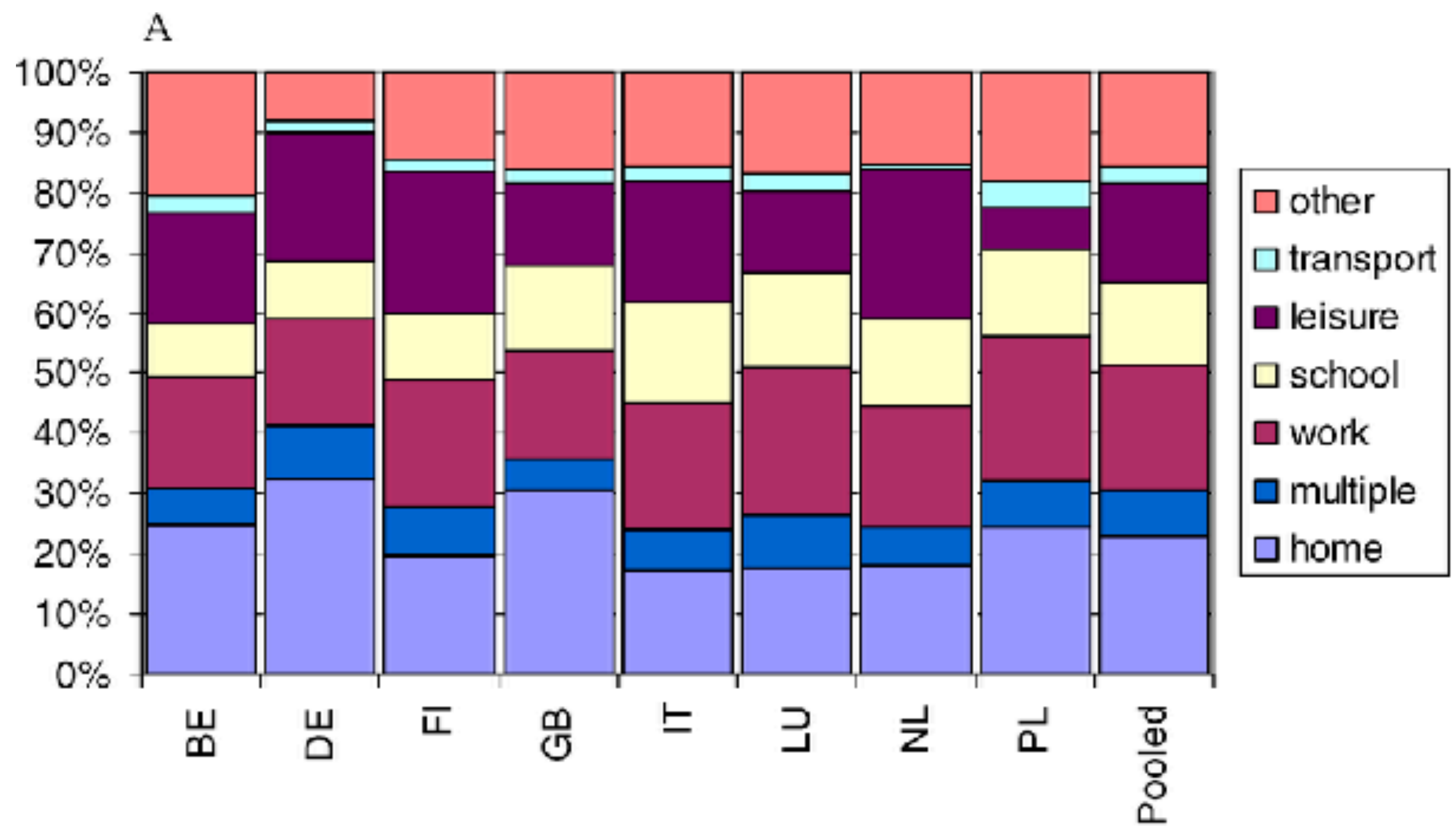
Surveys/diaries



7290 participants

"Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases", Mossong et al., PLOS Medicine (2008)

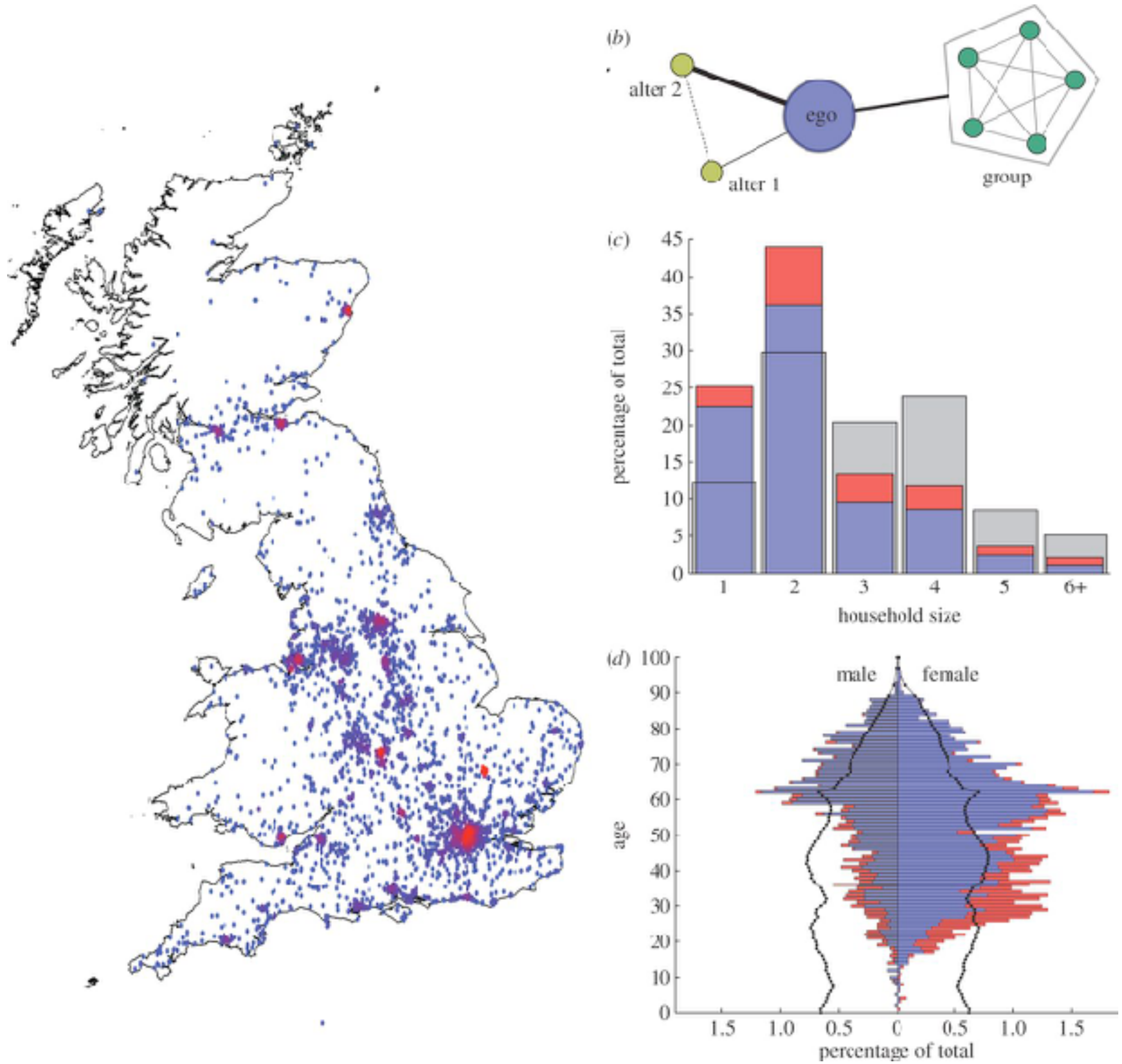
Surveys/diaries



"Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases", Mossong et al., PLOS Medicine (2008)

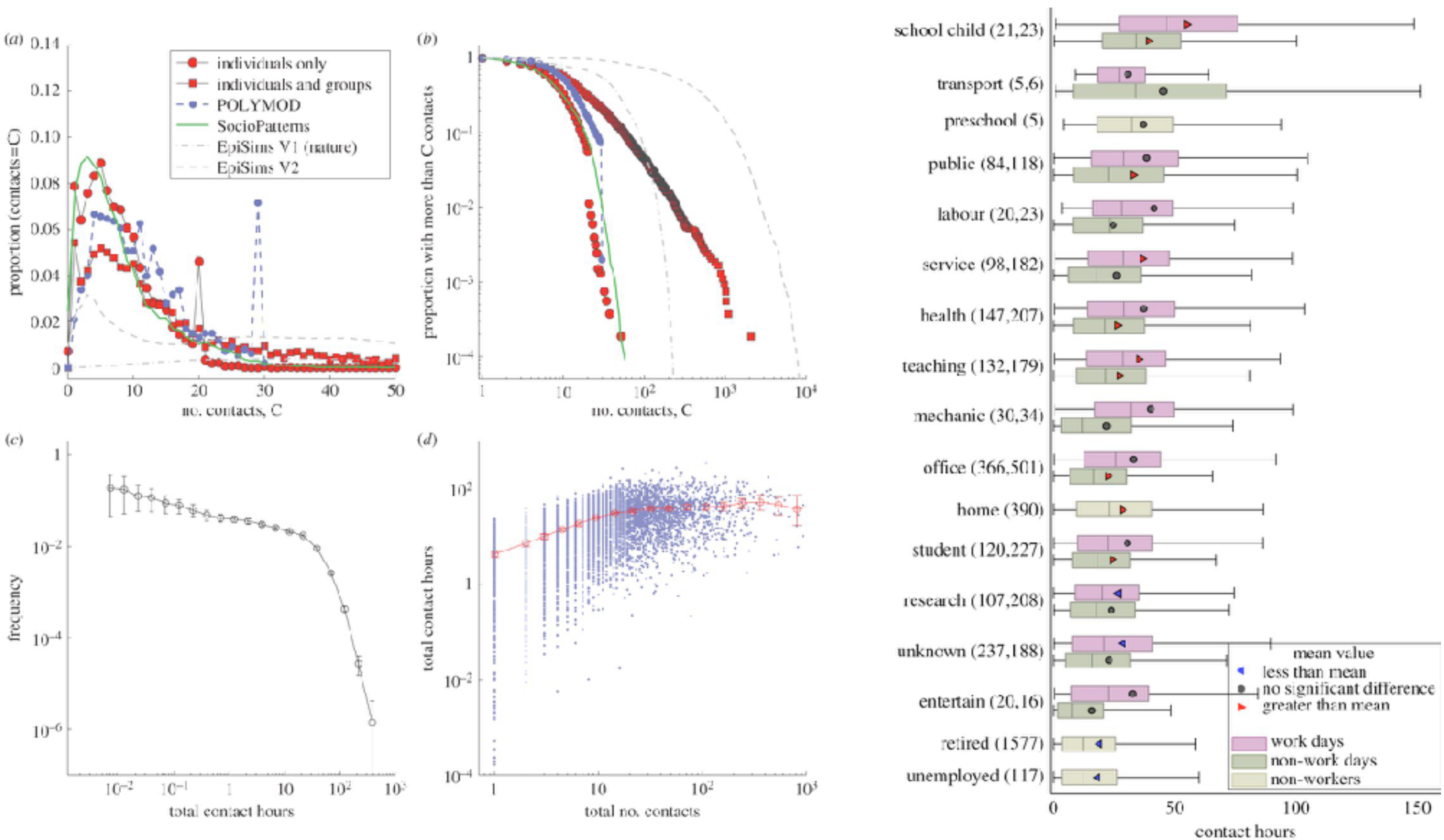
Surveys/diaries

140000 surveys sent
5000 answers



“Social encounter networks: characterizing Great Britain”, Danon et al., Proc. Roy. Soc. B (2013)

Surveys/diaries



“Social encounter networks: characterizing Great Britain”, Danon et al., Proc. Roy. Soc. B (2013)

Exploring human proximity patterns with wearable sensors

- Bluetooth, wifi (O' Neill et al. 2006, Scherrer et al. 2008, Eagle and Pentland 2009)
- MIT Reality Mining Project (sociometric badges)
- MOSAR european project (nosocomial infections)
- P. Polgreen's group
- M. Salathé group (Salathé et al. 2010)
- SMART study in schools (U. of Pittsburgh & CDC)
- smartphones, S. Lehmann's group at DTU

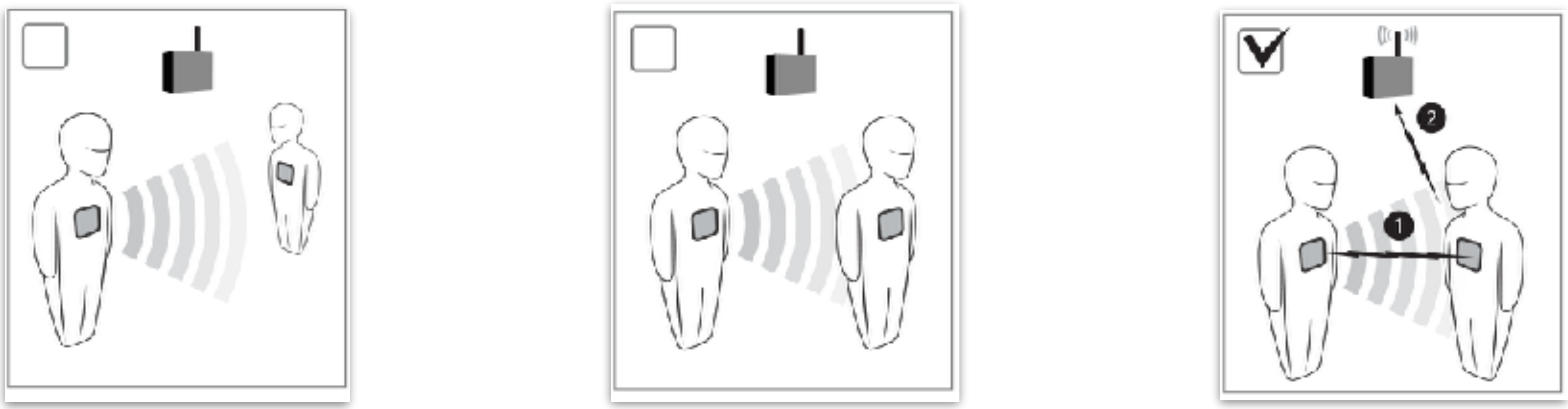
- SocioPatterns collaboration

SocioPatterns collaboration
(started 2008)

Ciro Cattuto
Jean François Pinton
Marco Quaggiotto
Alberto Tozzi
Juliette Stehlé
Lorenzo Isella
Anna Machens
Laetitia Gauvin
Mark Pachucki
Caterina Rizzo
Nicolas Voirin
Michele Starnini
Valerio Gemmetto
Kun Zhao
Julie Fournet
…

Wouter Van den Broeck
Alessandro Vespignani
Vittoria Colizza
Milosch Meriac
Brita Meriac
Martin Szomszor
Harith Alani
André Panisson
Romualdo Pastor-Satorras
Philippe Vanhems
Michele Tizzoni
Ginestra Bianconi
Mathieu Génois
Christian Vestergaard
Rossana Mastrandrea
…

Contact detection infrastructure: wearable sensors



Active RFID:
Emission of very low power data packets
=> Reception only at short distances (1-2m)

- Face to face situation
- Statistical detection => 20s time resolution
- Small
- Scalable

- (mostly) indoors

SocioPatterns.org

SocioPatterns.org

10 years, 35+ deployments, 12 countries, 50,000+ subjects

time-resolved proximity networks across a variety of relevant real-world settings

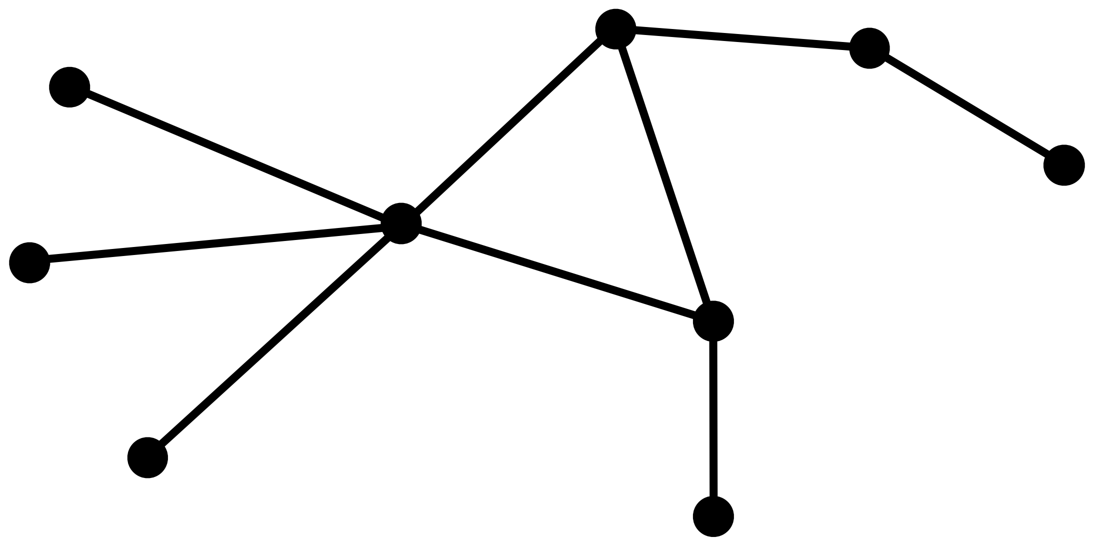
DATE	DEPLOYMENT	TYPE	# PERSONS	DURATION
Dec 2008	Chaos Communication Congress, Berlin, DE	conference	600	4 days
Apr-Jul 2009	INFECTIOUS, Science Gallery, Dublin, IE	museum	30 000	3 months
Jun 2009	ESWC2009, Crete, GR	conference	180	4 days
Jun 2009	Société Française d'Hygiène Hospitalière, Nice, FR	conference	350	2 days
Jul 20[illegible]	ACM Hypertext 2009 conference, Torino, IT	conference	120	3 days
Oct [illegible]	[illegible]imary school, Lyon, FR	school	250	2 days
N[illegible]	[illegible]ù Pediatric Hospital, Rome, IT	hospital	250	10 days
Jun 2010	[illegible]	conference	200	4 days
Apr 2010	Pra[illegible]ES	museum	100	10 days
Jul 2010	H-Farm Venture[illegible]	company	200	6 weeks
Dec 2010	hospital, Lyon, FR	hospital	100	1 week
May 2011	KEMRI Wellcome Trust, Kilifi, KE	[illegible]lds	100	2 weeks
Nov 2011	high school, Marseille, FR	[illegible]	[illegible]	1 week
Feb 2012	hospital, Lyon, FR	[illegible]	[illegible]	10 days
Mar-May 2012	primary schools, San Francisco, CA, US	school	[illegible]	[illegible]
Nov 2012	high school, Marseille, FR	school	200	[illegible]
Mar 2013	primary school, Hong Kong, HK	school	1000	[illegible]ks
June 2013	InVS, Paris, France	offices	100	2 weeks
Nov 2013	high school, Marseille, FR	school	300	10 days
July 2014	hospital, Marseille, FR	hospital	30	12 days
Mars 2015	InVS, Paris, France	offices	300	2 weeks
June 2015	hospital, Marseilles	hospital	40	4 weeks

www.sociopatterns.org/datasets

(Digression)
Data = Temporal networks

Static networks

set V of nodes joined by links (set E)

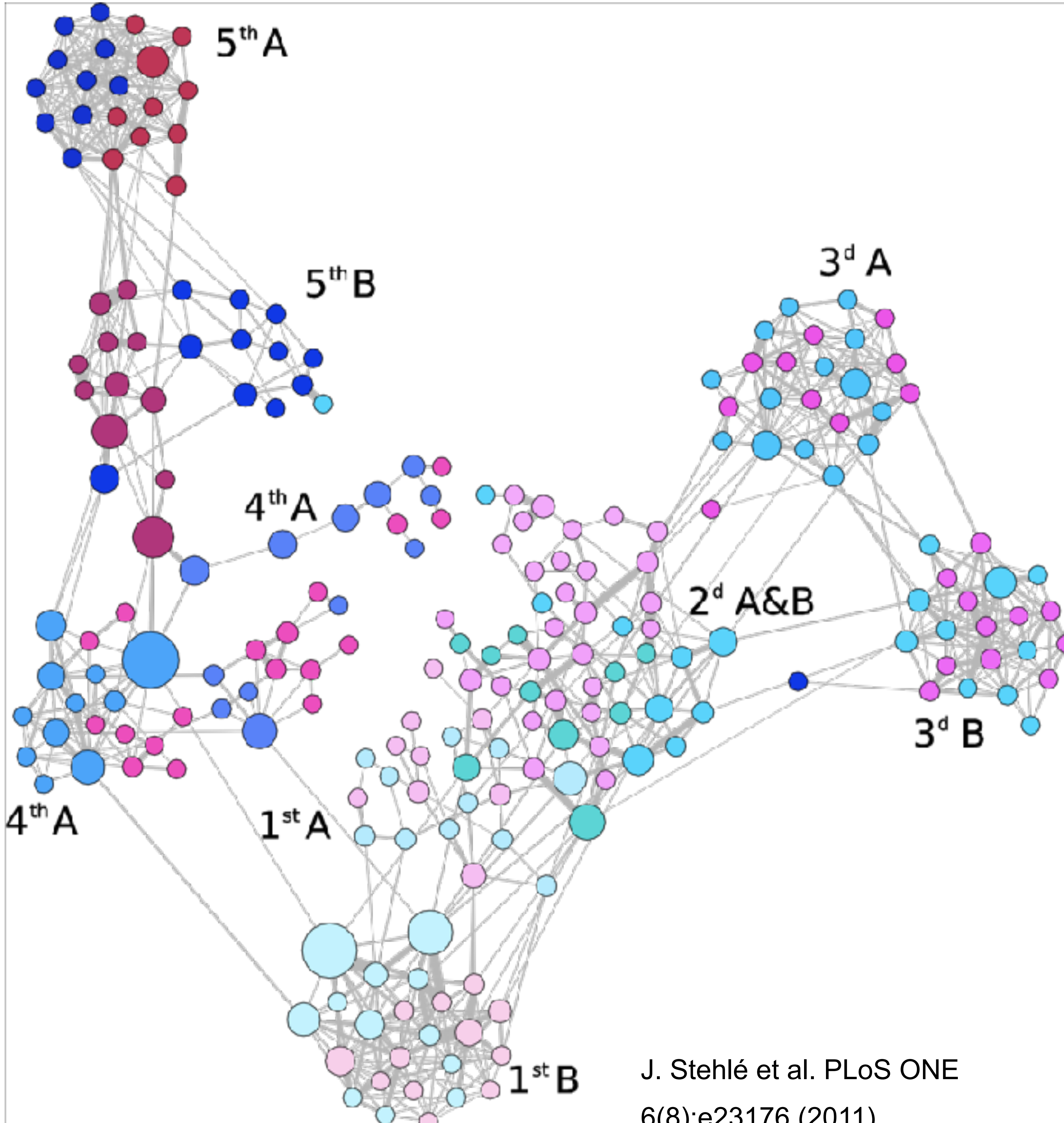


>Networks change over time

Exemples of temporal networks

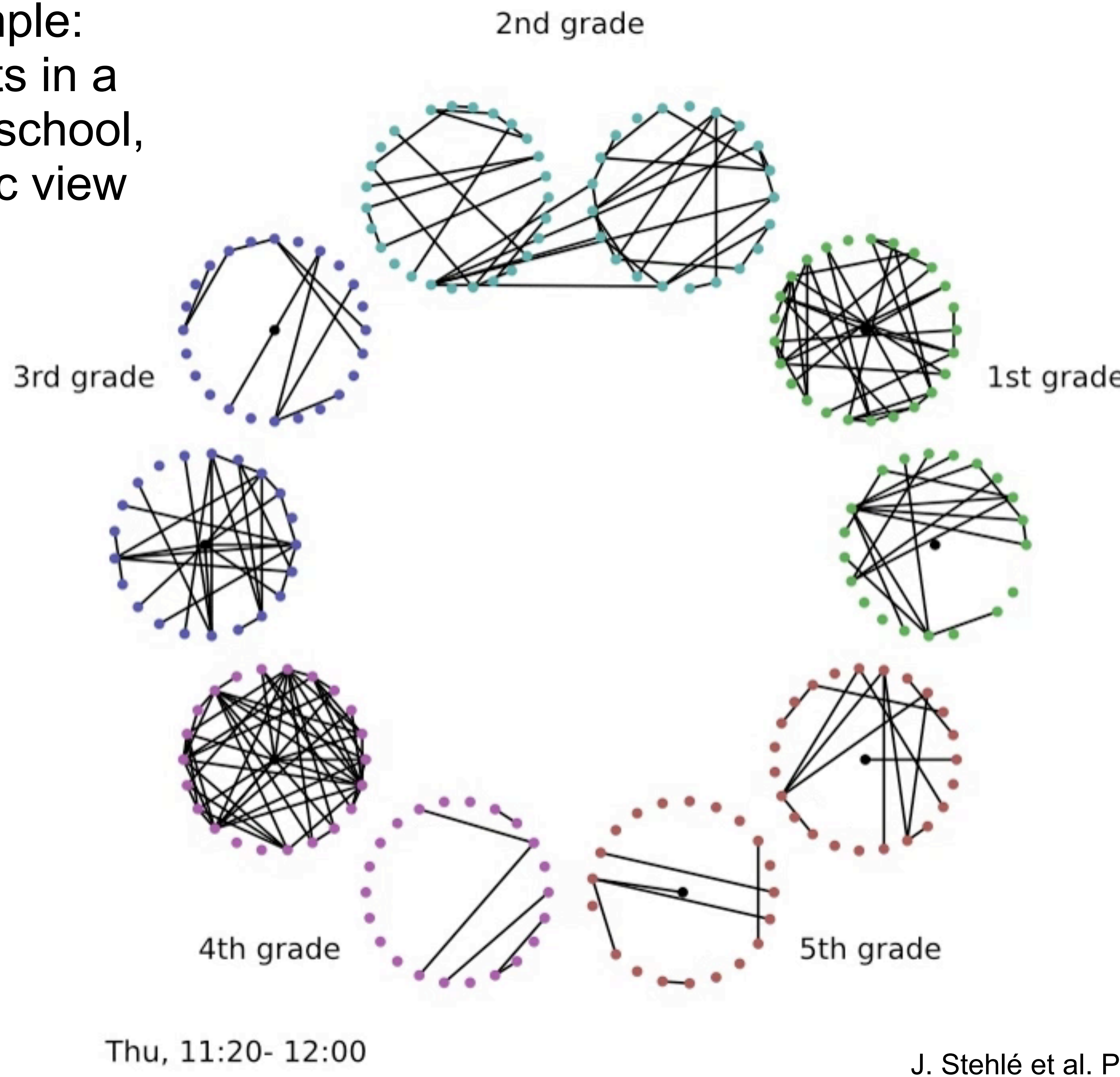
- Social networks
  - friendships
  - collaborations
- Communication networks
  - cell phone data
  - online social networks (Twitter, etc)
- Transportation networks
  - air transportation
  - animal transportation
- …

Example: contacts in a primary school, static view



J. Stehlé et al. PLoS ONE 6(8):e23176 (2011)

Example: contacts in a primary school, dynamic view



J. Stehlé et al. PLoS ONE 6(8):e23176 (2011)

Definition: temporal network

Temporal network: T=(V,S)

- V=set of nodes
- S=set of event sequences assigned to pairs of nodes

$$s_{ij} \in S: \; s_{ij} = \{(t_{ij}^{s,1}, t_{ij}^{e,1}) \cdots (t_{ij}^{s,\ell}, t_{ij}^{e,\ell})\}$$

Other representation:
time-dependent adjacency matrix:
a(i,j,t)= 1 <=> i and j connected at time t

Representations of temporal networks

Contact sequences

Time	ID1	ID2
2	2	4
2	1	5
3	2	4
3	1	6
4	2	3
5	2	4
5	1	4
8	4	6

....

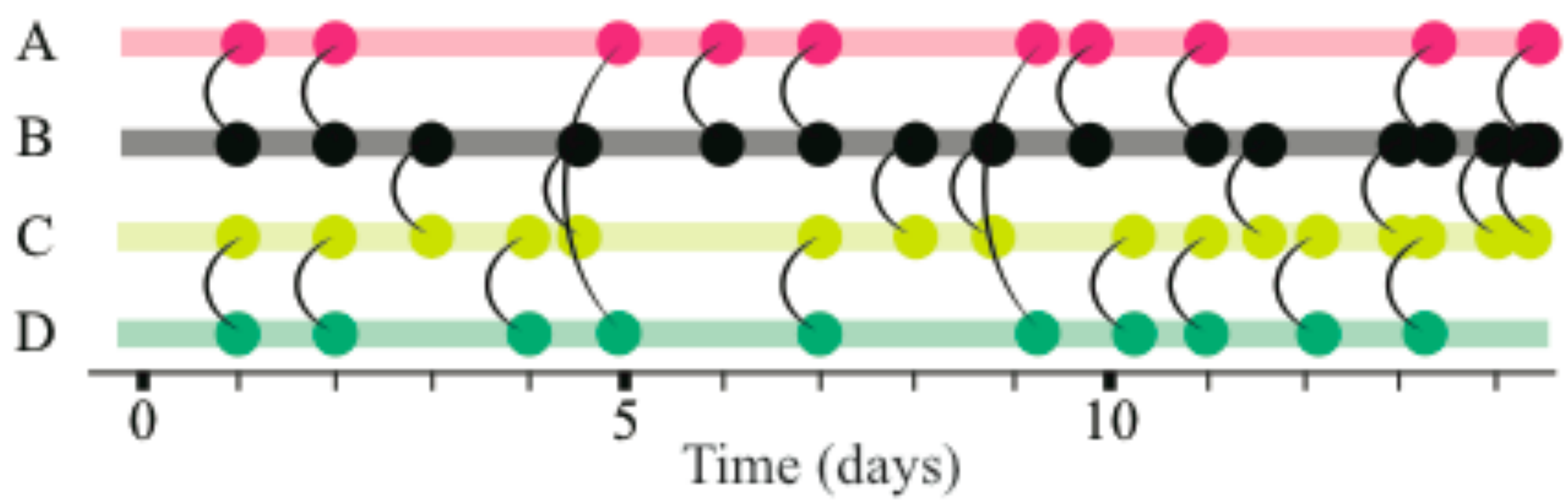
Contact intervals

ID1	ID2	Time interval
2	3	[1,5]
2	1	[2,4]
4	6	[5,9]
1	3	[7,15]
5	3	[7,9]

....

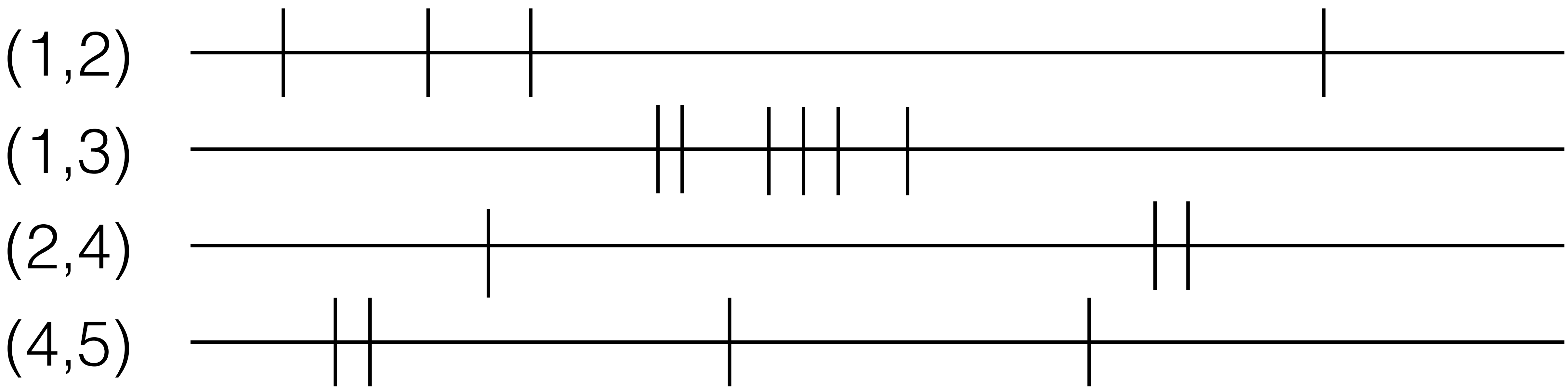
Representations of temporal networks

Timelines of nodes

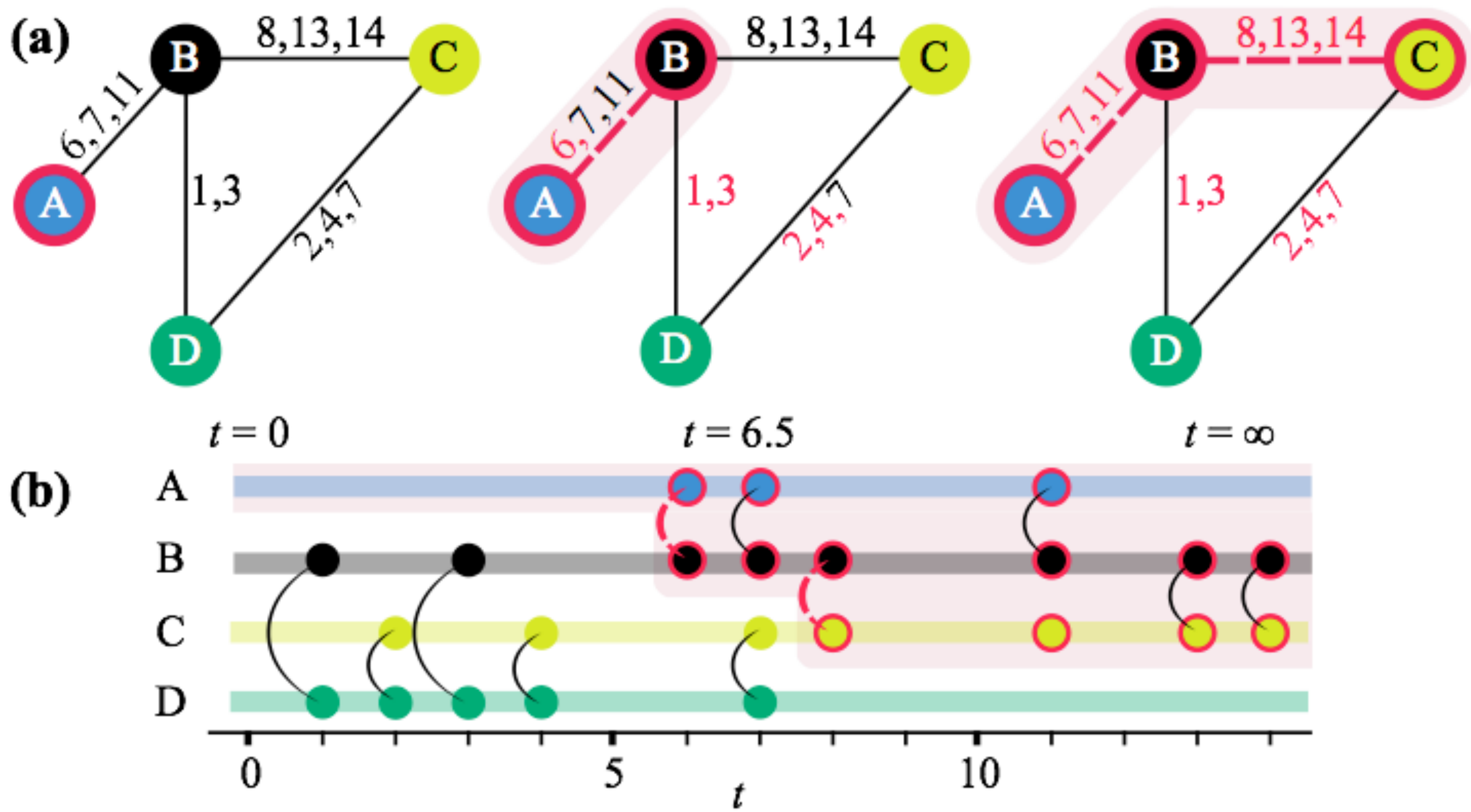


Representations of temporal networks

Timelines of links



Temporality matters: reachability issue



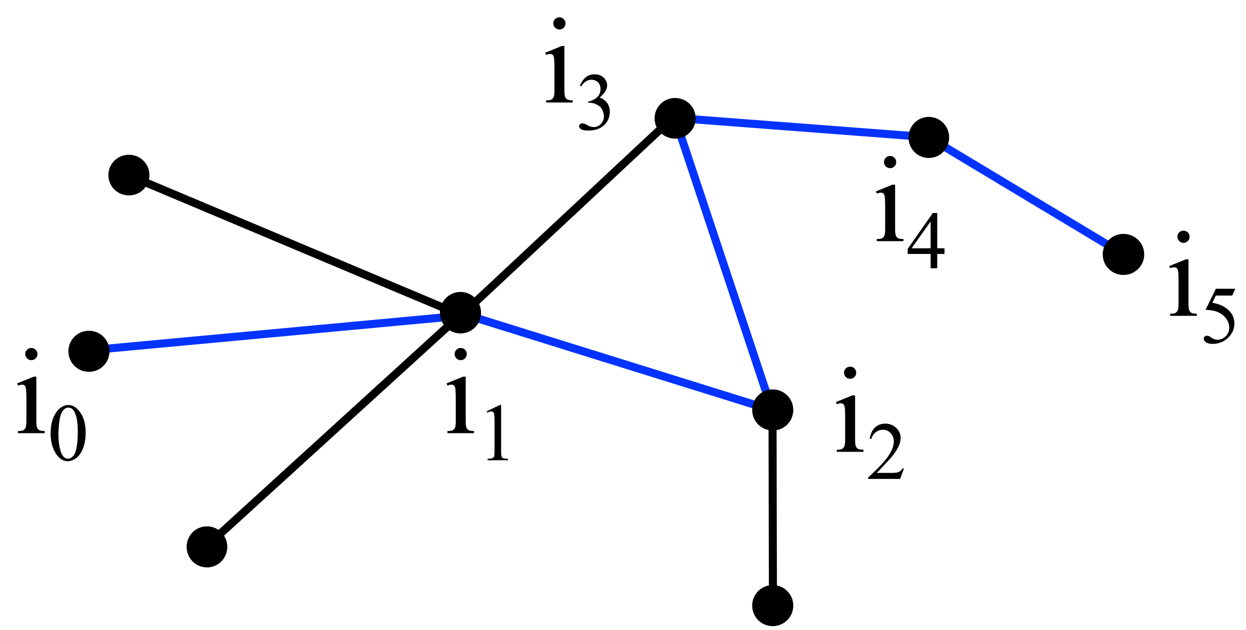
Review Holme-Saramaki, Phys. Rep. (2012), arXiv:1108.1780

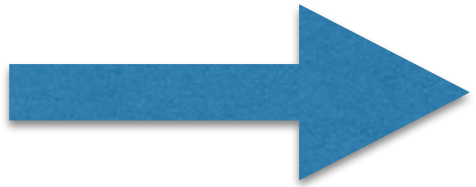
Paths in static networks

G=(V,E)

Path of length n = ordered collection of

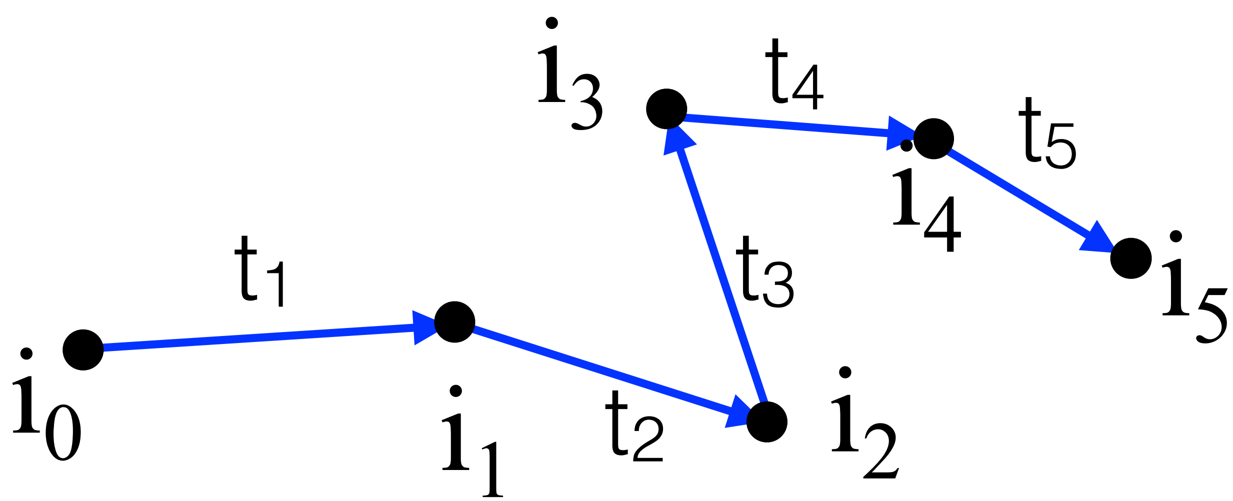
- n+1 vertices $i_0,i_1,\ldots,i_n \in V$
- n edges $(i_0,i_1), (i_1,i_2)\ldots,(i_{n-1},i_n) \in E$





Notions of shortest path, of connectedness

Time-respecting paths in temporal networks



Sequence of ***events***

Path = $\{(i_0,i_1,t_1),(i_1,i_2,t_2),\ldots,(i_{n-1},i_n,t_n) \mid t_1 < t_2 < \ldots < t_n)\}$

Length of path: n
Duration of path: $t_n - t_1$

Notions of ***shortest*** path and of ***fastest*** path

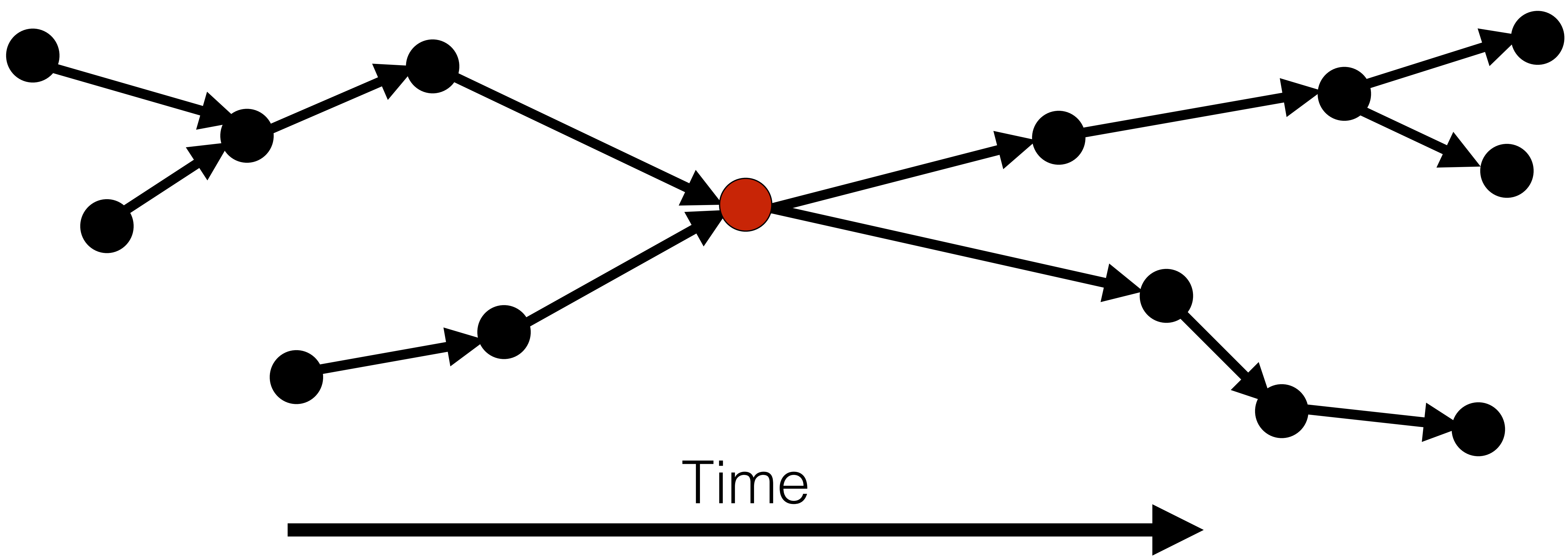
Reachability

Node i at time t

"Light cone"

Source set:
set of nodes that
can reach i at time t

Reachable set:
set of nodes that
can be reached from i
starting at time t



Time-respecting paths in temporal networks

- **Not always reciprocal**: the existence of a path from i to j does not guarantee the existence of a path from j to i

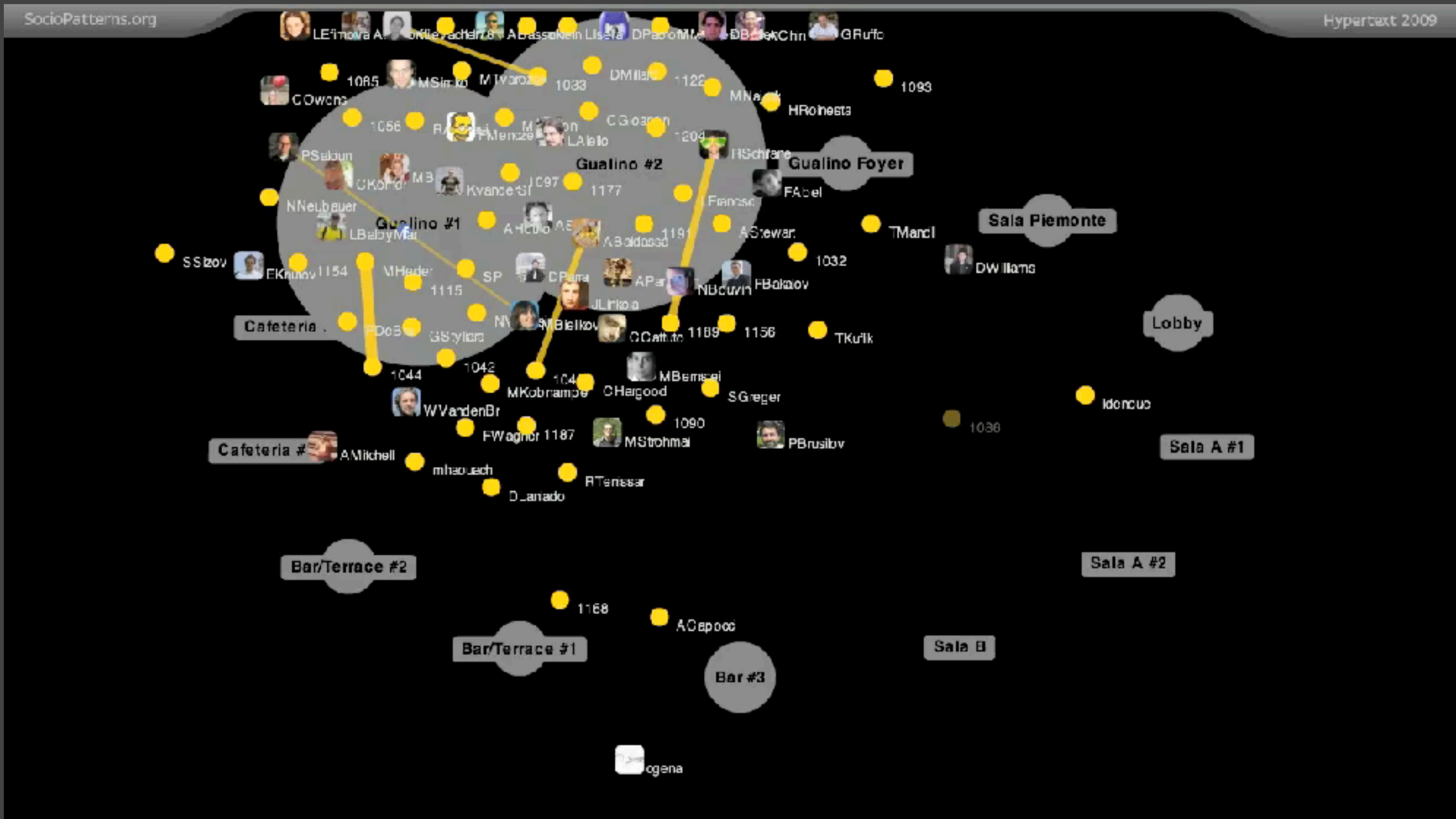
- **Not always transitive**: the existence of paths from i to j and from j to k does not guarantee the existence of a path from i (to j) to k

- **Time-dependence**:
  - there can be a path from i to j starting at t but no path starting at $t' > t$
  - shortest and fastest paths can differ
  - length of shortest path can depend on starting time
  - duration of fastest path can depend on starting time
  - there can be a path starting from i at t_0, reaching j at t_1, and another path starting form i at $t'_0 > t_0$ reaching j at $t'_1 > t_1$, with $t'_1 - t'_0 < t_1 - t_0$ (i.e., smaller duration but arriving later), and/or of shorter length (smaller number of hops)

> a glimpse of SocioPatterns data

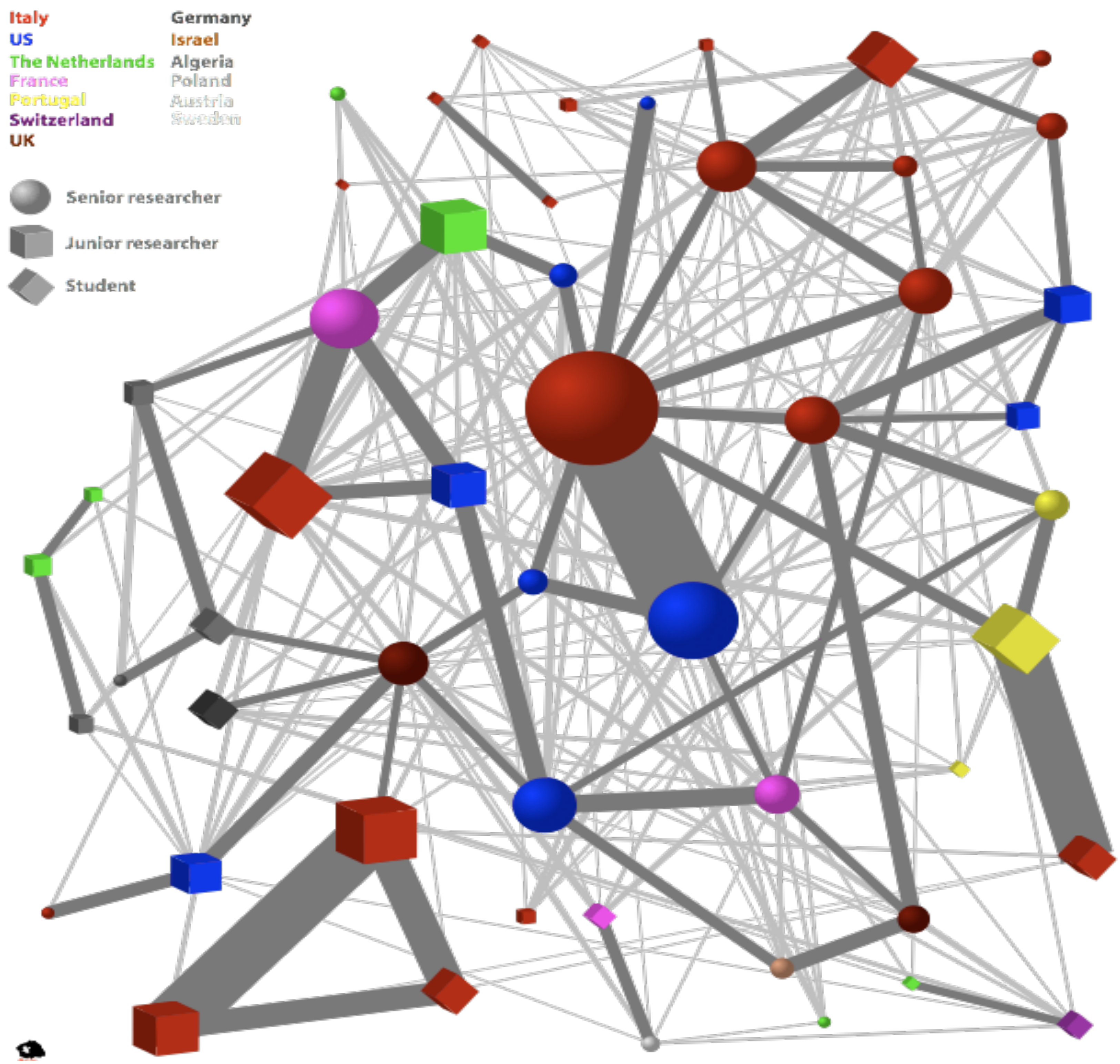
Many data sets available at
www.sociopatterns.org

conference: dynamical network of f2f proximity



http://www.vimeo.com/6590604

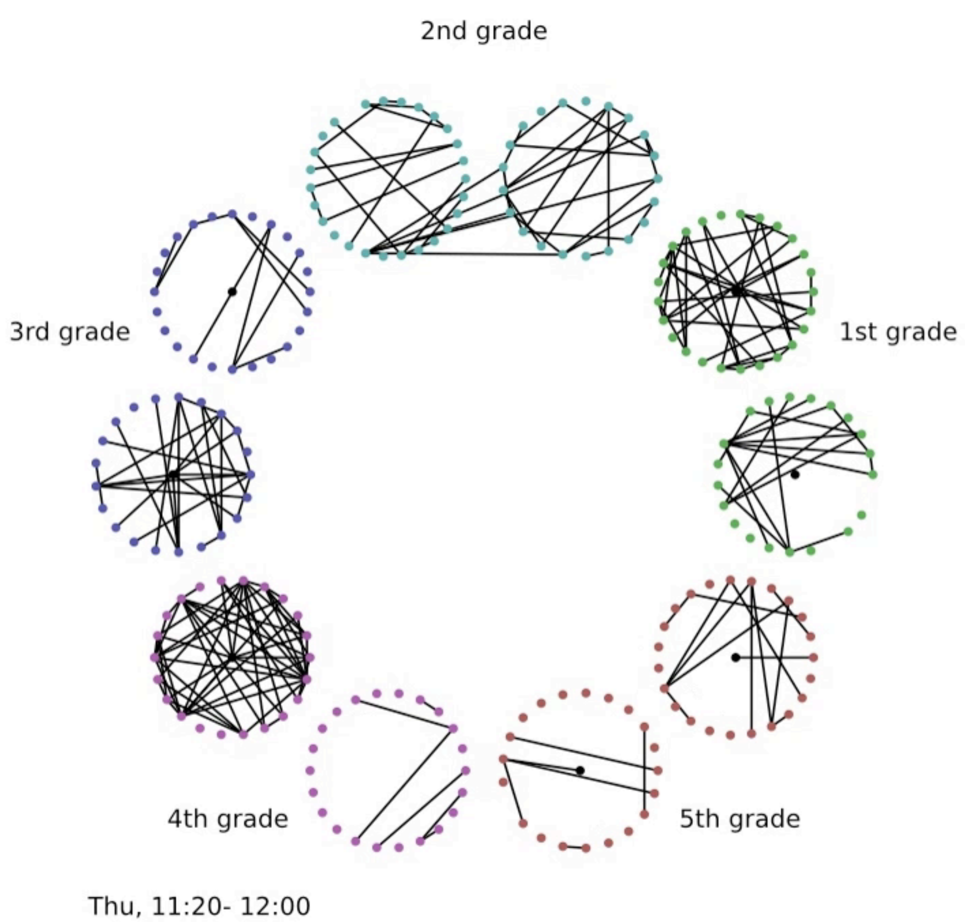
Conference contact network



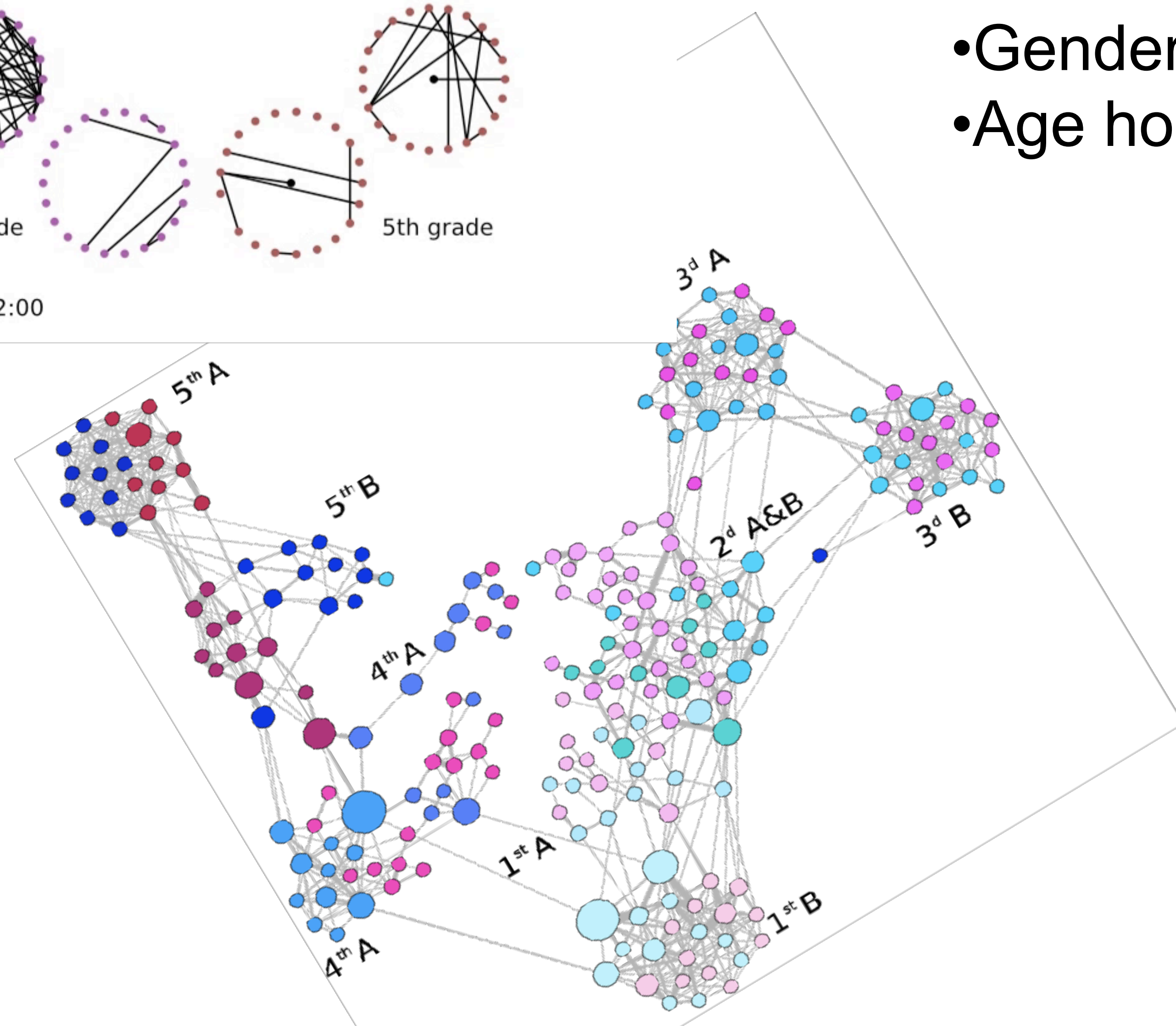
SocioPatterns

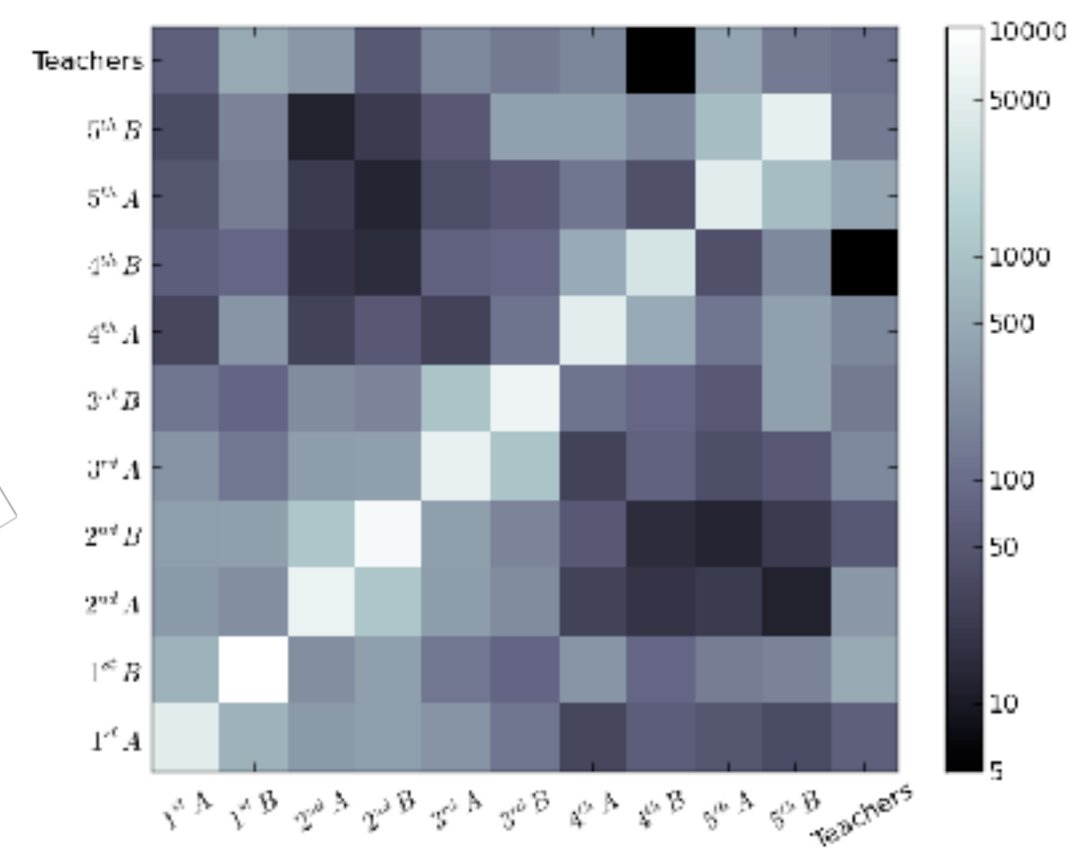
Contacts in a primary school

J. Stehlé et al.
PLoS ONE 6(8):e23176 (2011)
Soc. Networks 35:604 (2013)

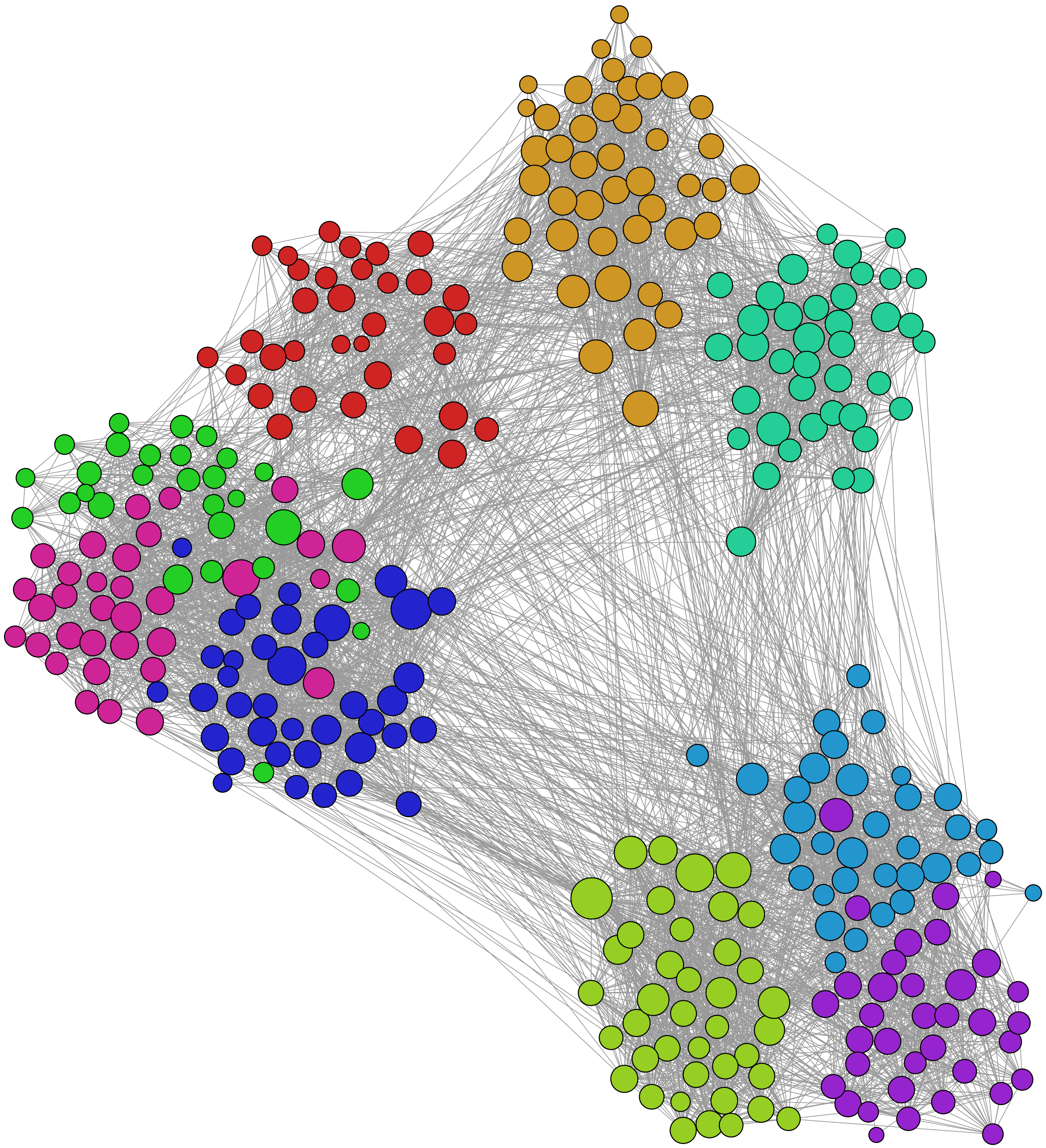


- Epidemiology:
  - Information of models
  - Design and efficiency of containment measures
- Social sciences:
  - Gender segregation
  - Age homophily



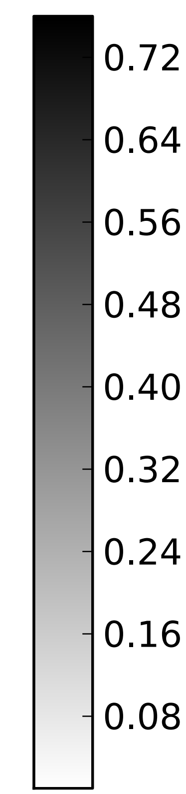


Contacts in a high-school



Class structure

	2BIO1	2BIO2	2BIO3	MP	MP*1	MP*2	PC	PC*	PSI*
2BIO1	0.61	0.11	0.09	0.01	0.01	0.02	0.01	0.01	0.01
2BIO2	0.11	0.64	0.12	0.02	0.01	0.02	0.01	0.03	0.02
2BIO3	0.09	0.12	0.76	0.03	0.01	0.03	0.02	0.02	0.01
MP	0.01	0.02	0.03	0.68	0.09	0.1	0.02	0.02	0.03
MP*1	0.01	0.01	0.01	0.09	0.52	0.12	0.02	0.01	0.03
MP*2	0.02	0.02	0.03	0.1	0.12	0.69	0.04	0.01	0.05
PC	0.01	0.01	0.02	0.02	0.02	0.04	0.72	0.09	0.04
PC*	0.01	0.03	0.02	0.02	0.01	0.01	0.09	0.65	0.04
PSI*	0.01	0.02	0.01	0.03	0.03	0.05	0.04	0.04	0.61

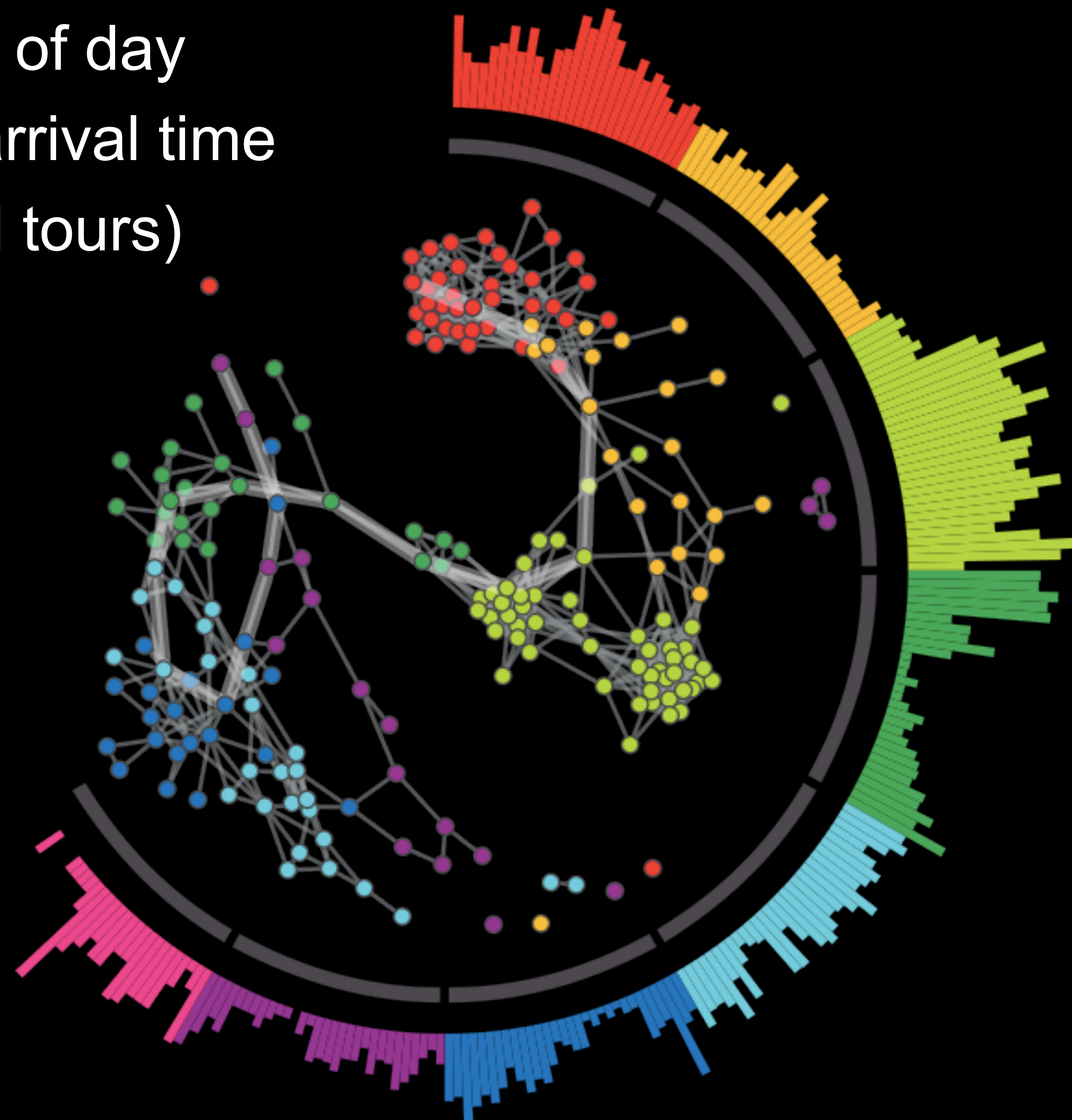


J. Fournet, A. Barrat, PLoS ONE (2014)

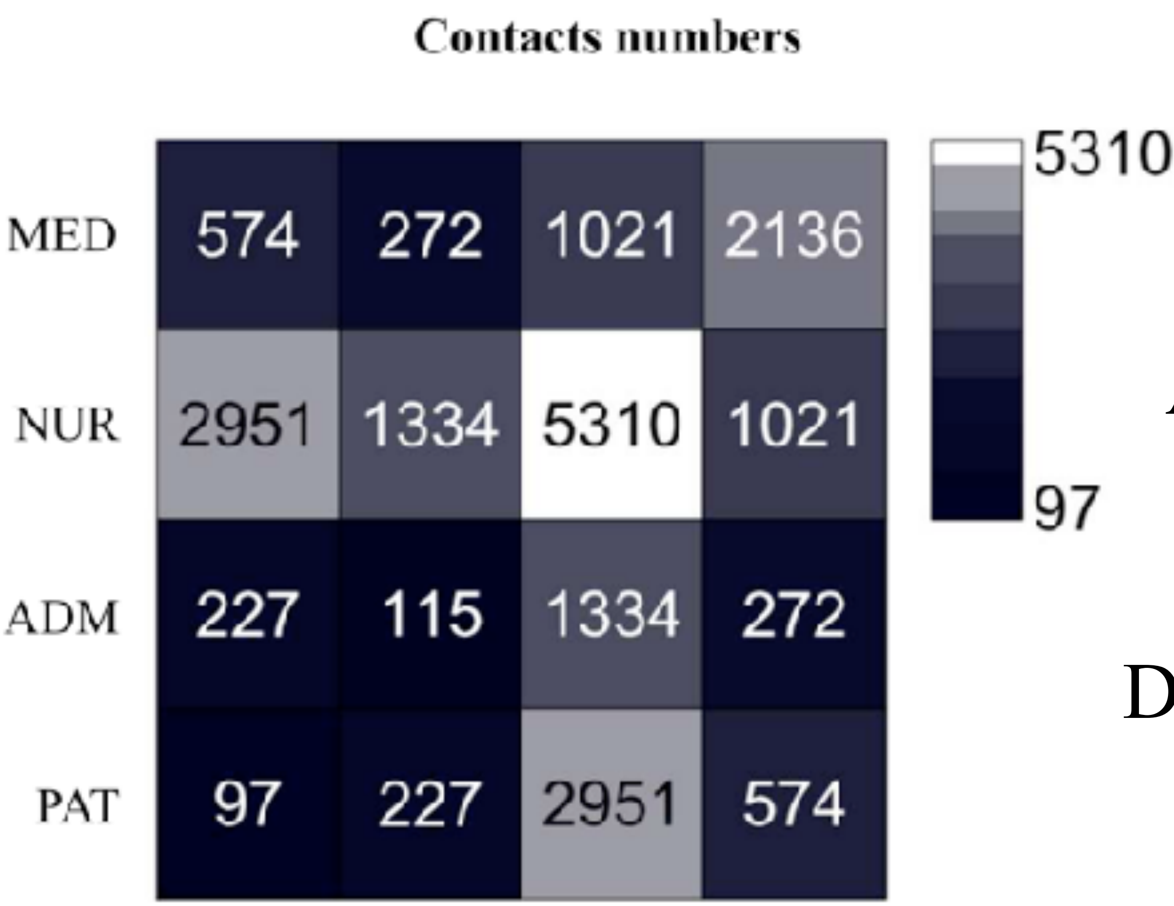
R. Mastrandrea, J. Fournet, A. Barrat, PLoS ONE (2015)

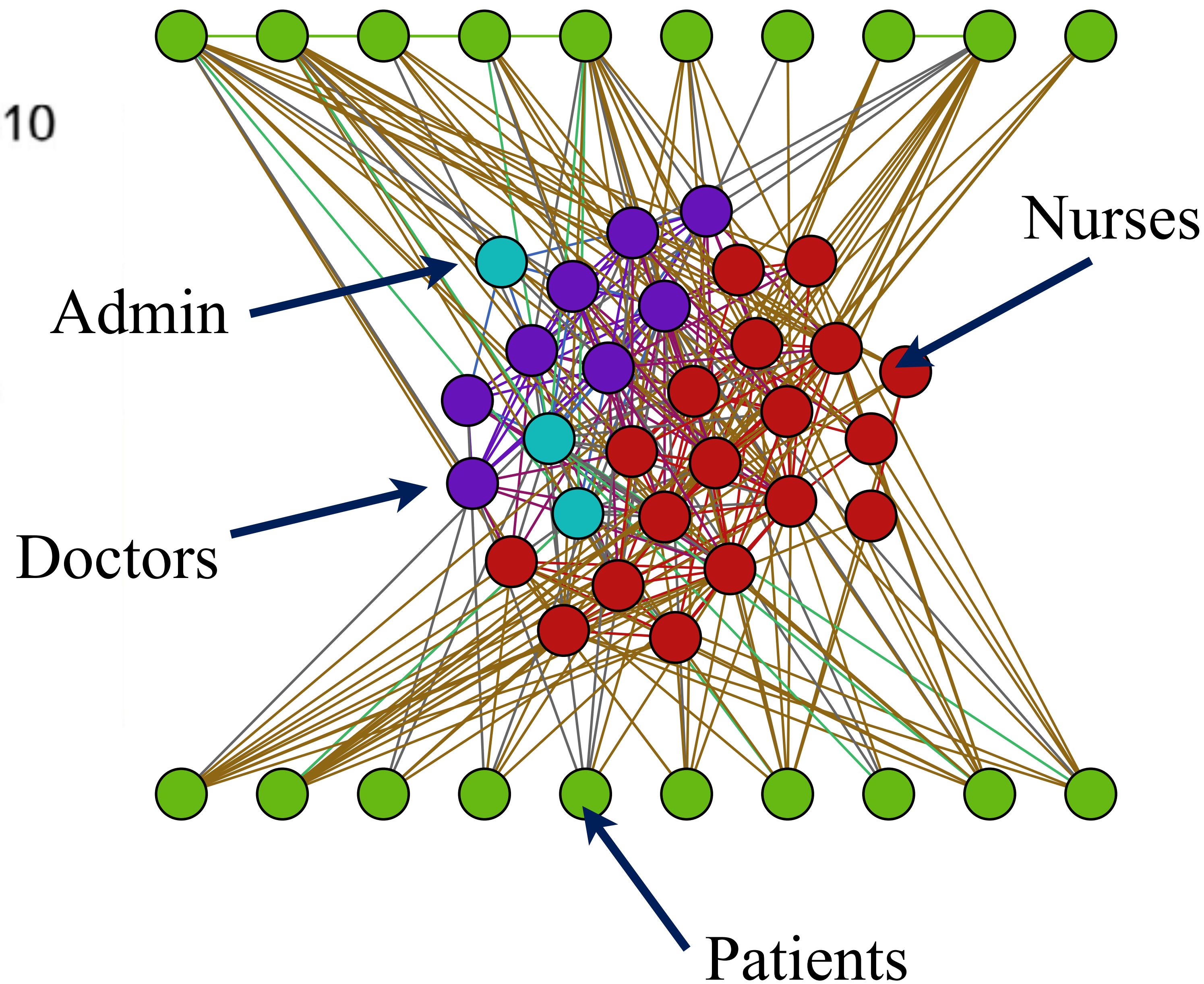
Museum

- color encodes the time of day
- nodes are colored by arrival time
- several groups (guided tours)



Lyon hospital, geriatric ward



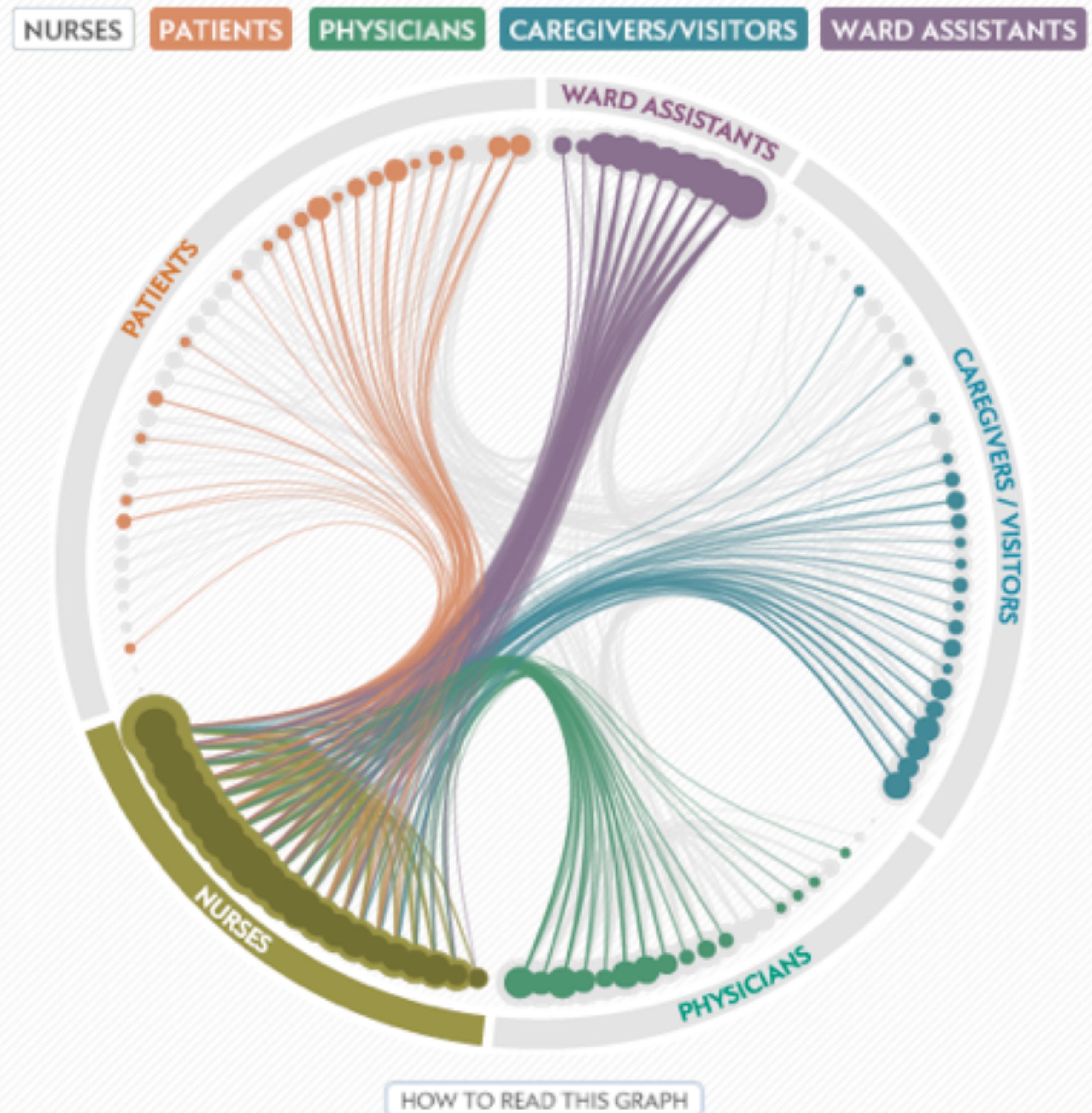


Most contacts involve nurses

P. Vanhems et al., PLOS ONE (2013)

Contacts in a hospital ward (Rome)

Click the labels on the right to filter the data displayed



Hospitals shouldn't make you sicker. But plenty of people acquire illnesses while hospitalized—in some countries, such so-called nosocomial infections afflict more than 10 percent of patients.

To investigate transmission pathways, European researchers of the SocioPatterns collaboration fitted 119 people in a ward of the Bambino Gesù Children's Hospital with radio-frequency identification (RFID) badges. The tags registered face-to-face interactions—and the potential spreading of airborne pathogens.

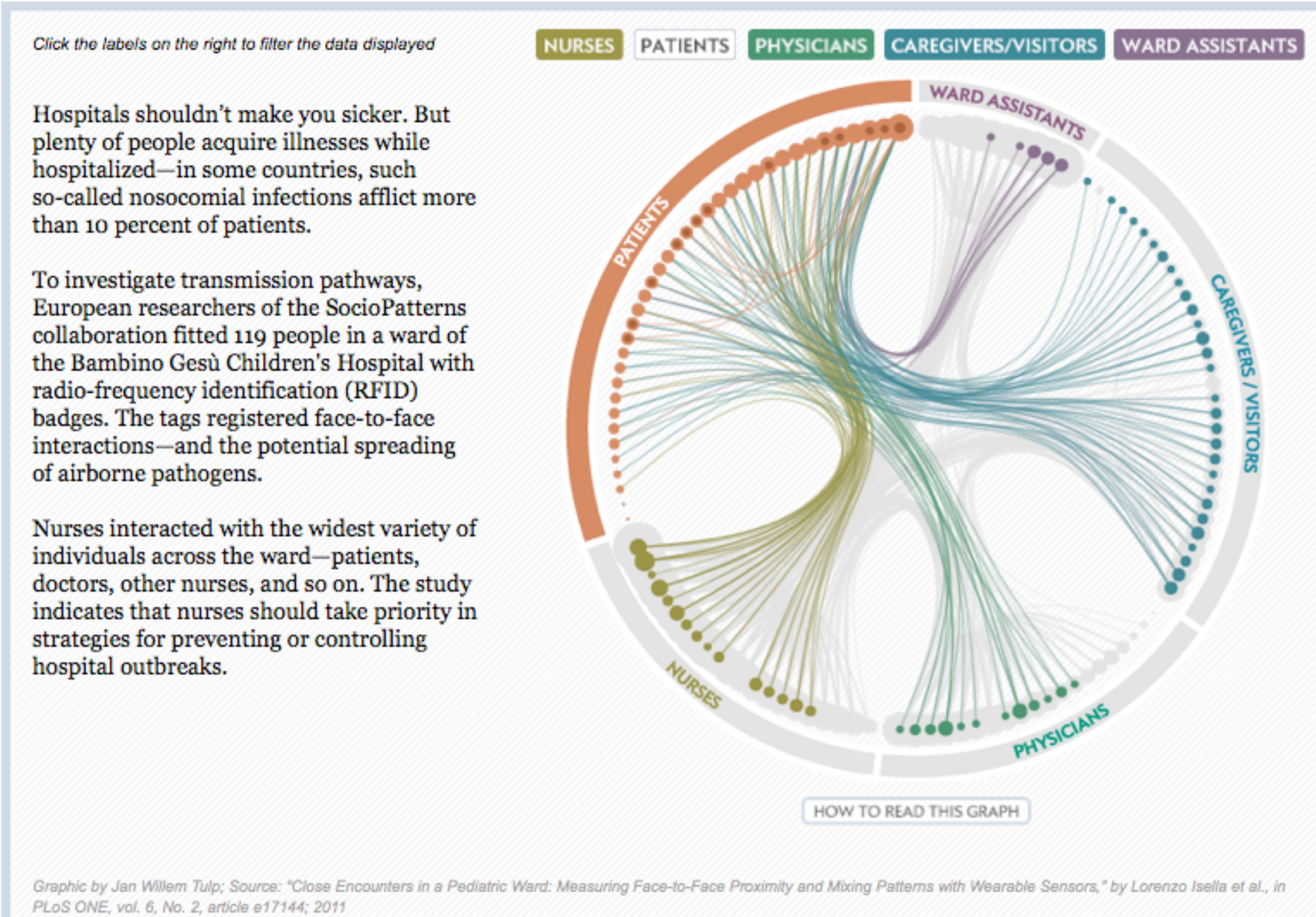
Nurses interacted with the widest variety of individuals across the ward—patients, doctors, other nurses, and so on. The study indicates that nurses should take priority in strategies for preventing or controlling hospital outbreaks.

Graphic by Jan Willem Tulp; Source: "Close Encounters in a Pediatric Ward: Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors," by Lorenzo Isella et al., in PLoS ONE, vol. 6, No. 2, article e17144; 2011

Scientific American Magazine, *Graphic Science*, 12 November 2012

L. Isella et al. PLoS ONE 6(2):e17144 (2011)

Contacts in a hospital ward (Rome)



Hospitals shouldn't make you sicker. But plenty of people acquire illnesses while hospitalized—in some countries, such so-called nosocomial infections afflict more than 10 percent of patients.

To investigate transmission pathways, European researchers of the SocioPatterns collaboration fitted 119 people in a ward of the Bambino Gesù Children's Hospital with radio-frequency identification (RFID) badges. The tags registered face-to-face interactions—and the potential spreading of airborne pathogens.

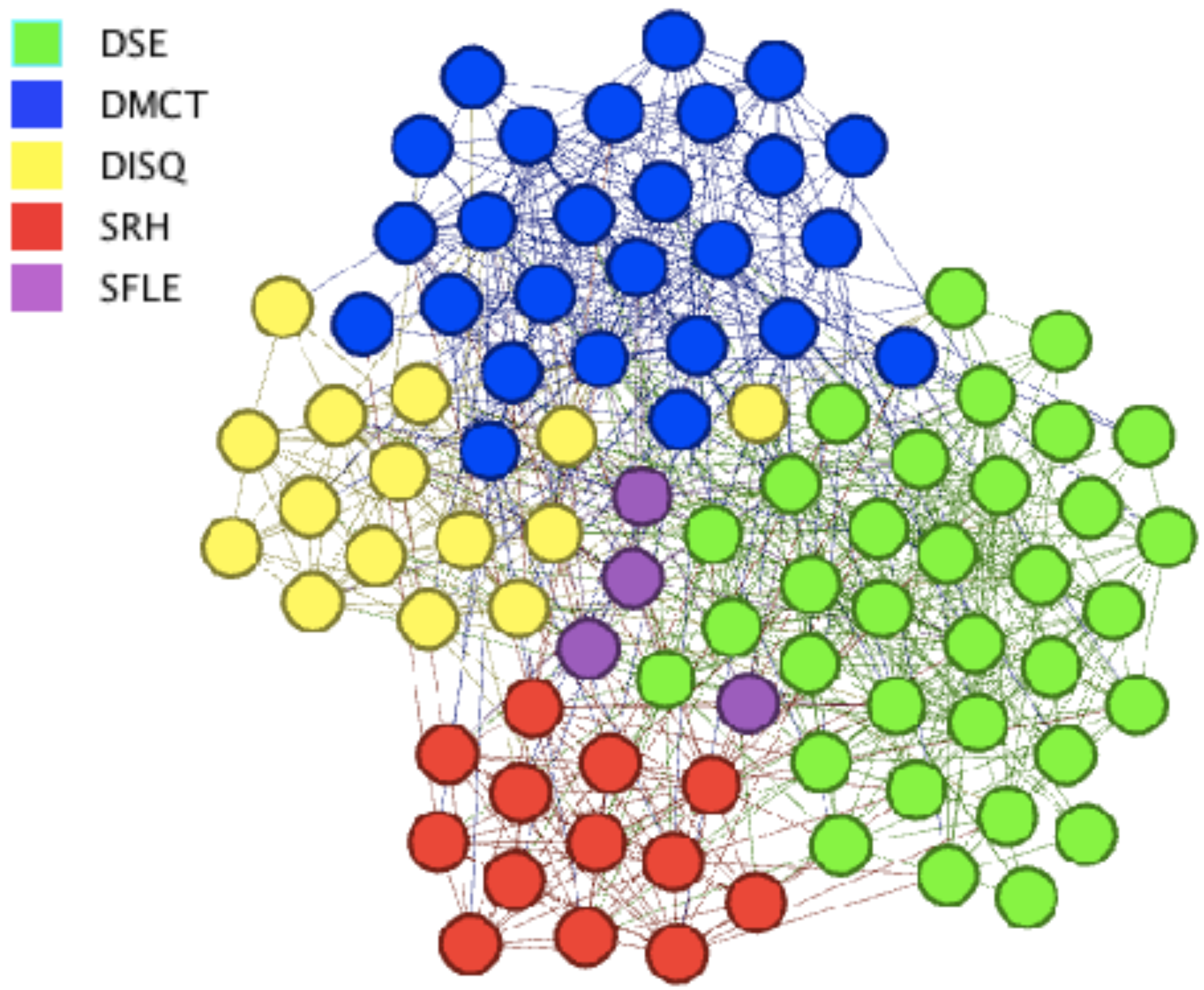
Nurses interacted with the widest variety of individuals across the ward—patients, doctors, other nurses, and so on. The study indicates that nurses should take priority in strategies for preventing or controlling hospital outbreaks.

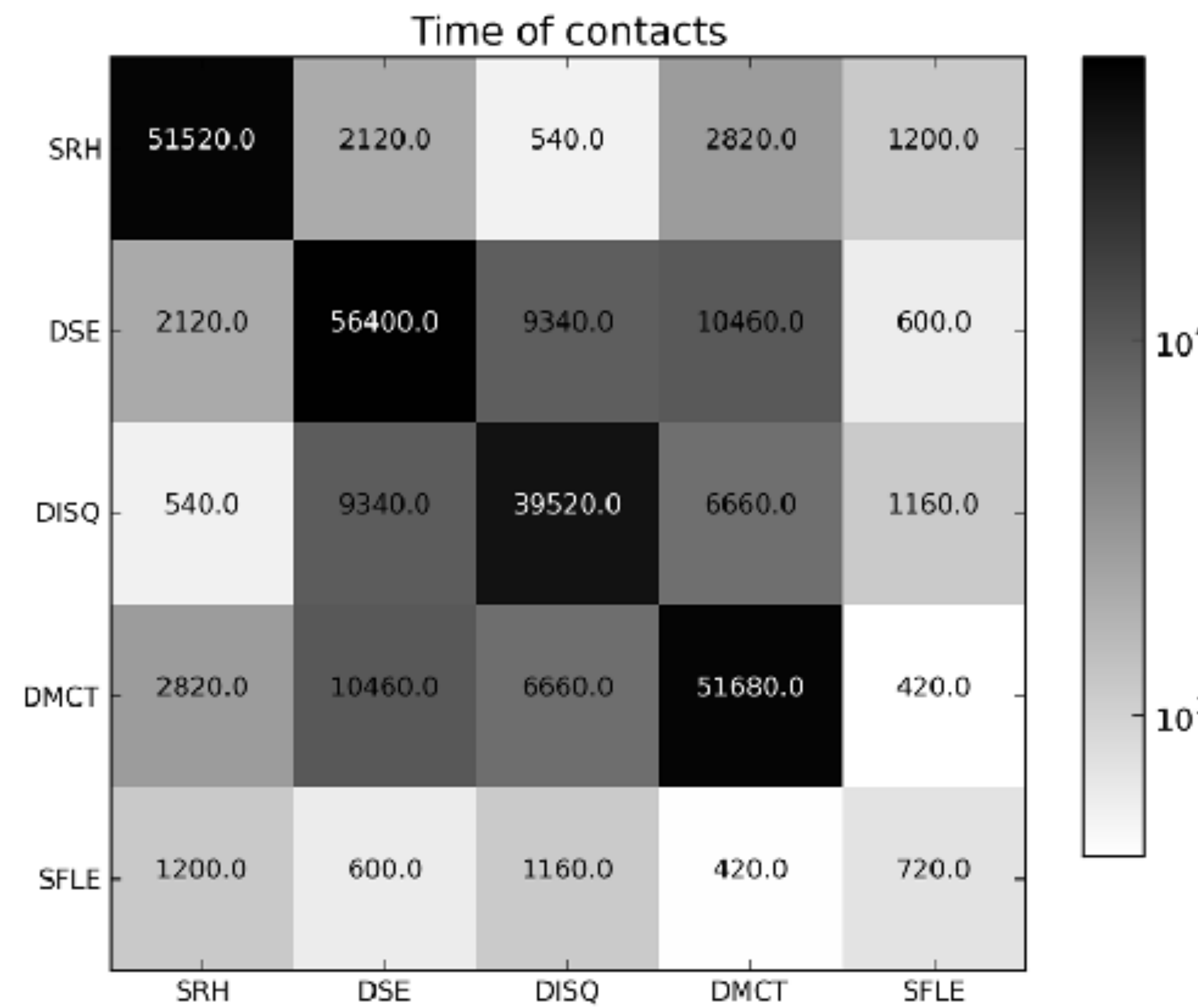
Graphic by Jan Willem Tulp; Source: "Close Encounters in a Pediatric Ward: Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors," by Lorenzo Isella et al., in PLoS ONE, vol. 6, No. 2, article e17144; 2011

Scientific American Magazine, *Graphic Science*, 12 November 2012

L. Isella et al. PLoS ONE 6(2):e17144 (2011)

Offices, two weeks aggregated data





Non-homogeneous mixing
Strong homophily by department, except for SFLE (logistics)

M. Génois et al., Network Science (2015)

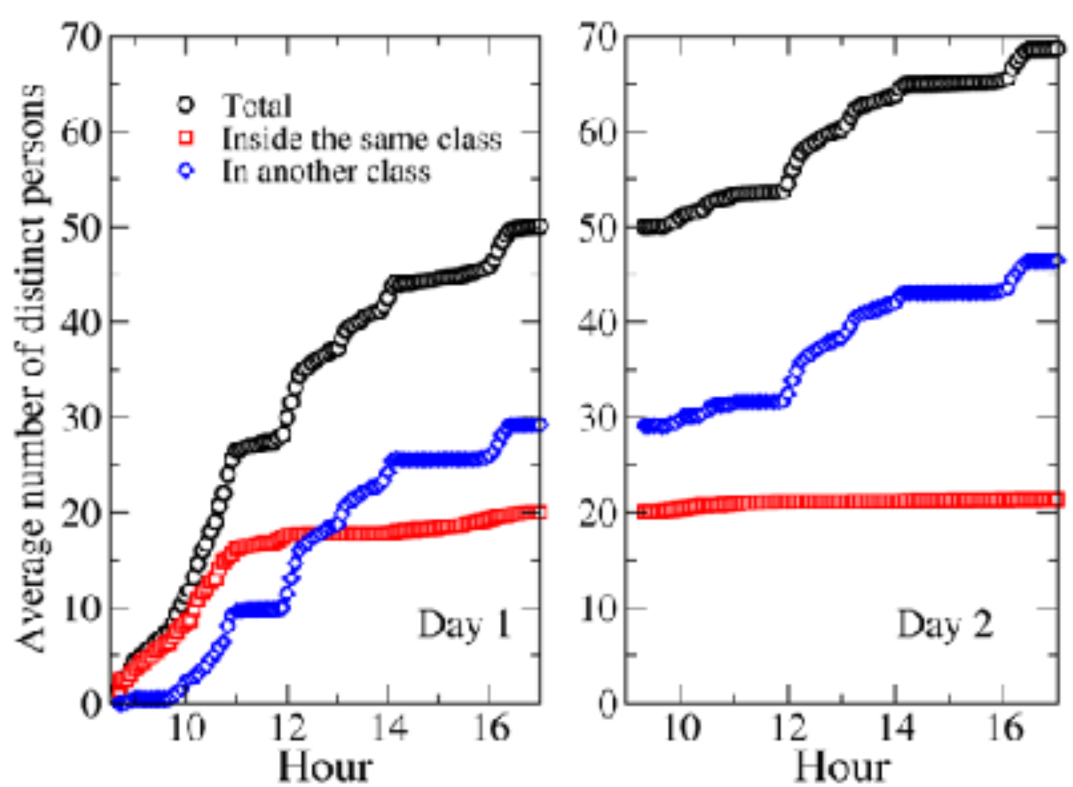
> looking at and comparing datasets

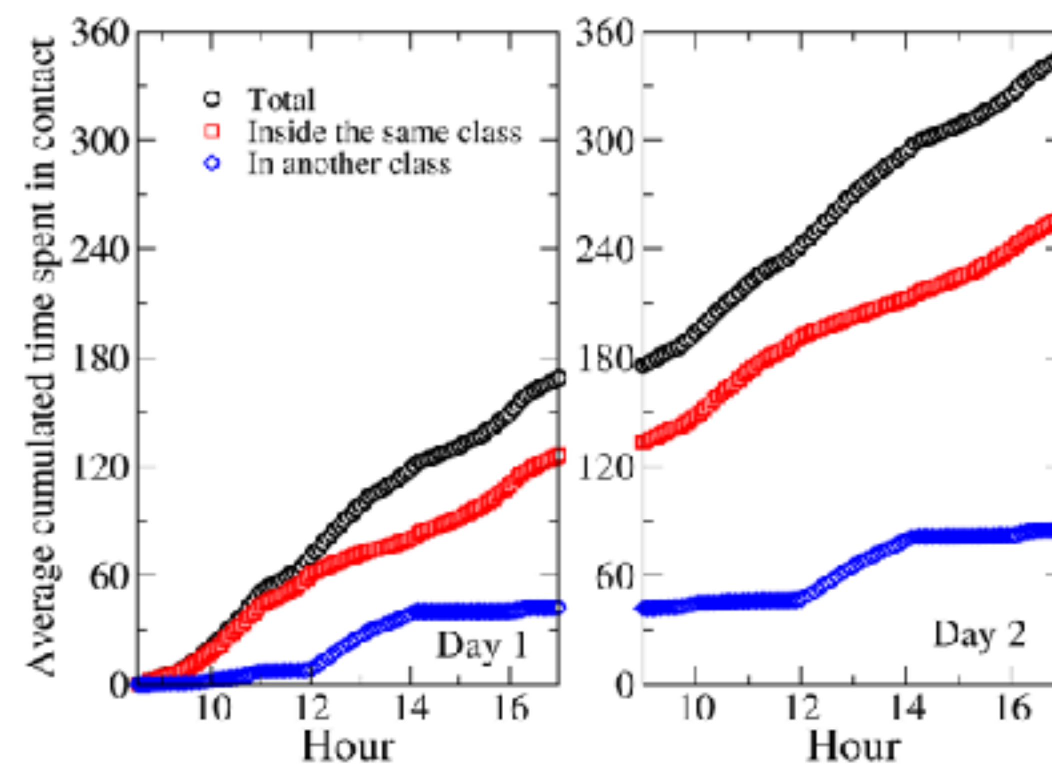
Static networks

- Degree distribution P(k)
- Clustering coefficient (of nodes of degree k)
- Average degree of nearest neighbours k_{nn}
- Motifs
- Communities
- Distribution of link weights P(w)
- Distribution of node strengths P(s)
- ….

Temporal networks

- Properties of networks aggregated on different time windows: P(k), P(w), P(s), etc…
- Evolution of averaged properties when time window length increases (e.g., <k>(t), <s>(t))





J. Stehlé et al. PLoS ONE 6(8):e23176 (2011)

Temporal networks

Temporal statistical properties

- distribution of contact numbers $P(n)$,
- distribution of contact durations $P(\tau)$,
- distribution of inter-contact times $P(\Delta t)$

for nodes and links

Stationarity of distributions
(Non-)stationarity of activity

Burstiness

Different contexts

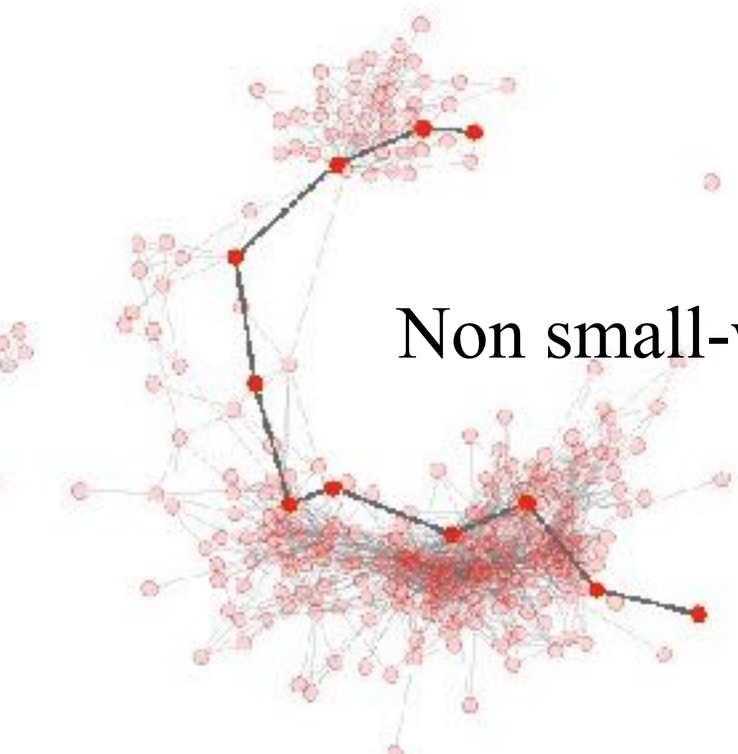
- Conferences
- Museum
- School
- Offices
- Hospital wards
- …

Similarities/differences in the f2f proximity patterns?

L. Isella *et al.*, Journal of Theoretical Biology 271, 166 (2011)

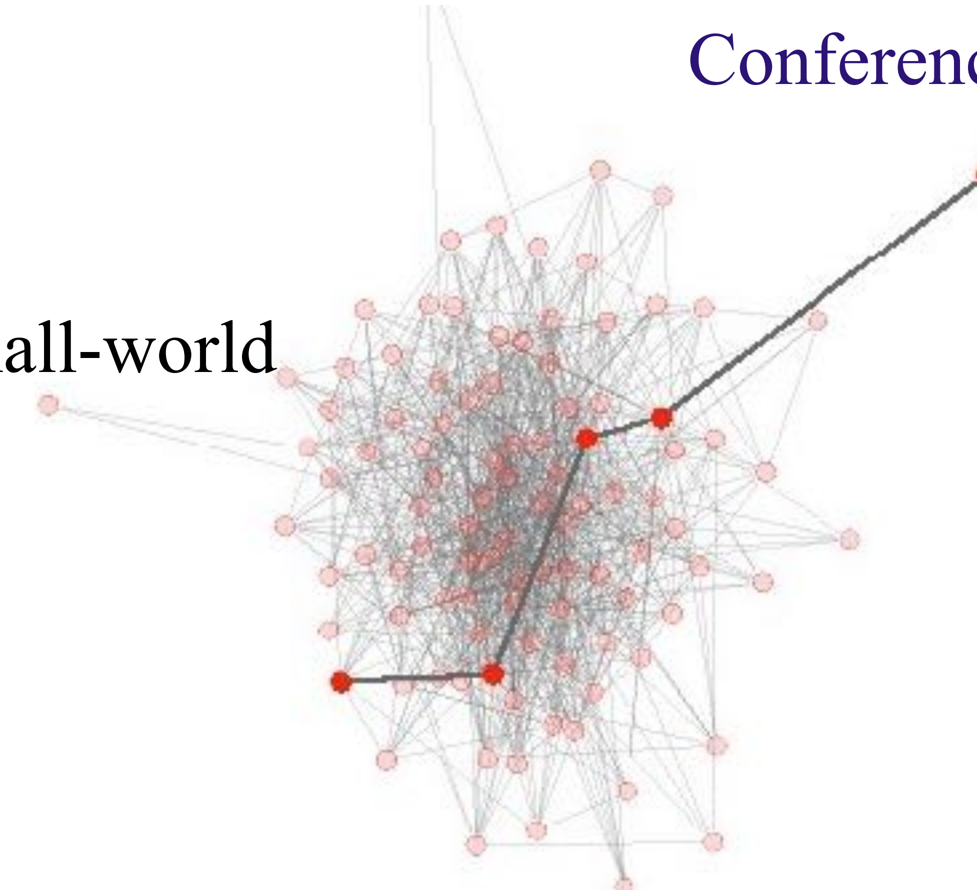
Different structures of aggregated networks

Museum



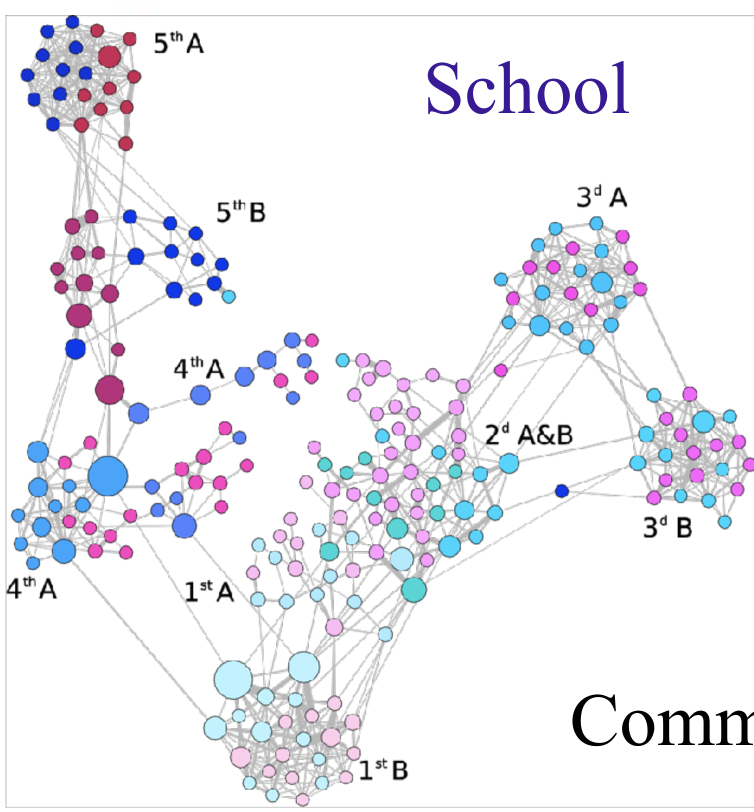
Non small-world

Conference



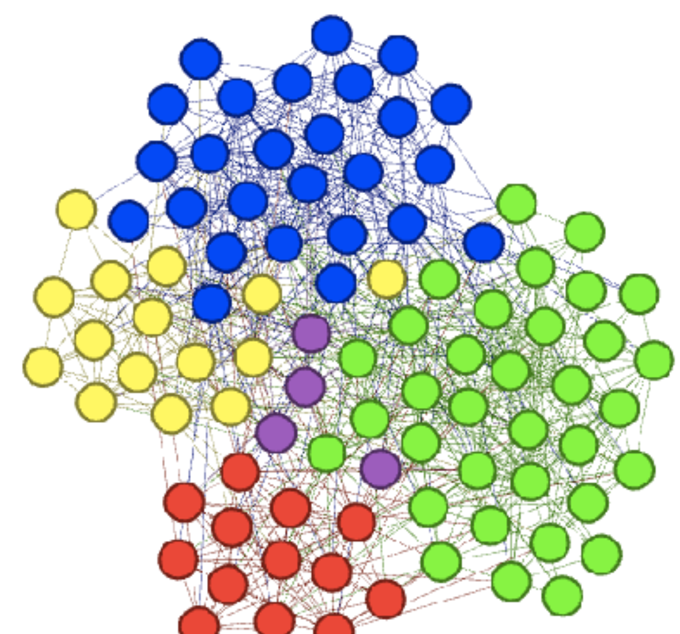
Small-world

School



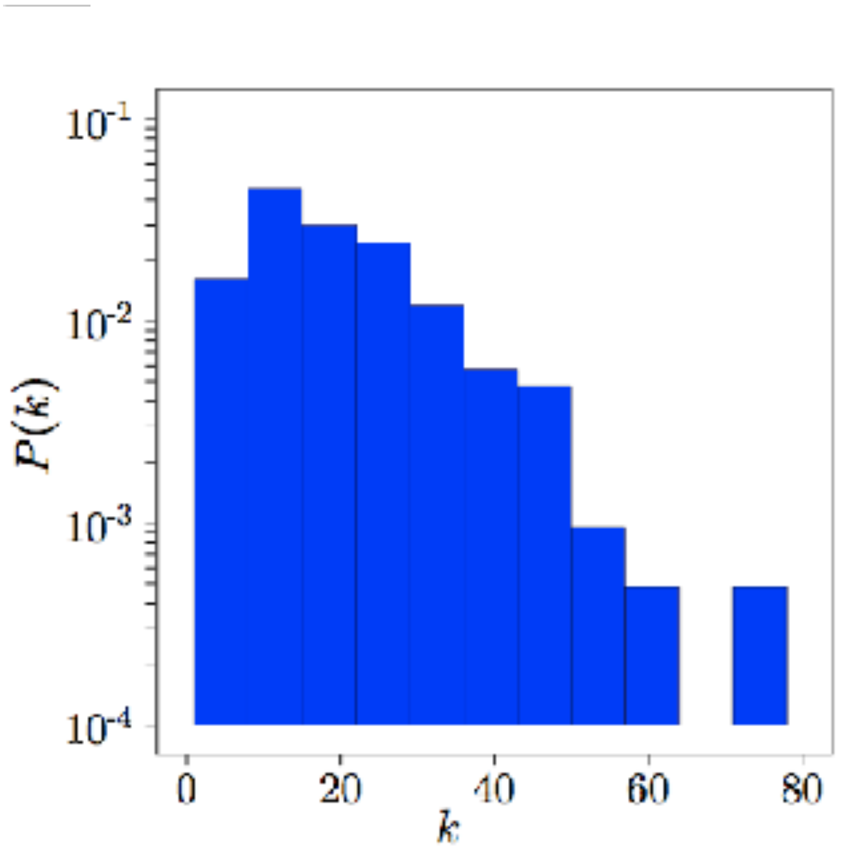
Communities

Offices



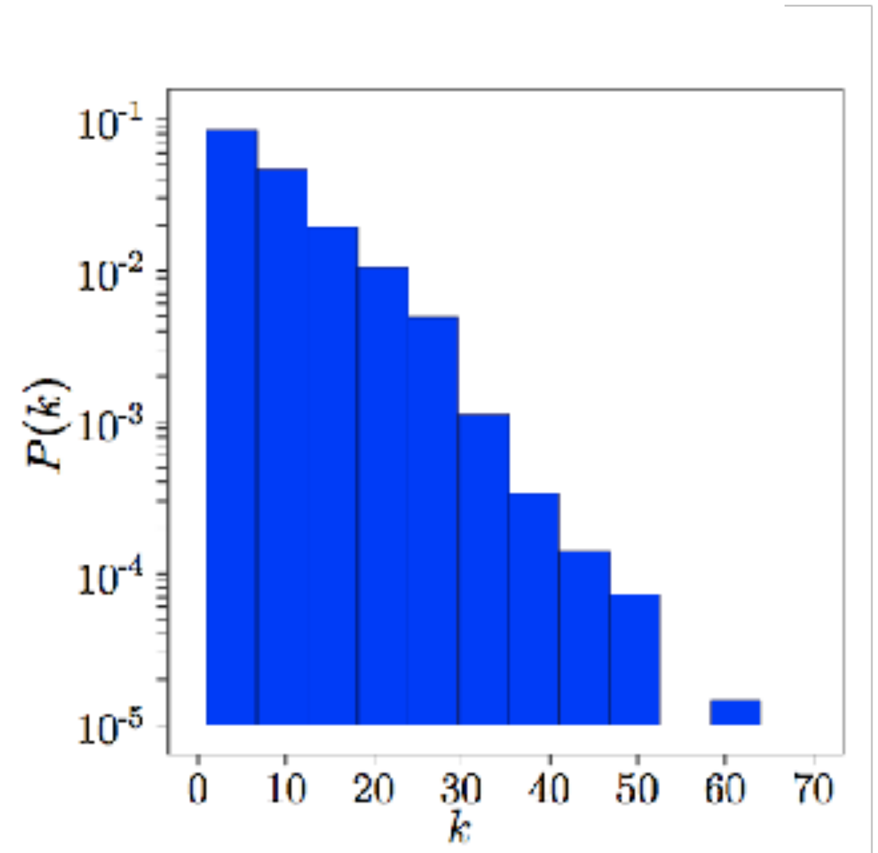
Degree distributions

Conference



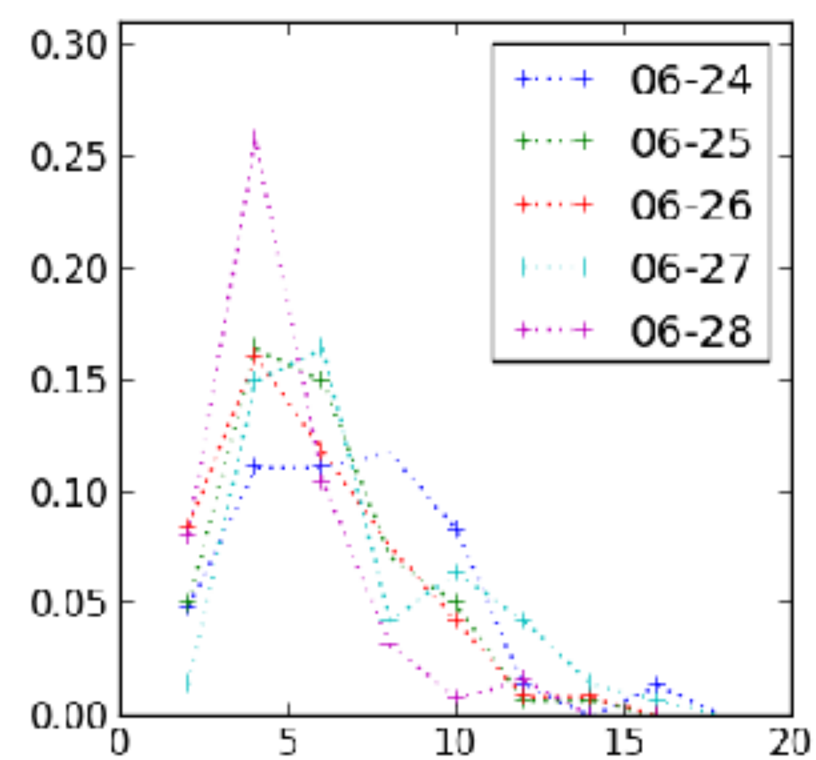
$$< k > \simeq 20$$

Museum



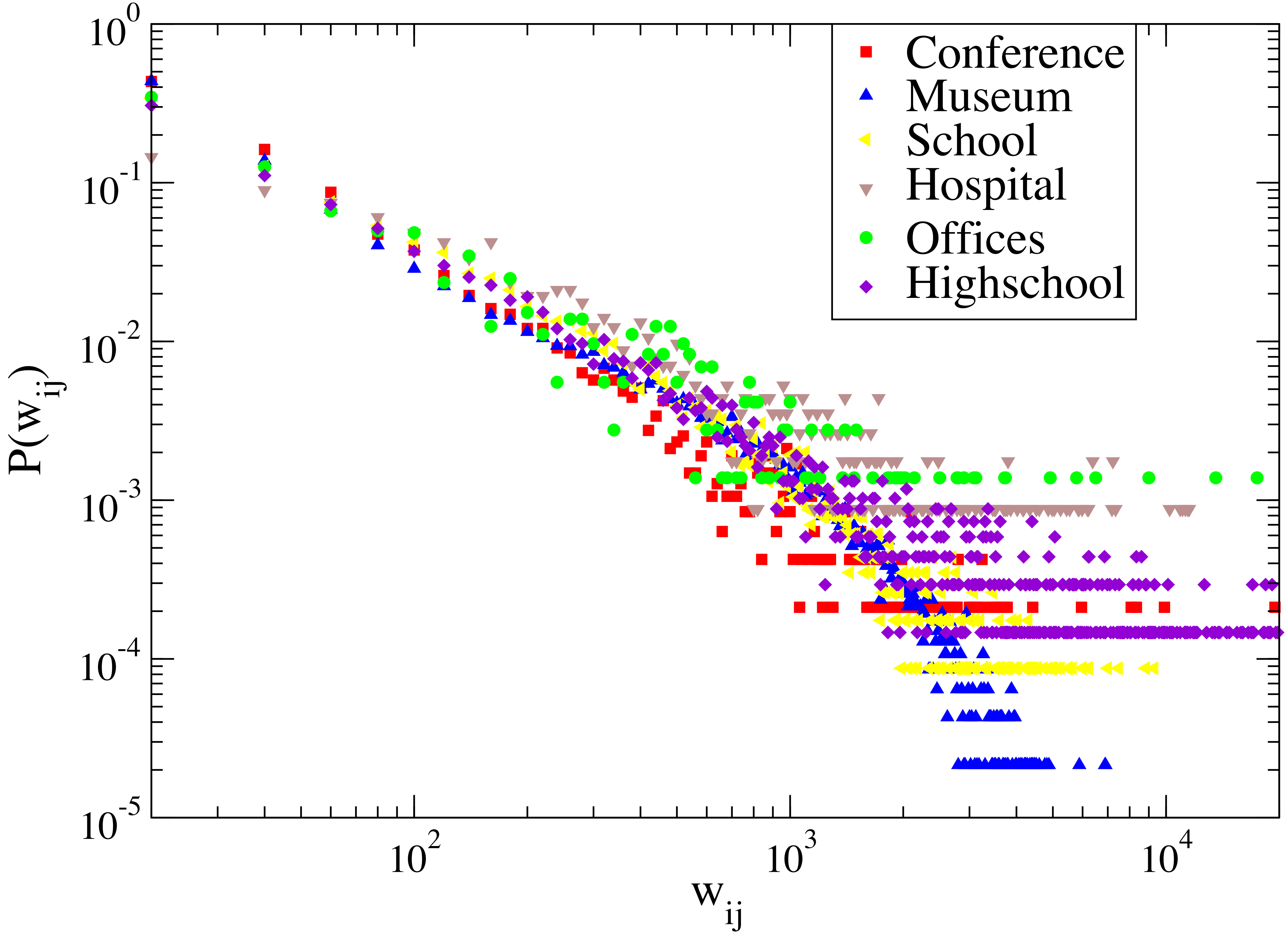
$$< k > \simeq 8$$

Offices

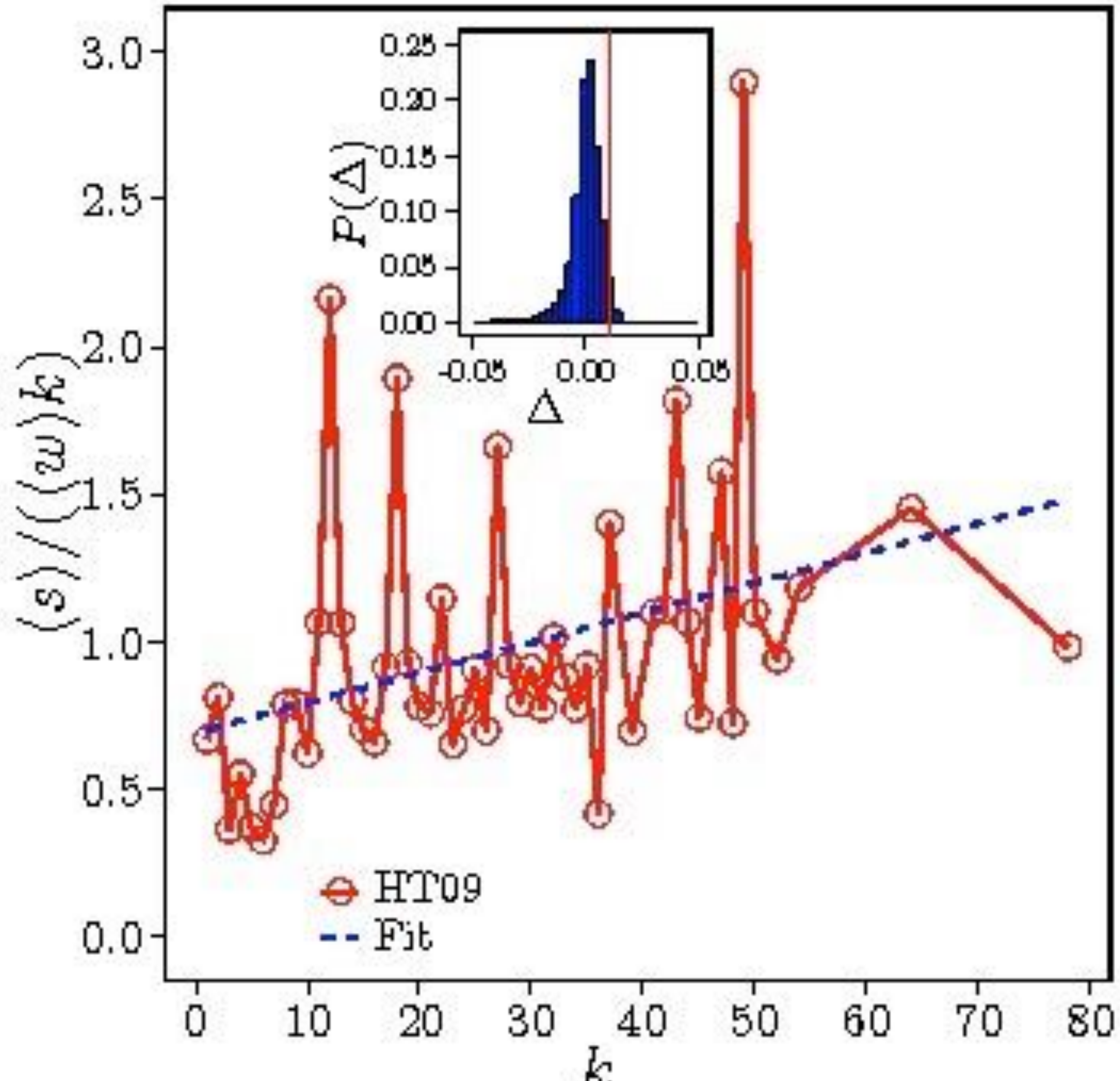


Similar degree distributions

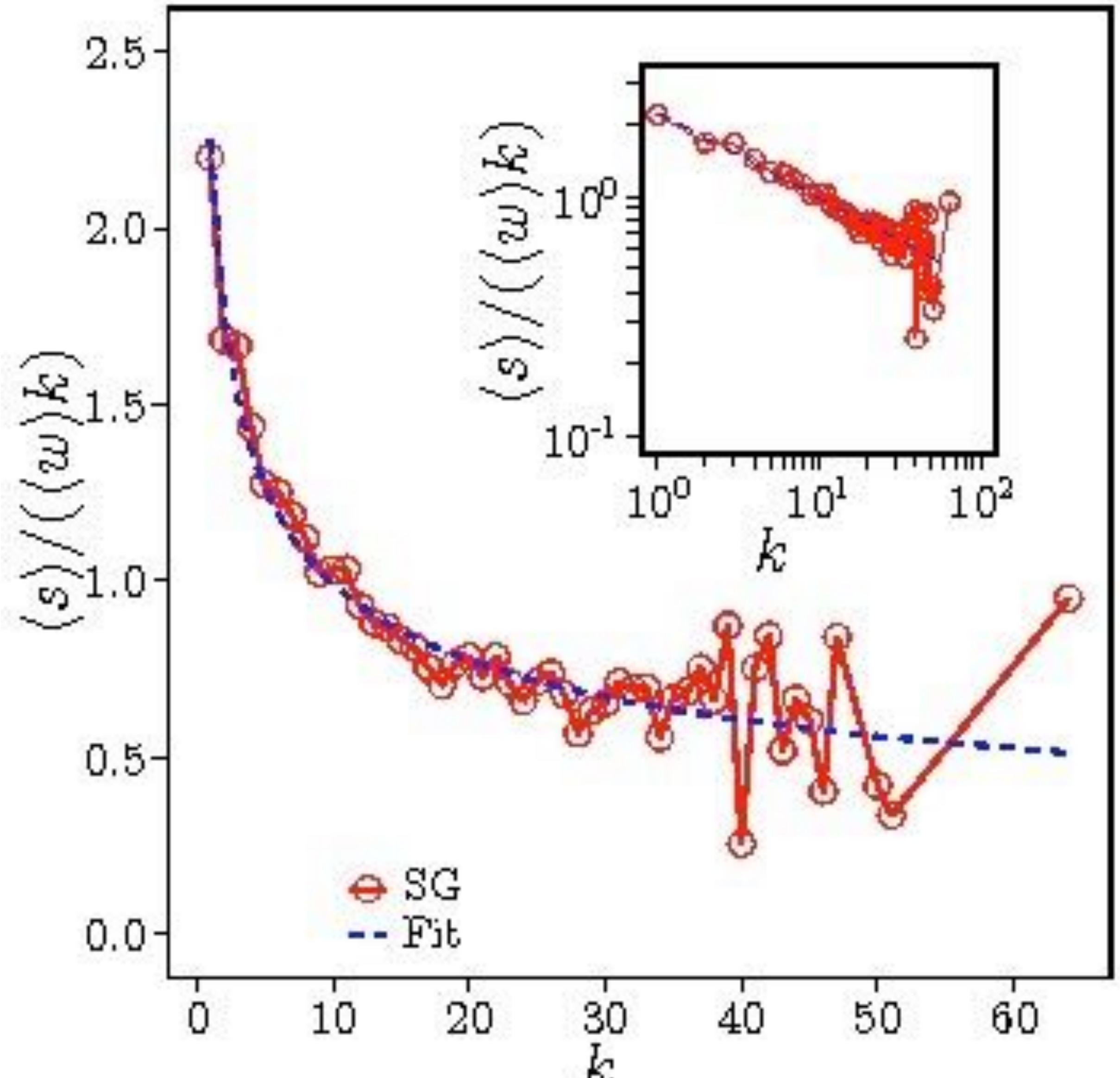
Daily cumulated contact times (=link weights)



Different “supercontacting” patterns
(topology-weights correlations)



Superspreading

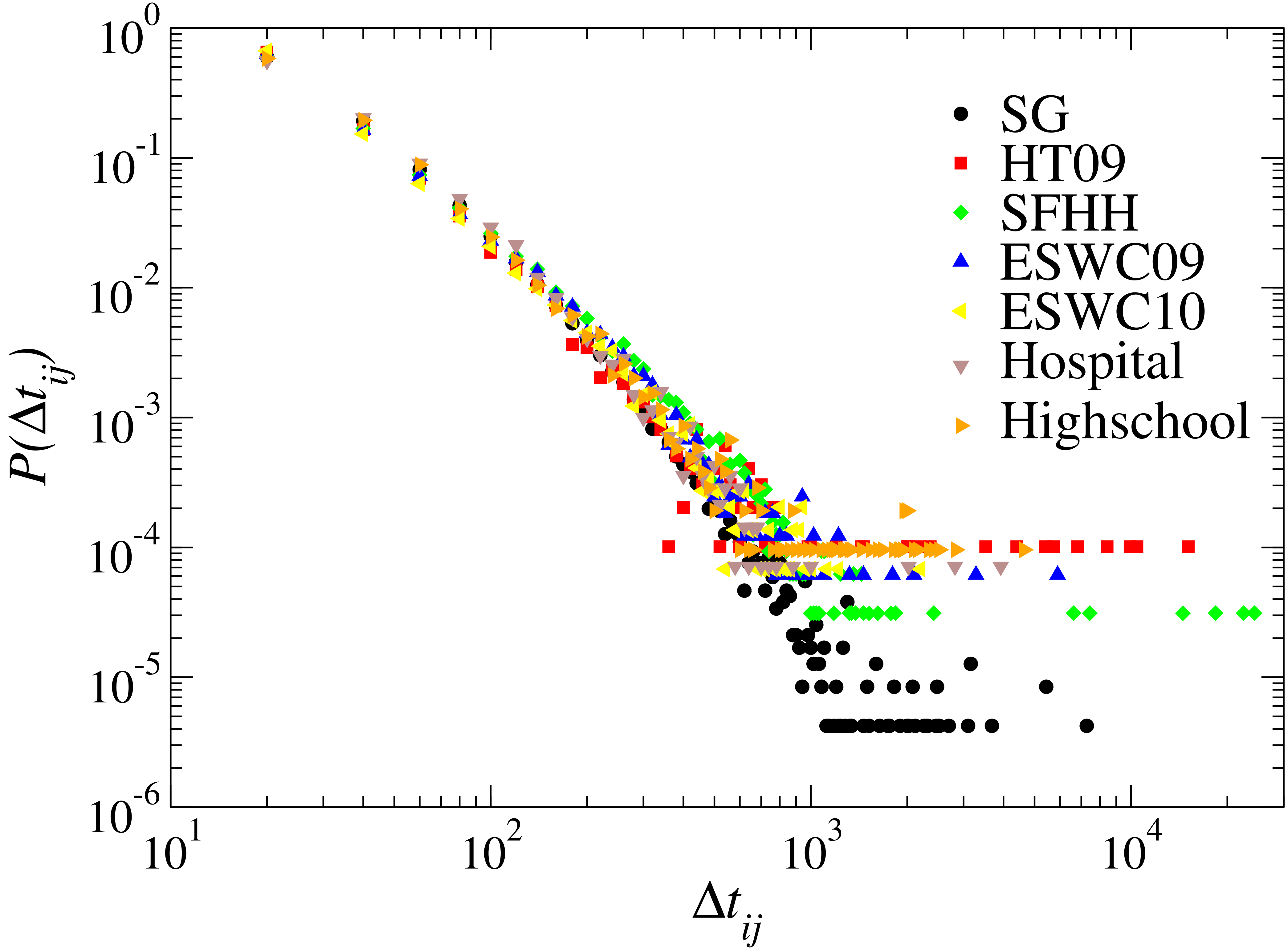


Opposite trend

k = number of distinct persons contacted
s = total time spent in contact
Random weights: s ~ <w>k

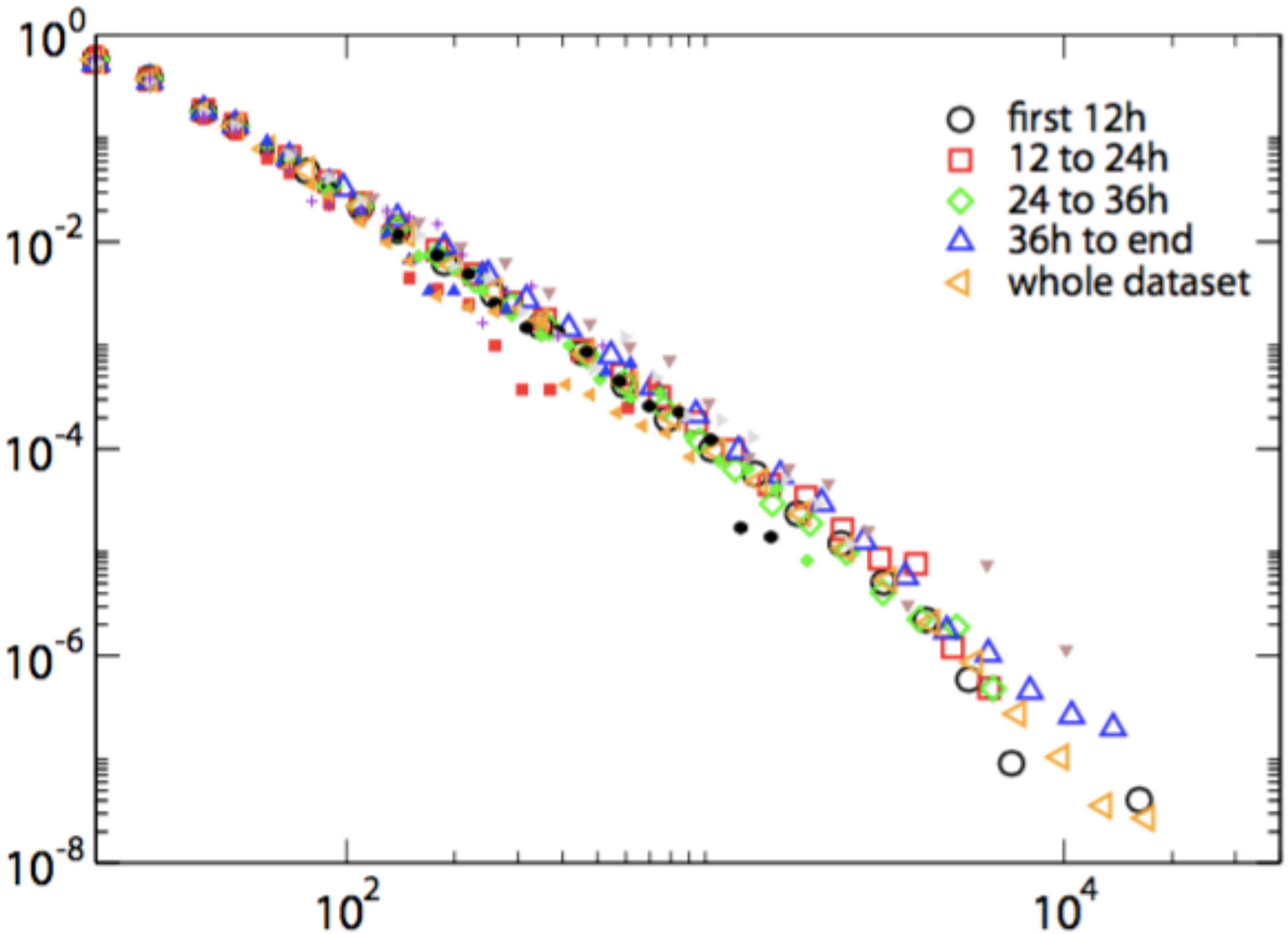
L. Isella *et al.*, Journal of Theoretical Biology 271, 166 (2011)

Contact duration distributions

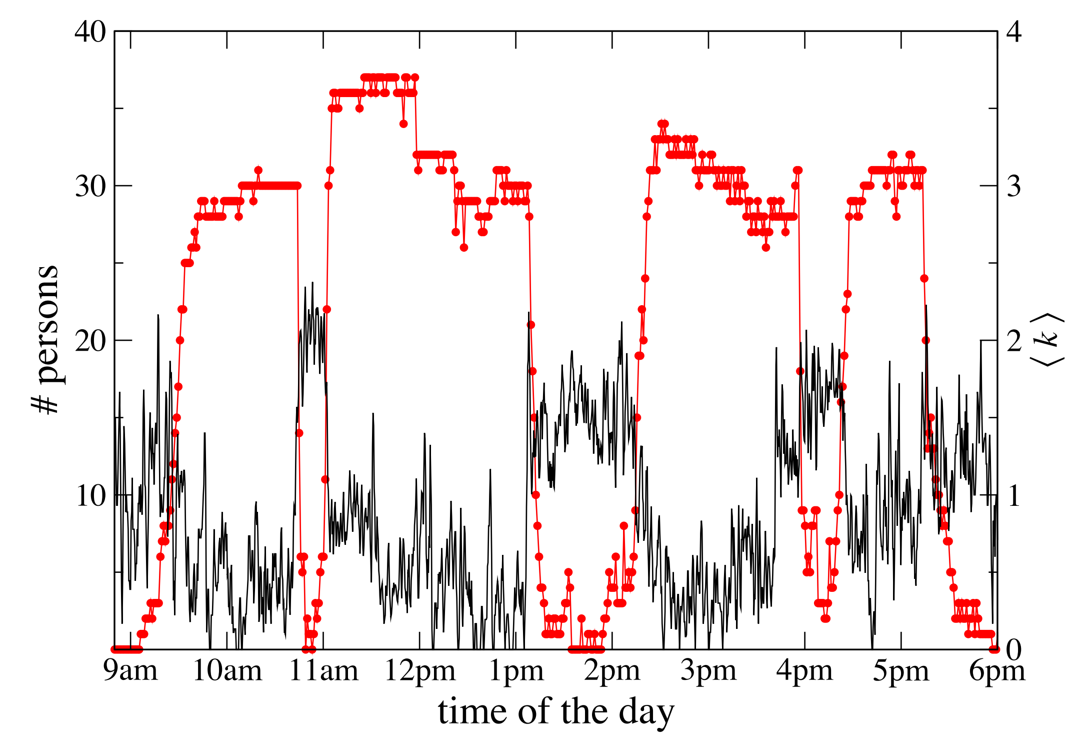


Contact durations

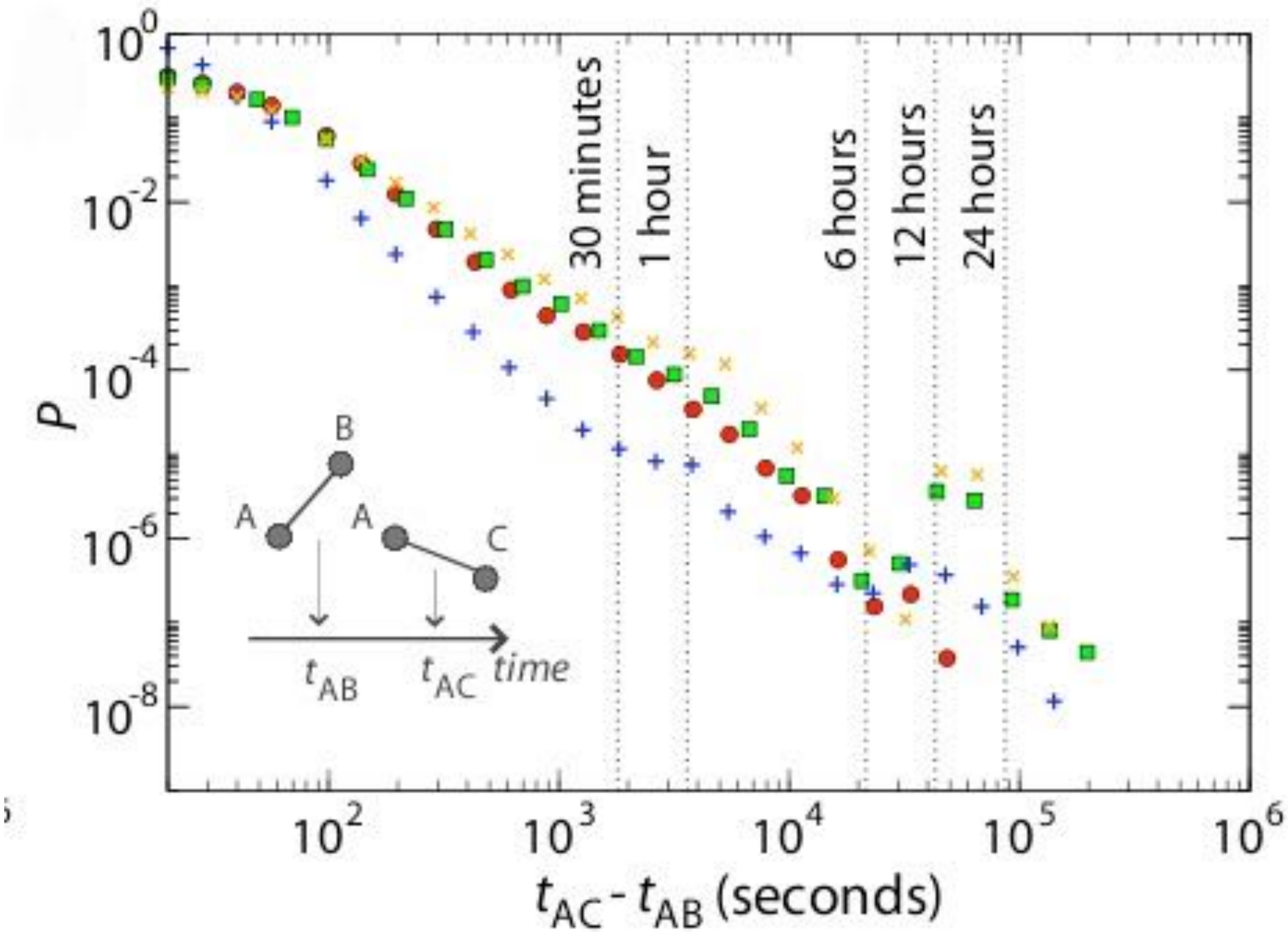
Stationary distributions



Contact duration



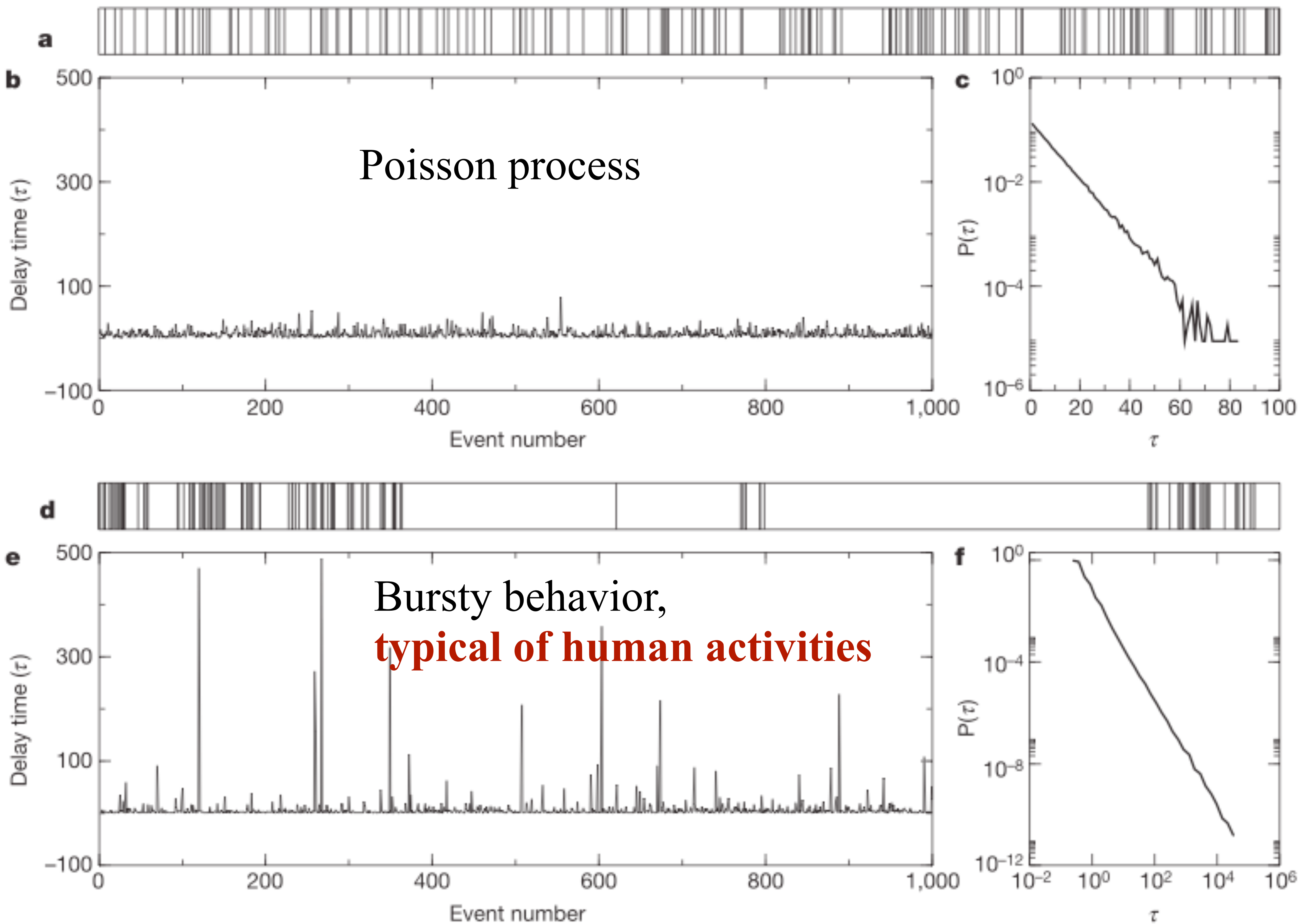
Time between contacts



Burstiness

C. Cattuto *et al.*, PLoS ONE 5(7), e11596 (2010)

Digression: Burstiness



A.-L. Barabasi, Nature (2006)

Measure of burstiness

$$B = \frac{\sigma_\tau - m_\tau}{\sigma_\tau + m_\tau}$$

where m_τ is the mean and σ_τ the std deviation of the inter-event time distribution

Poisson: B = 0 (exponential distribution)

Periodic: B = -1 (Delta distribution)

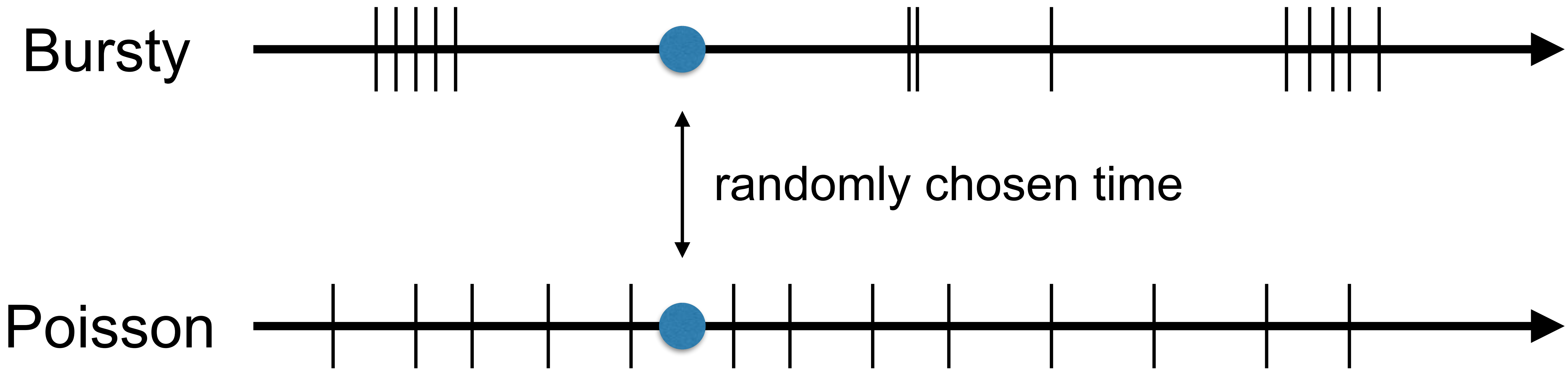
Broad distribution: B = 1 (if σ_τ diverges)

Burstiness = clustering of events in time

Goh, Barabási, EPL 2008

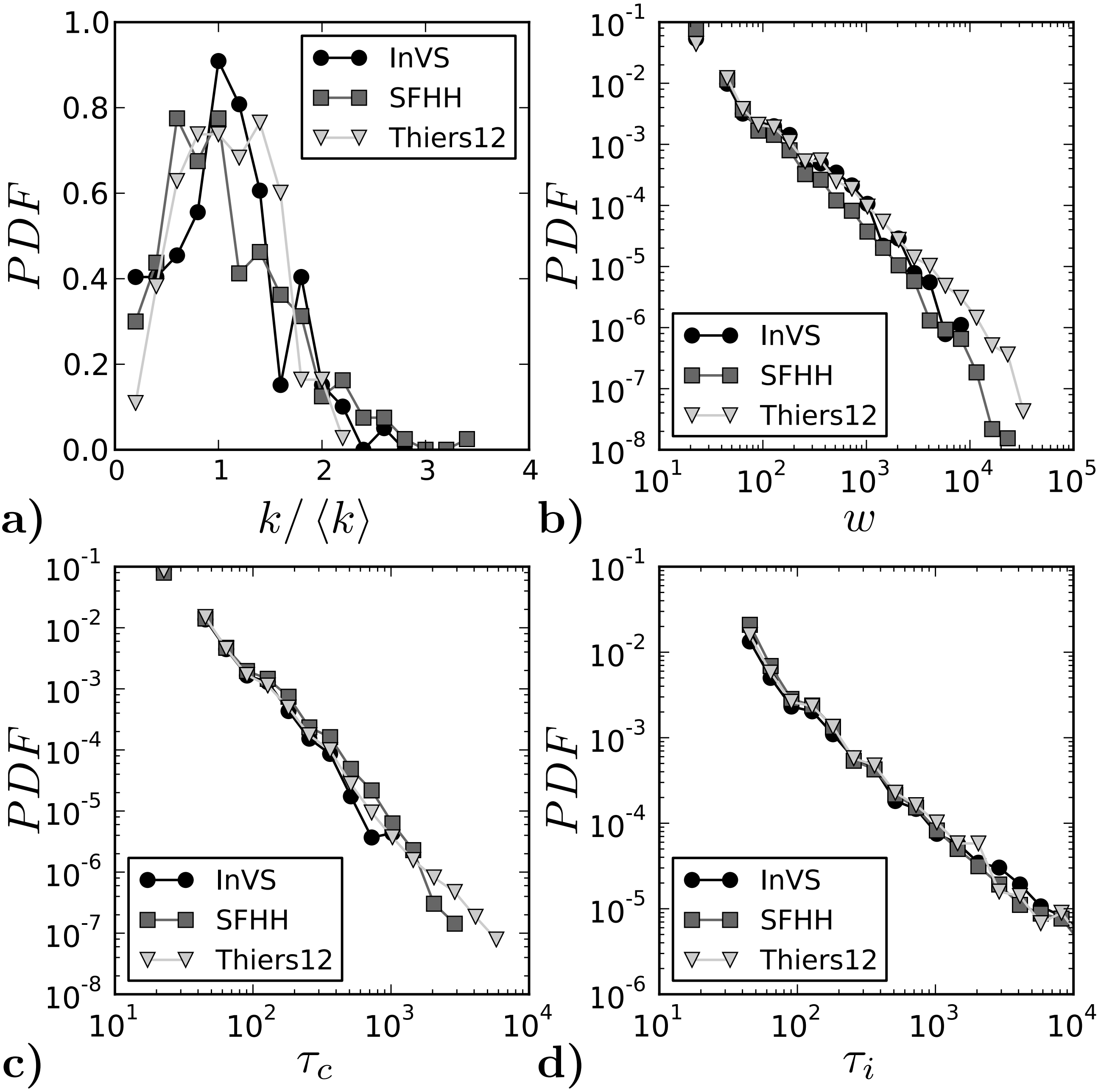
See also: Karsai et al., Sci Rep 2, 397 (2012)

Consequence of burstiness



Bursty timeline implies larger waiting time with higher probability
=> typically slows down diffusion (if no correlations)

Main data properties



Common:

- Sparse
- Large clustering
- Broad distributions of
  - contact durations
  - inter-contact durations (burstiness)
  - aggregate durations

Context-dependent:

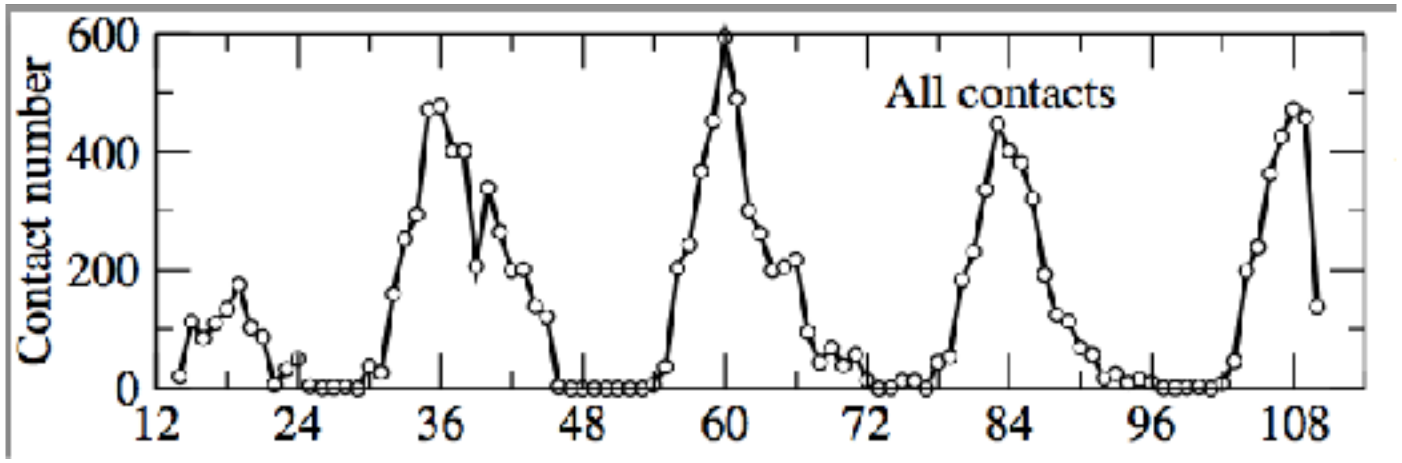
- Mixing patterns (classes for high school, departments for offices)
- Correlations (weight-topology; activity-structure)
- Longitudinal aspects

M. Génois et al., Network Science (2015)

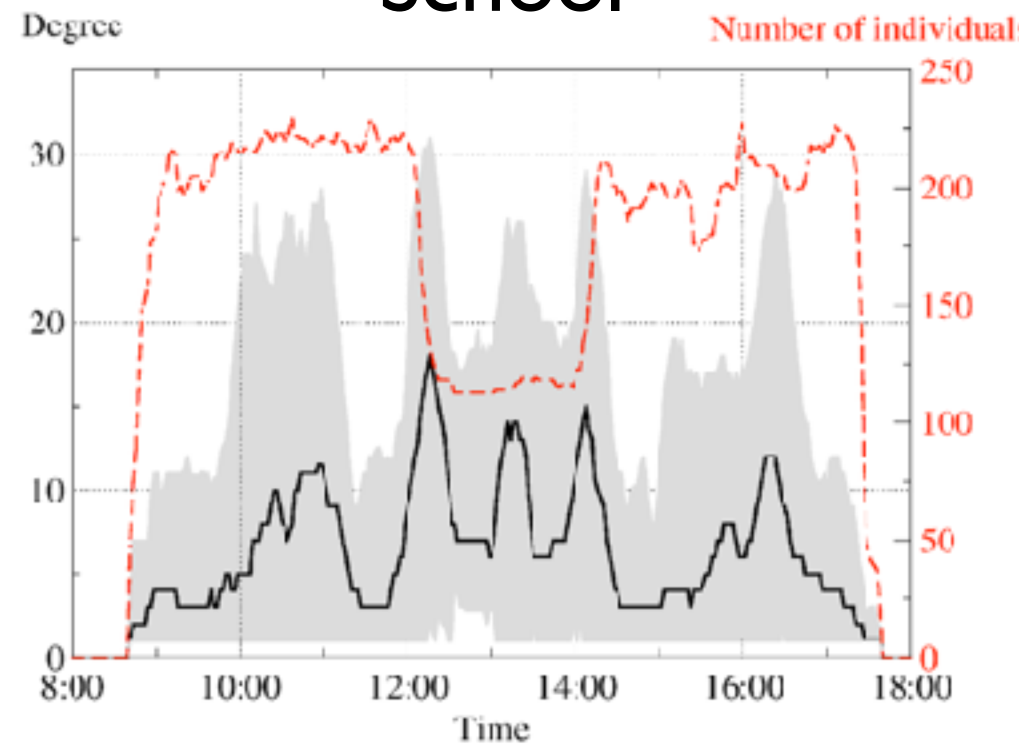
> Longitudinal aspects

Different timelines

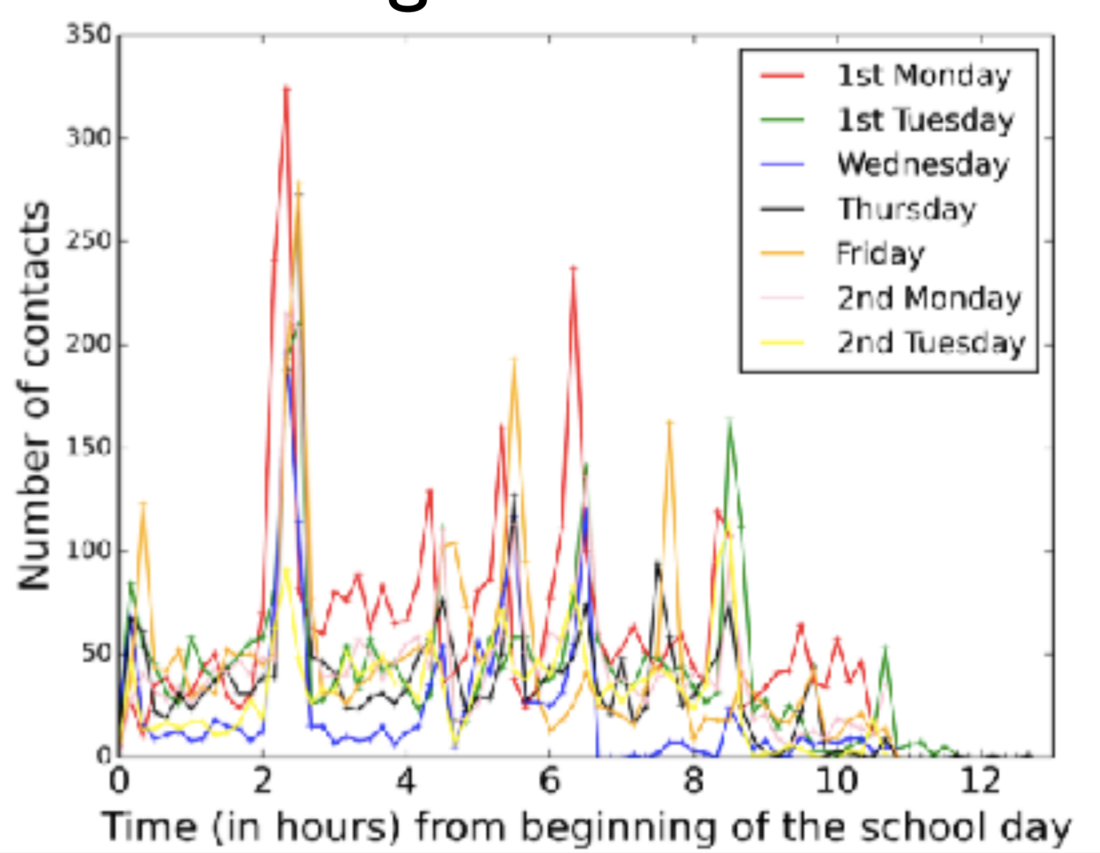
Hospital

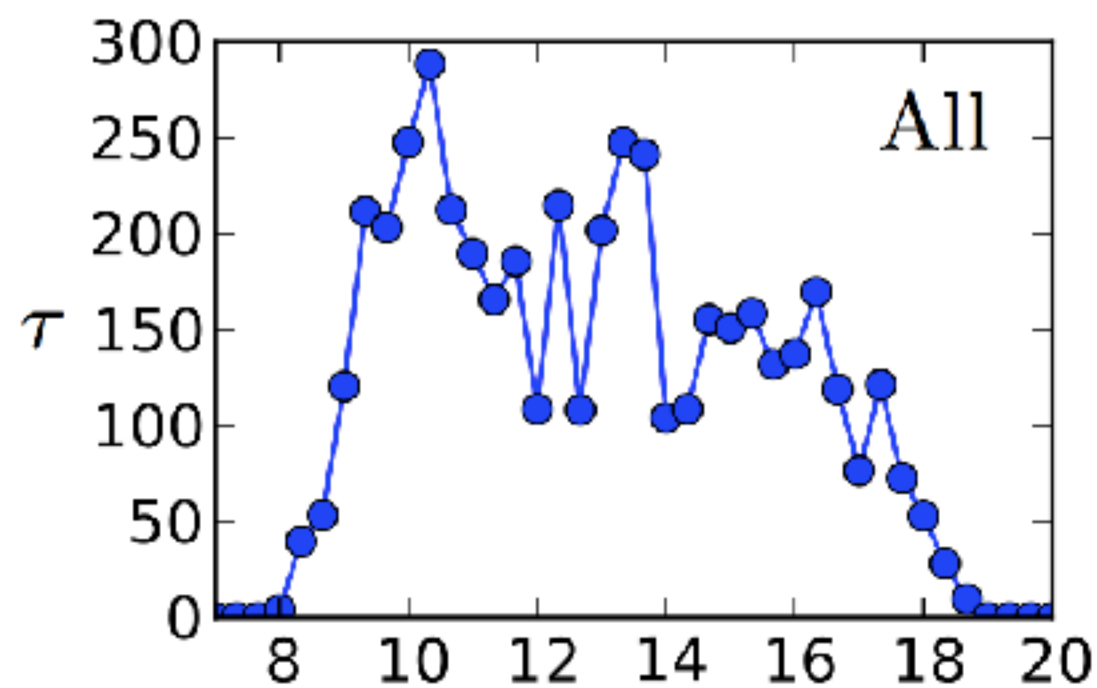


School



High-school

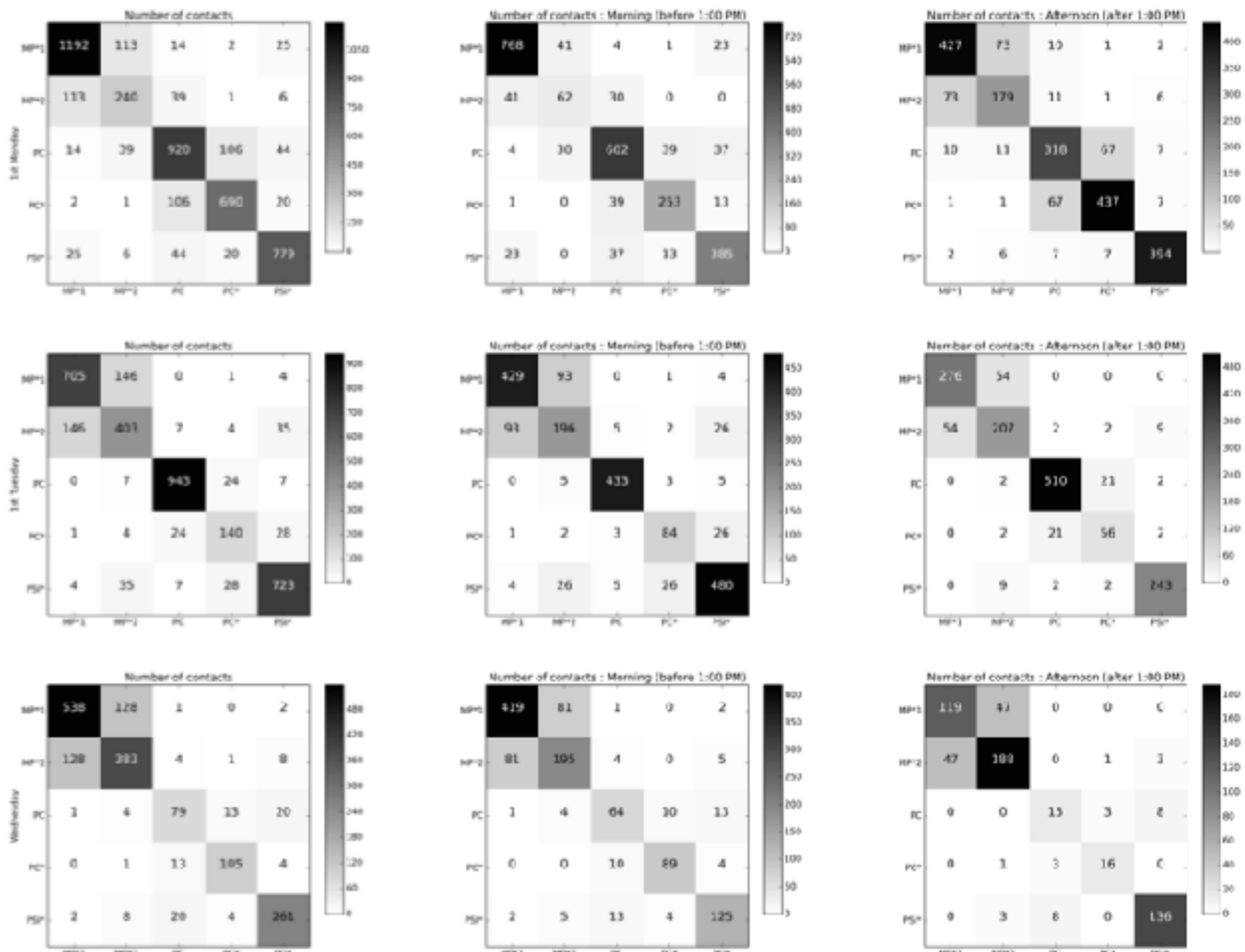




Offices

Stability of patterns and contact matrices

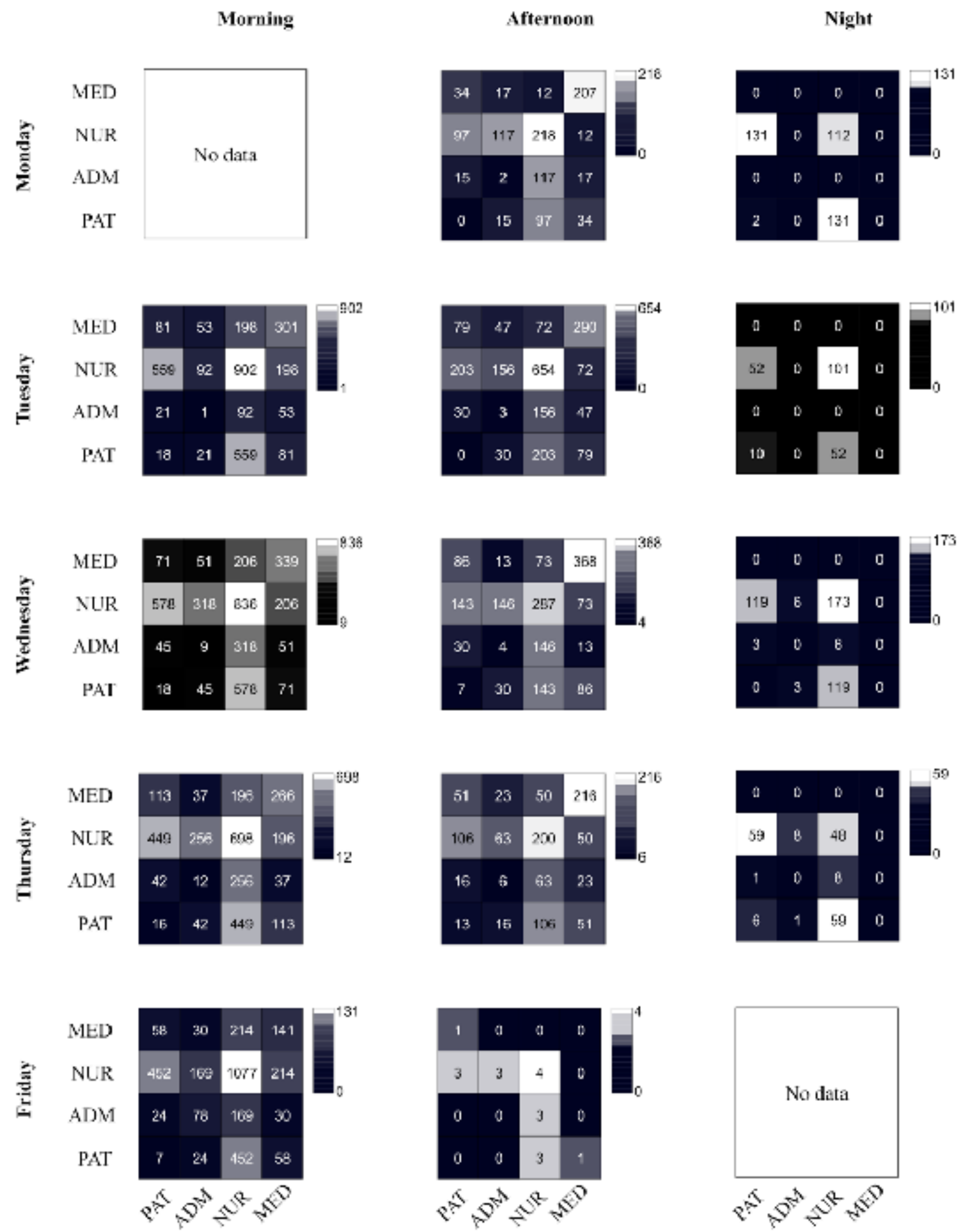
High-school, 3 days
morning & afternoons



Hospital

Morning, afternoon and night shifts

Lyon, geriatric ward

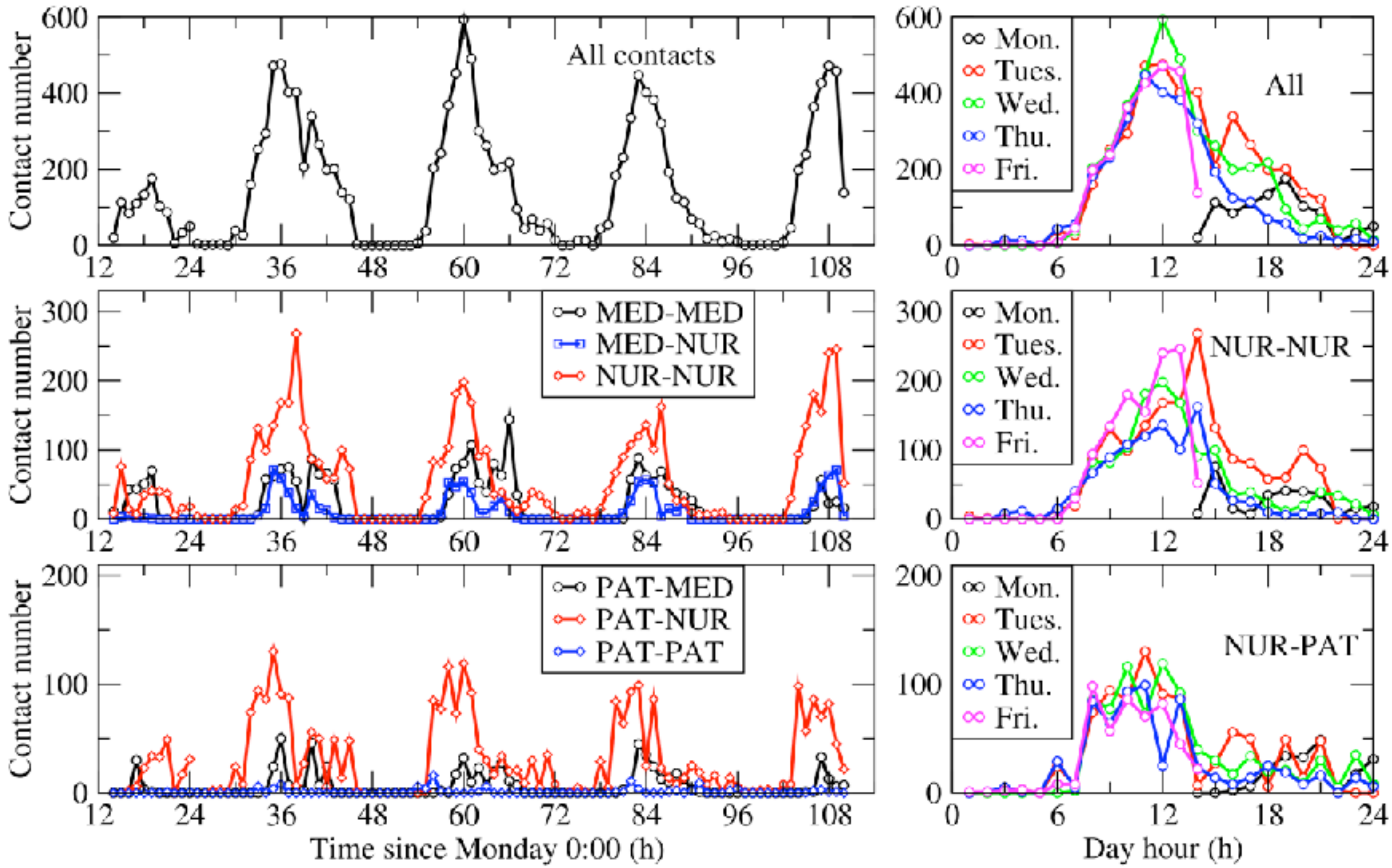


P. Vanhems et al., PLOS ONE (2013)

Longitudinal evolution

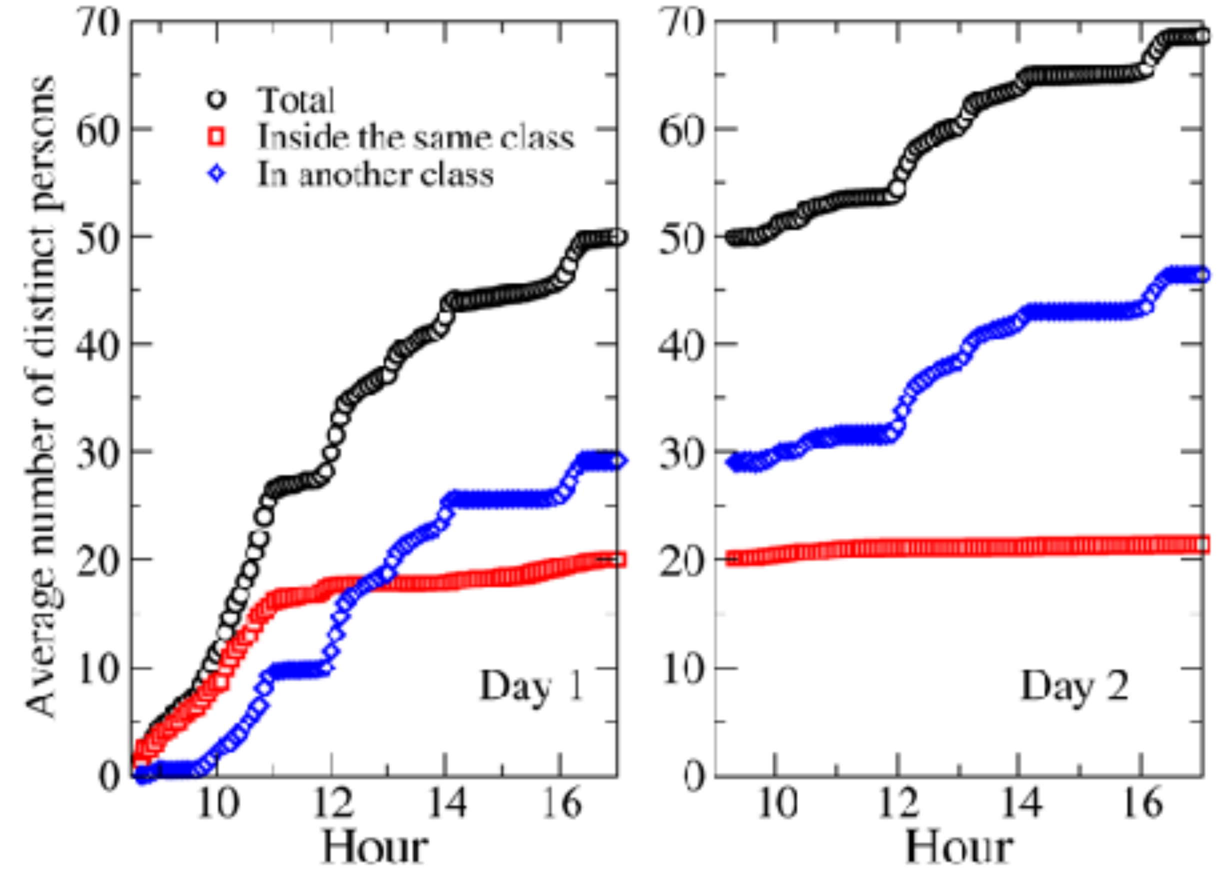
Lyon, geriatric ward

Robustness of daily patterns

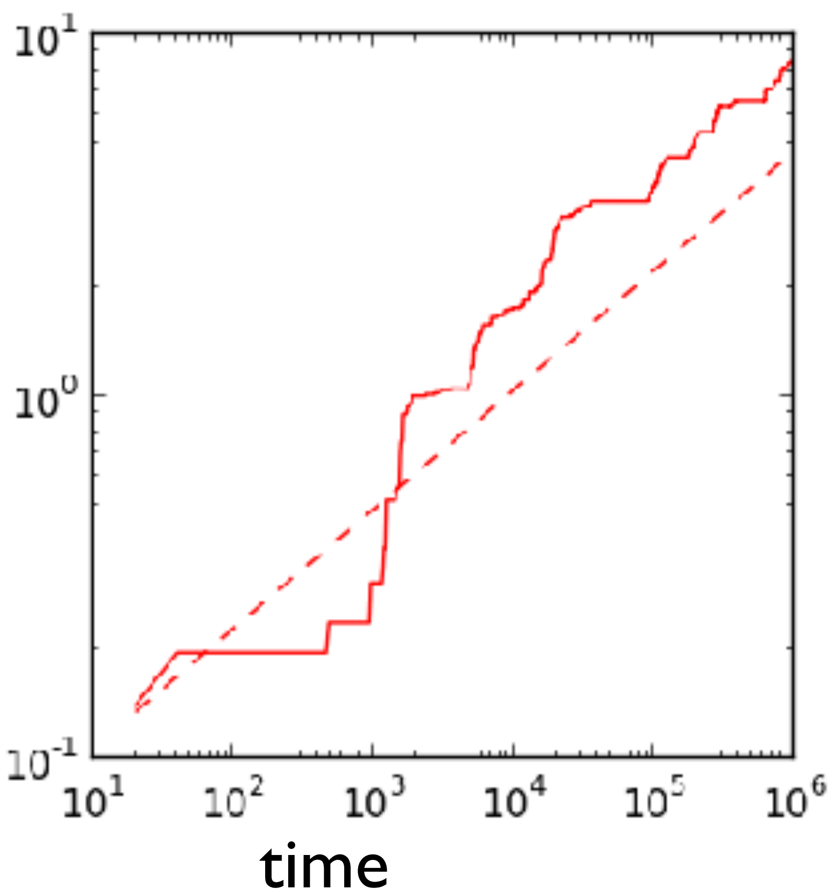


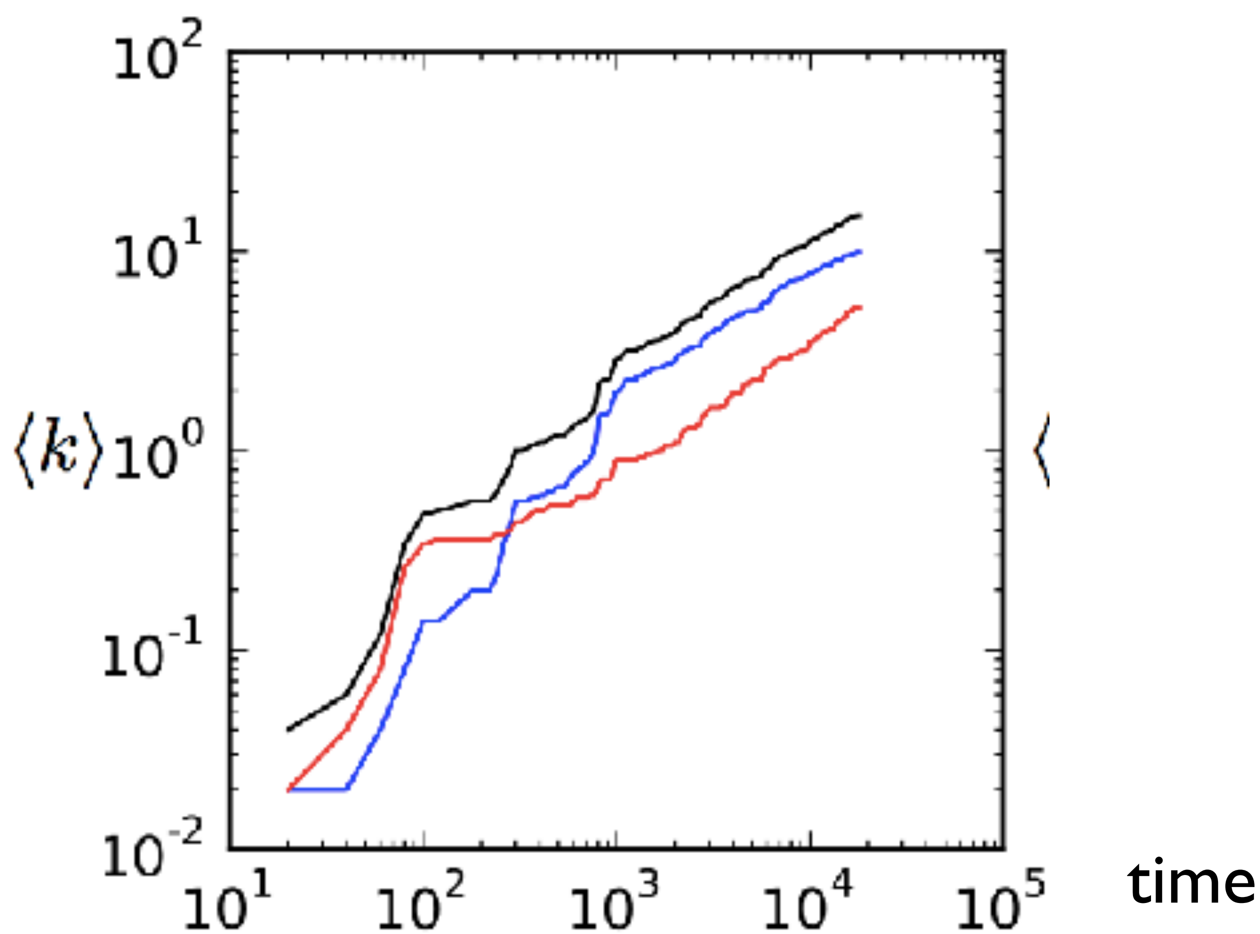
P. Vanhems et al., PLOS ONE (2013)

Stability in global patterns, but substantial changes in neighbourhoods



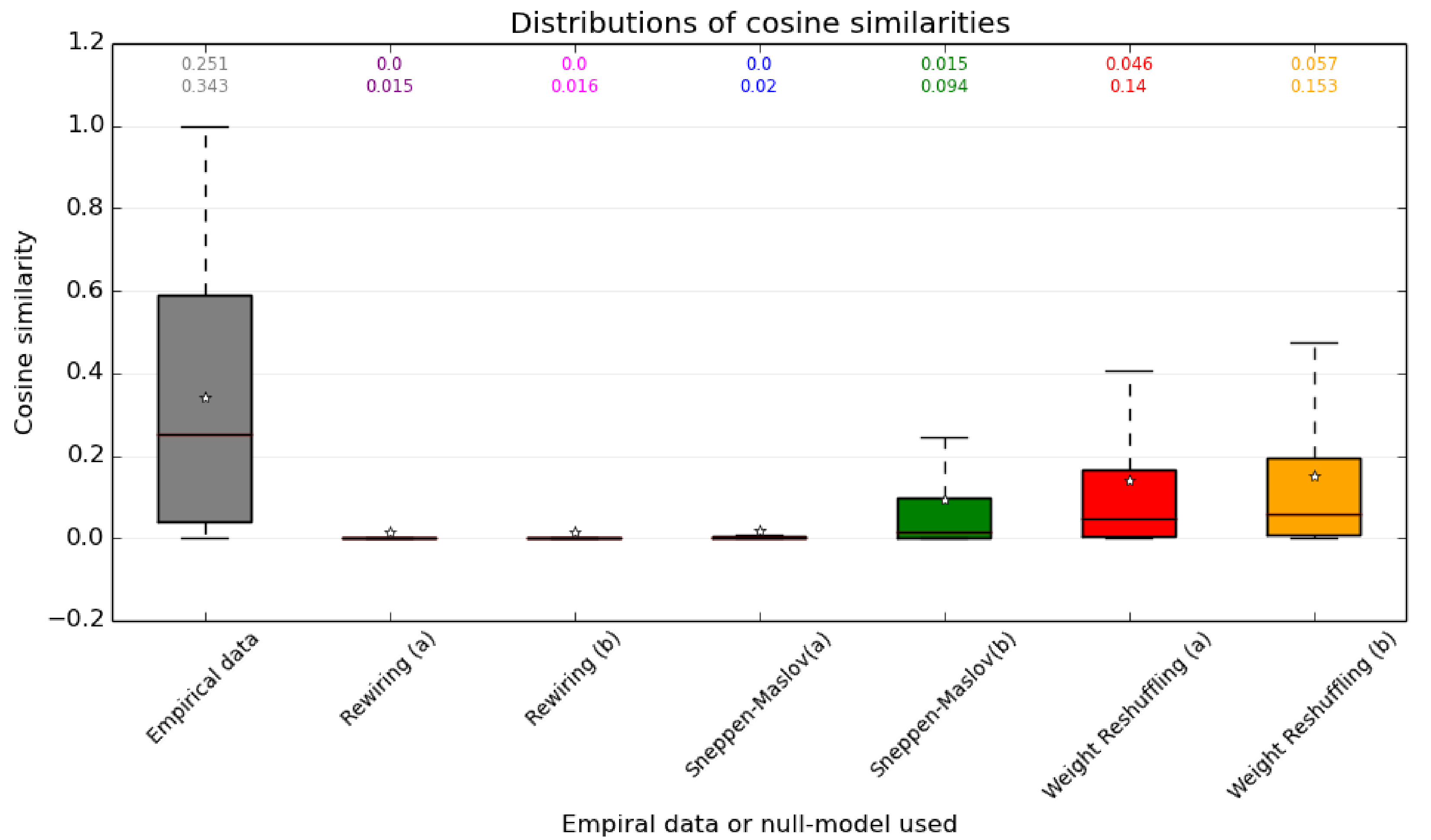
Number of distinct persons contacted





Quantifying neighborhood similarity

High-school data



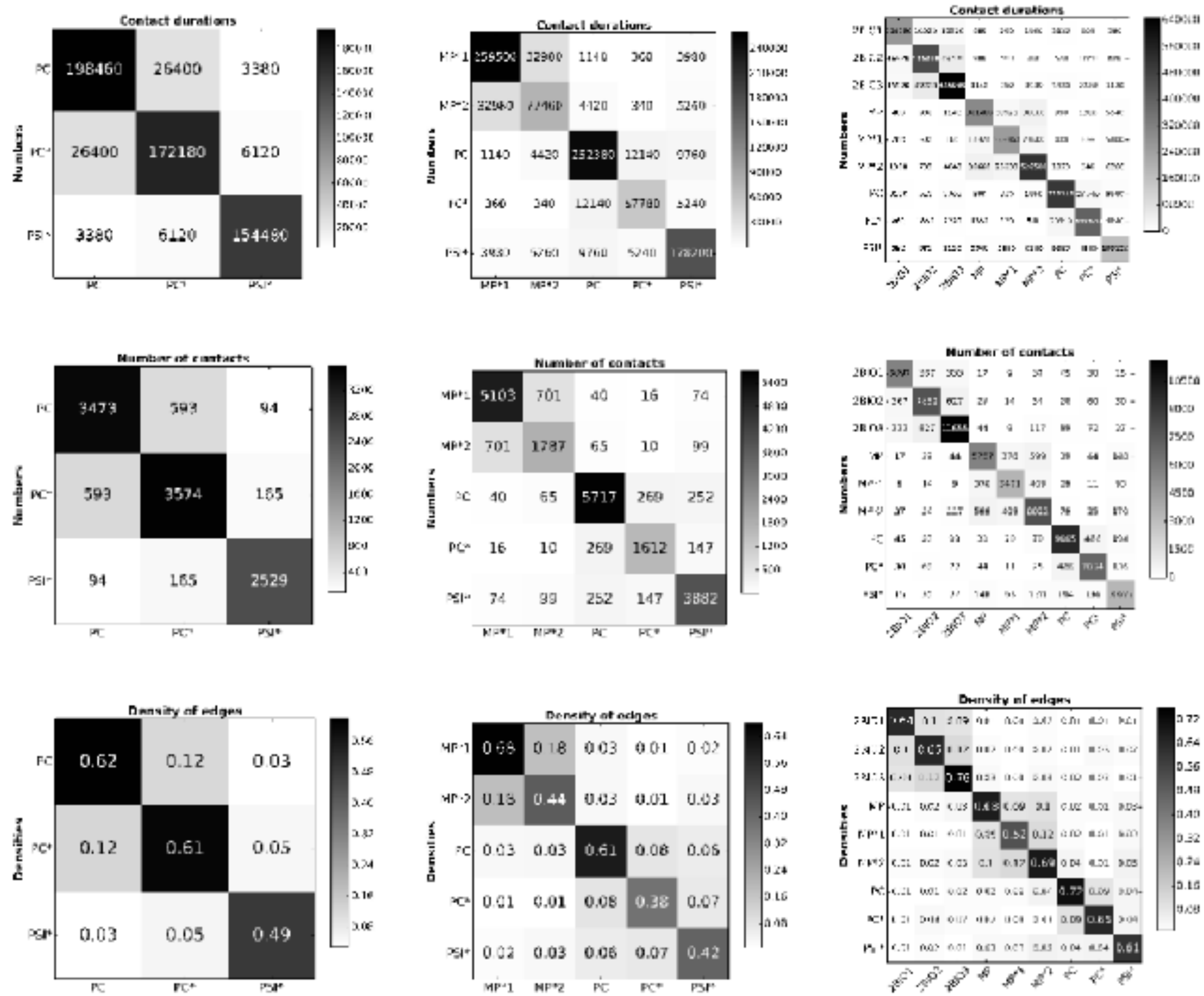
- **Substantial differences** in neighborhood from one day to the next
- **Much less than by random chance**
- True for all investigated contexts
- Similar rates of neighbourhood changes in different years (high-school)

(this plays an important role e.g.
in data-driven simulations of epidemic spreading)

Long timescales

Stability across years
(Highschool)

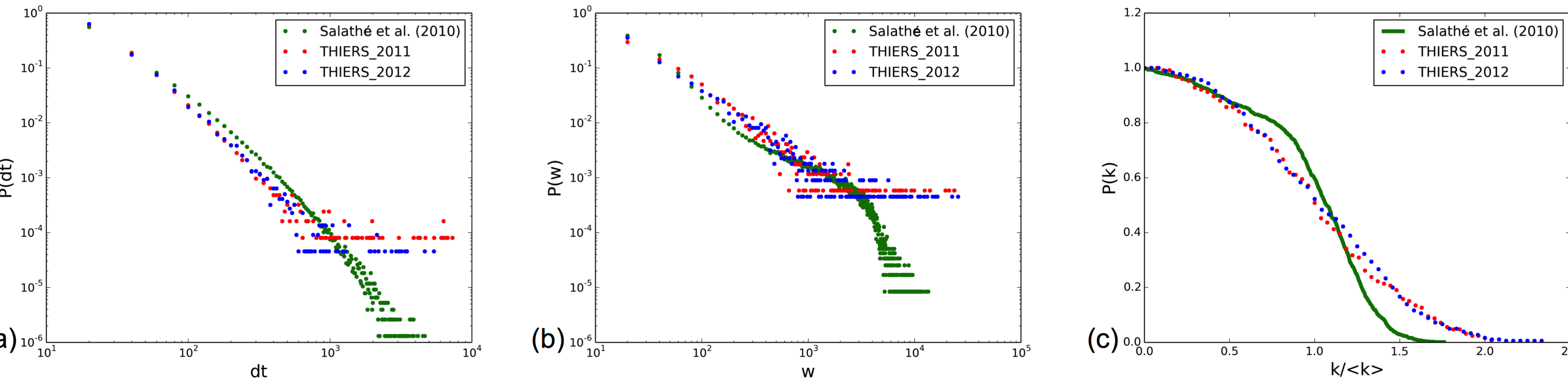
- Very large similarity of contact matrices in both years
- Similar values of cosine similarities between neighbourhoods of nodes in different days, for different years



Stability of statistical properties

-Marseilles high school: 2011, 2012
same context, different individuals

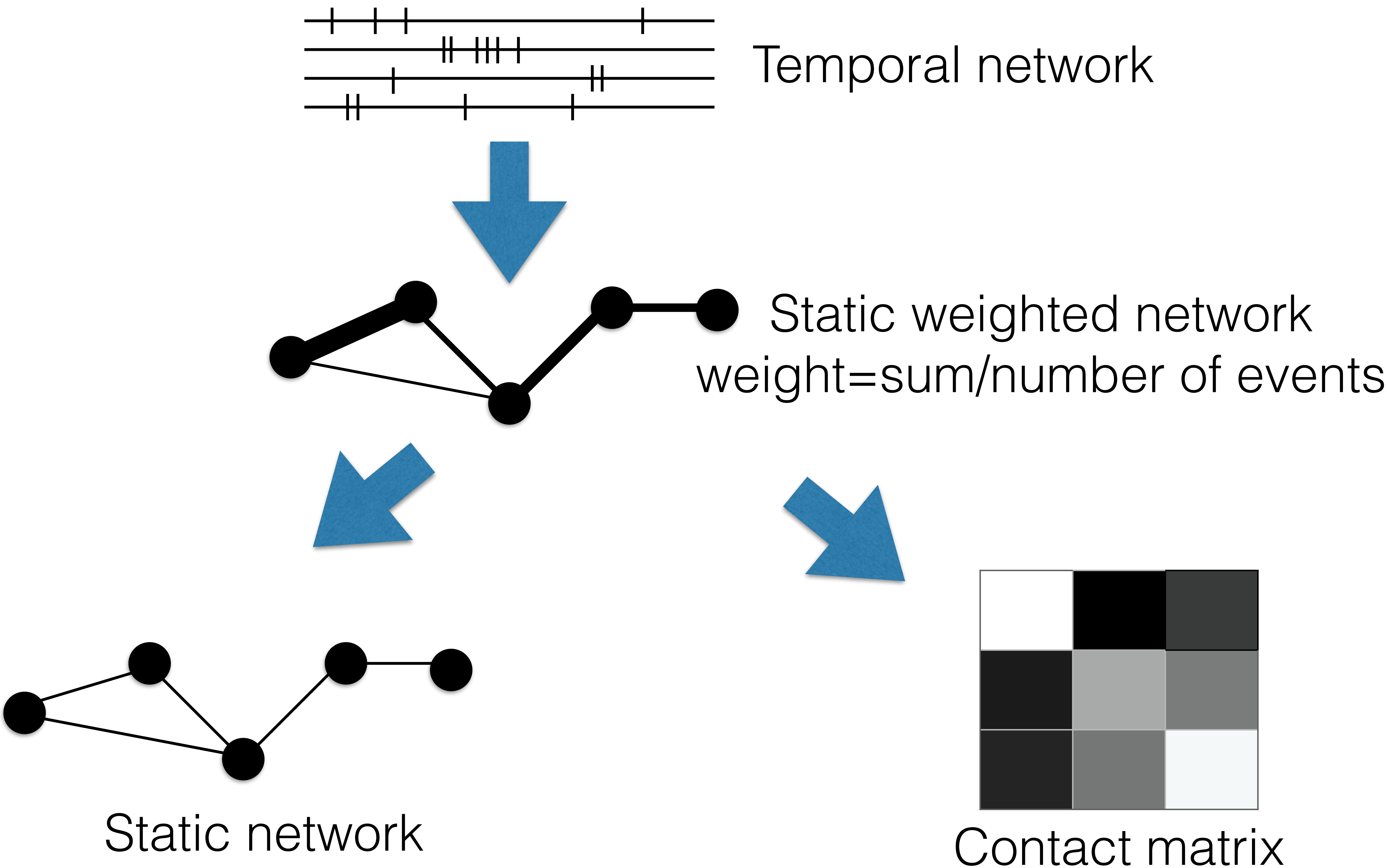
-US high school (Salathe et al., 2010)



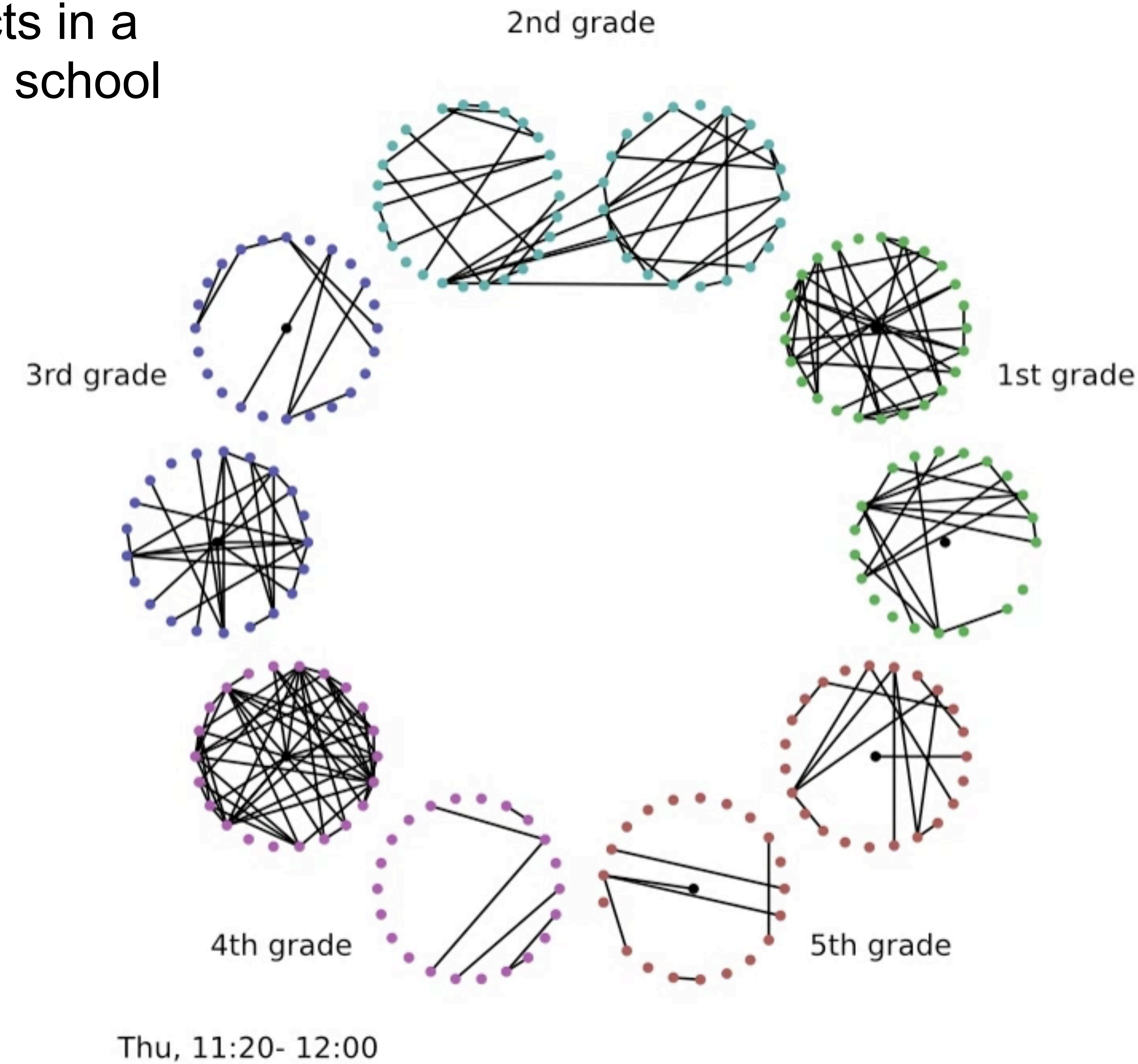
J. Fournet, A. Barrat, PLoS ONE (2014)

Representing data, finding structures and features

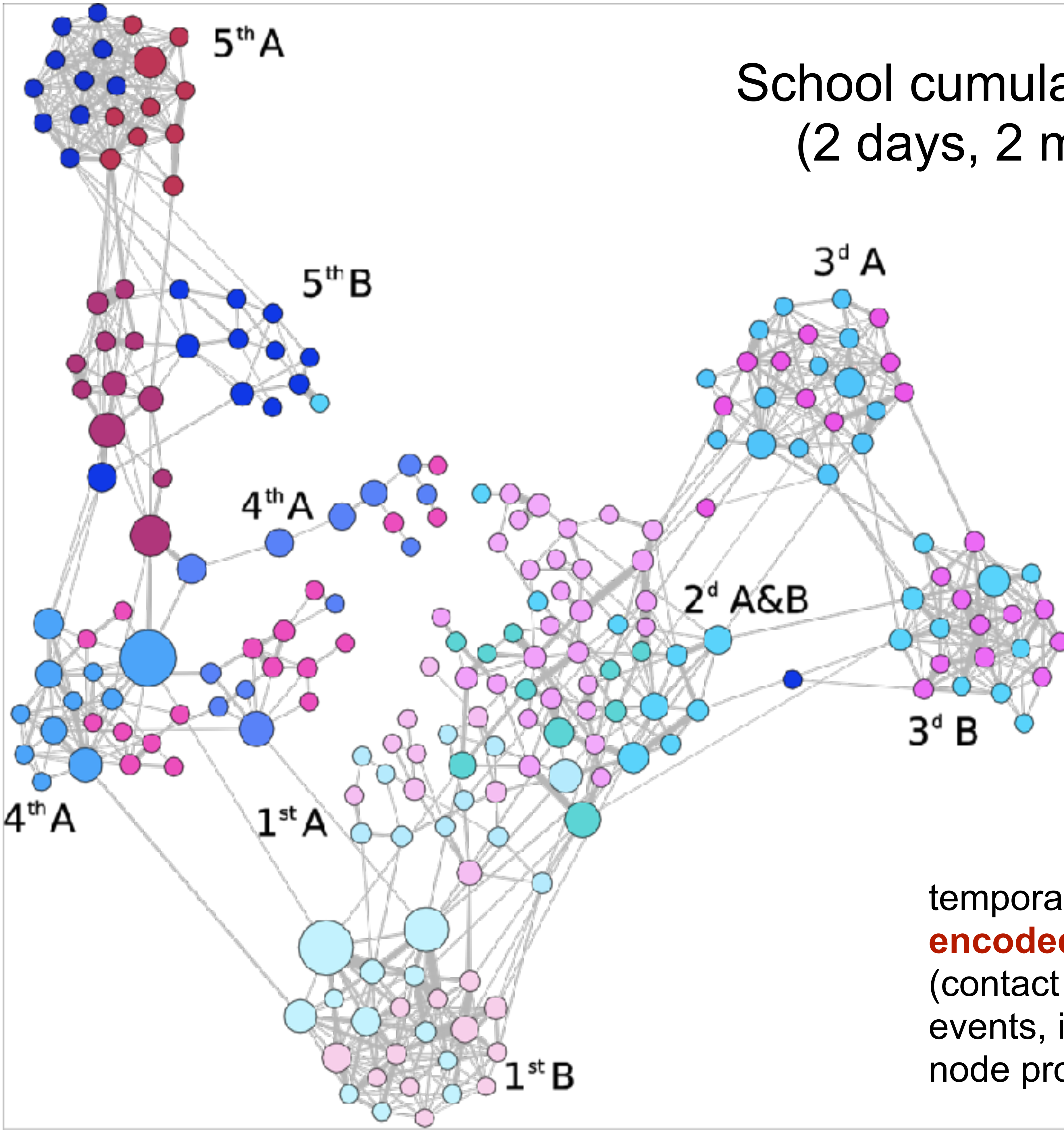
Aggregation of temporal networks



contacts in a
primary school



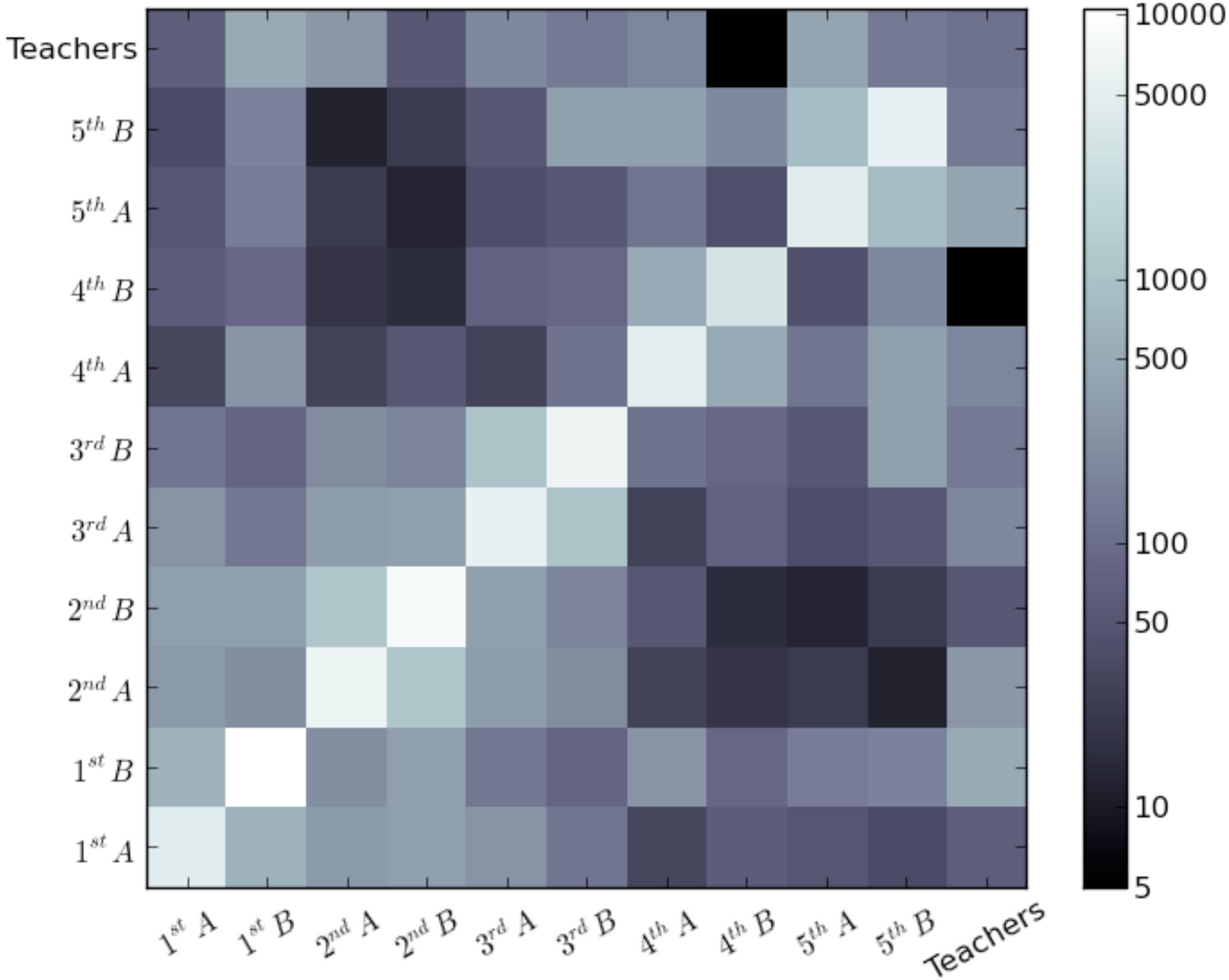
School cumulative f2f network (2 days, 2 min threshold)



J. Stehlé et al. PLoS ONE 6(8):e23176 (2011)

temporal information:
encoded in edges' weights
(contact duration(s), number of events, intermittency, etc...) and node properties

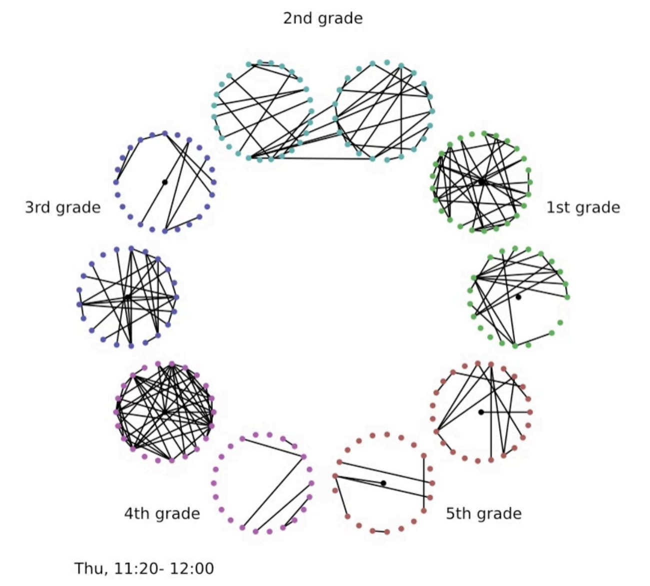
contact matrices



Average values: little information

J. Stehle, et al.
High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School
PLoS ONE 6(8), e23176 (2011)

Which level of detail?



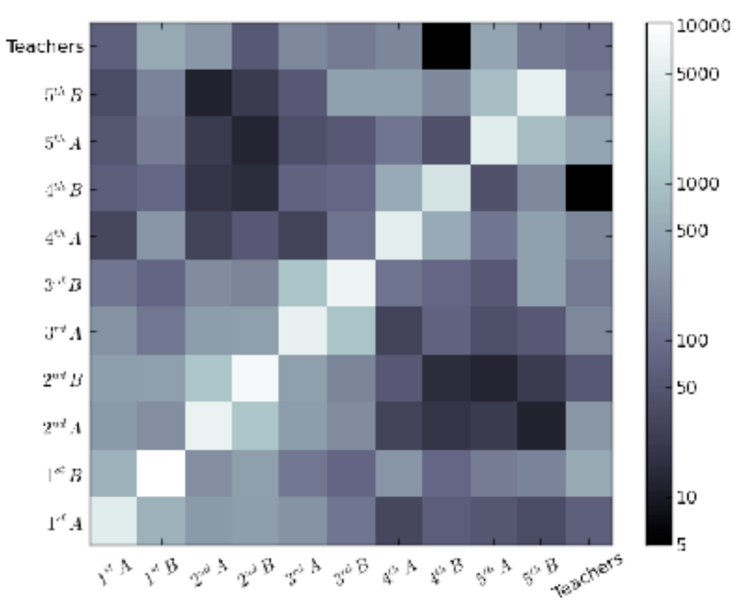
Detailed dynamic network

• very detailed ✔

• very realistic ✔

• takes into account individual heterogeneities of behavior ✔

• very specific (context+period), not easy to generalize ✕



Contact matrix

• coarse-grained ✕

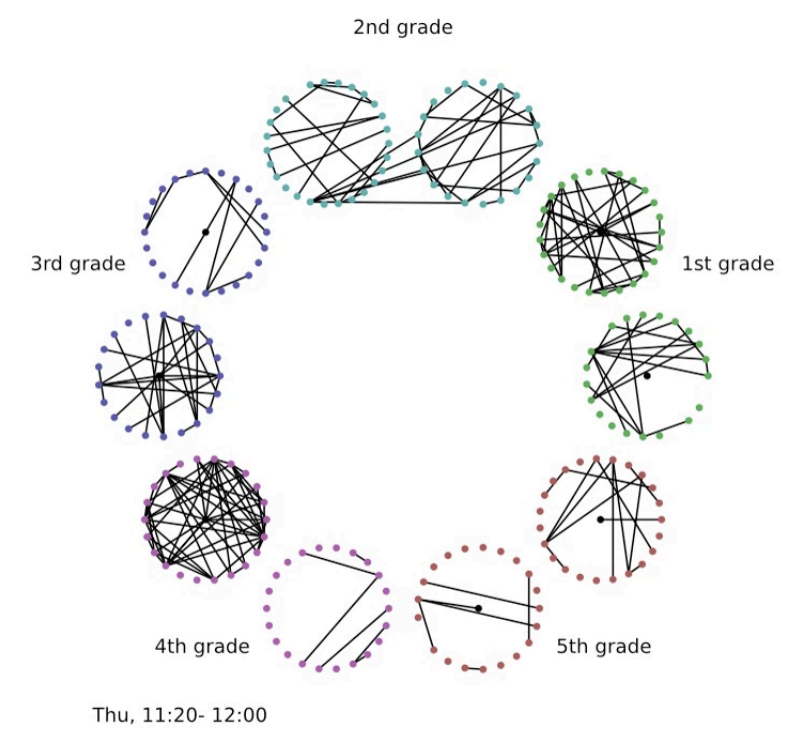
• fully connected structure ✕

• only heterogeneities between groups ✕

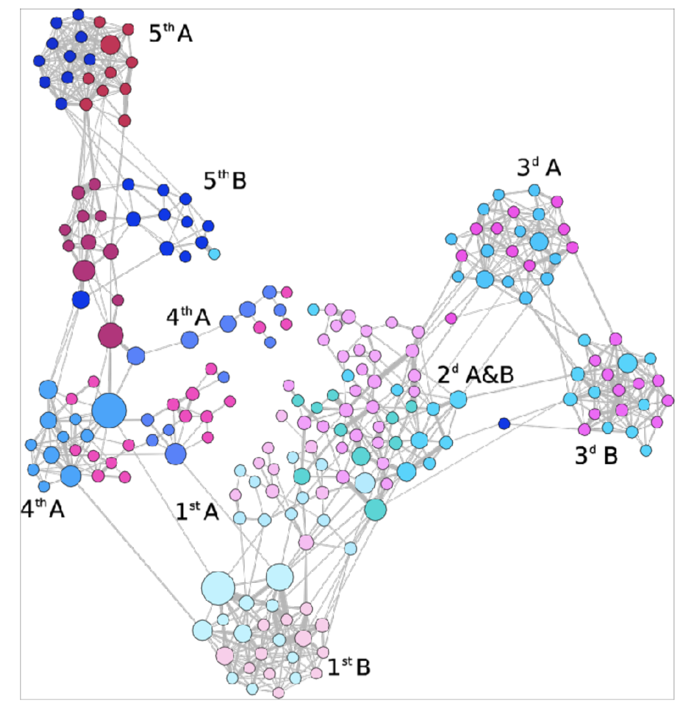
• very easy to generalize ✔

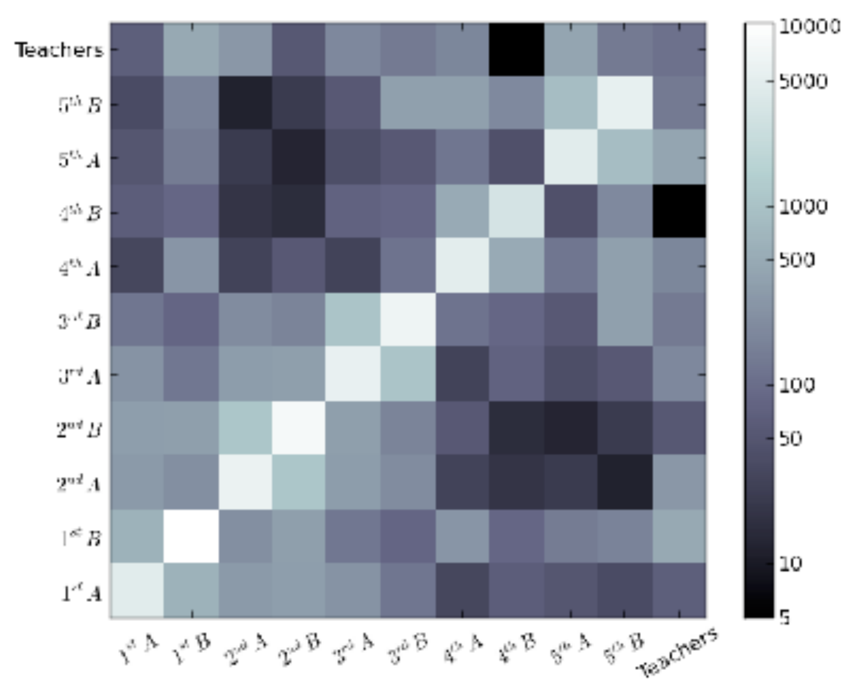
“synopsis” of dynamic network data

Temporal network



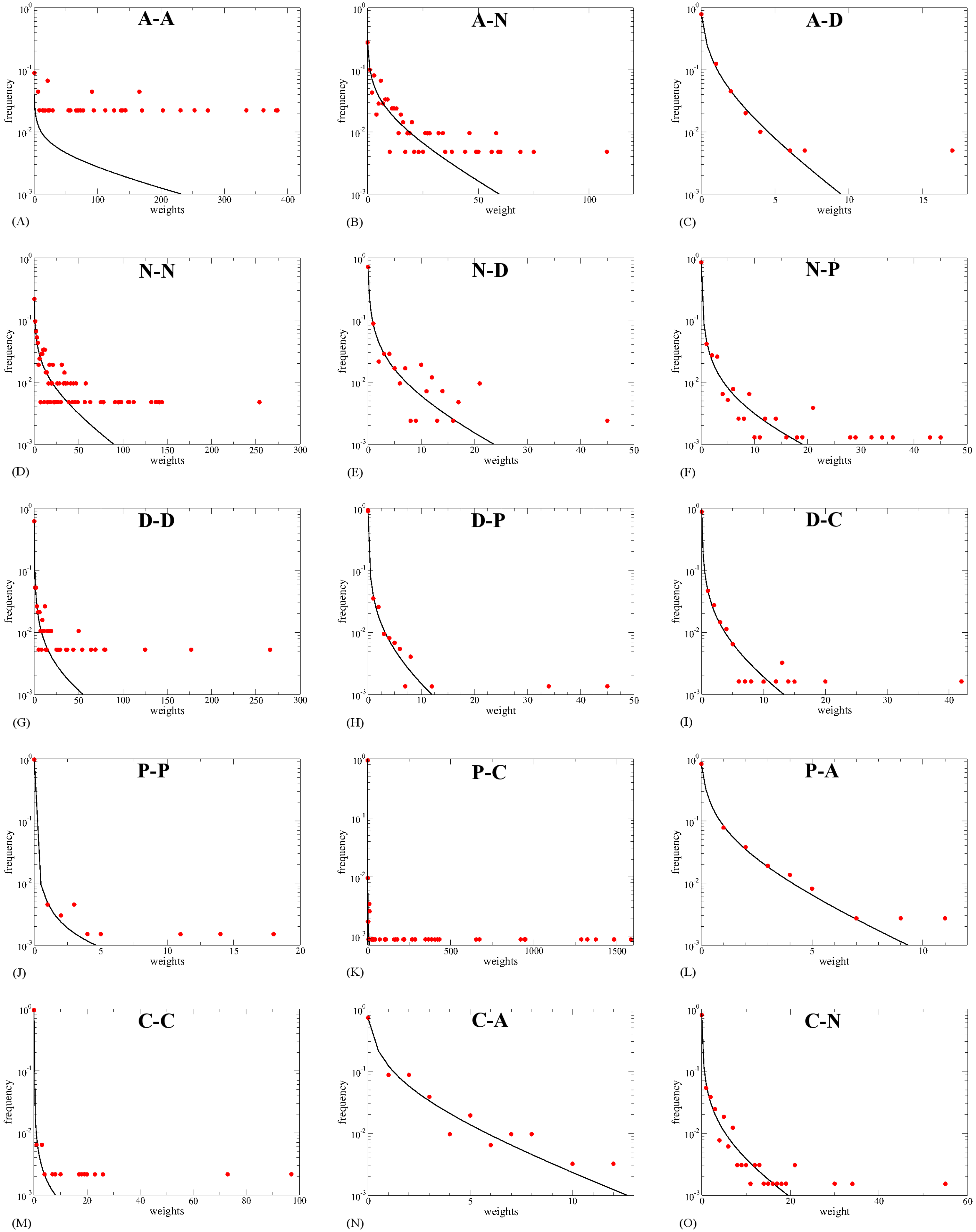
Static network





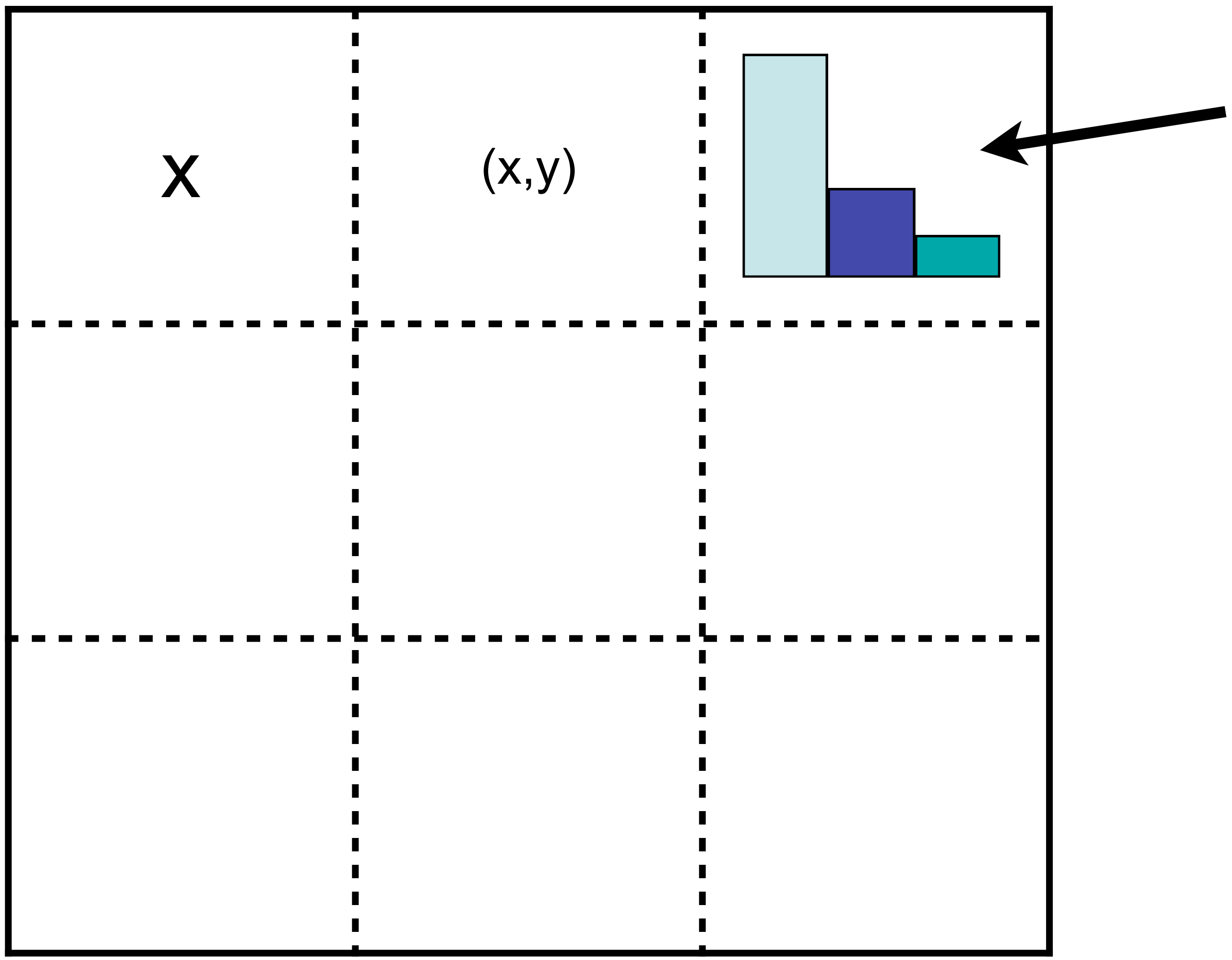
Contact matrix
(underlying fully connected assumption + **no within-class heterogeneity**)

Example: cumulated contact duration distributions between individuals of different roles in the hospital (and neg. binomial fits)



A. Machens et al., BMC Inf. Dis. (2013)

Intermediate representations? example:



Distribution of weights. of numbers and durations of events,…

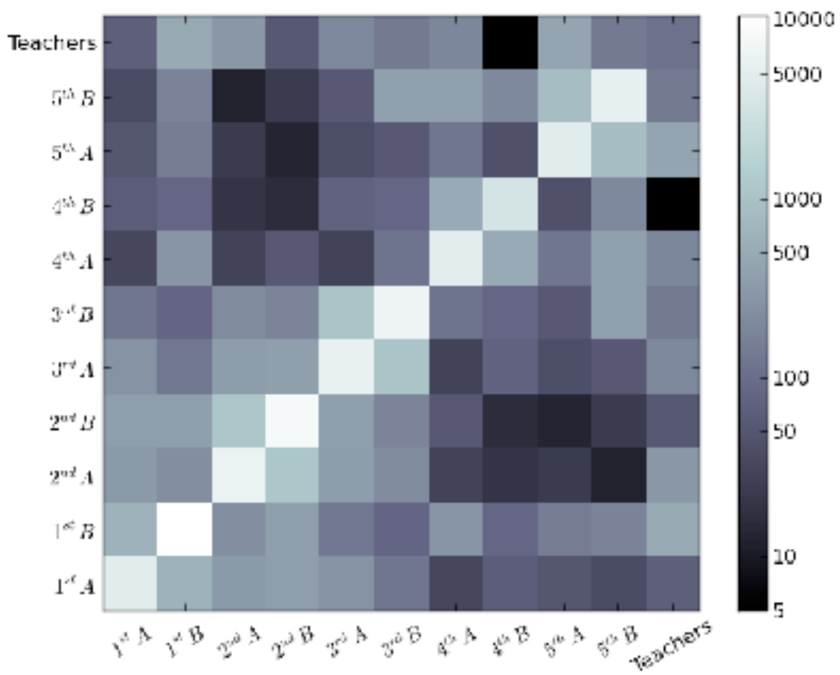
Contact matrices of distributions

A. Machens et al., BMC Inf. Dis. (2013)
M. Génois et al., Nat Comms (2015)

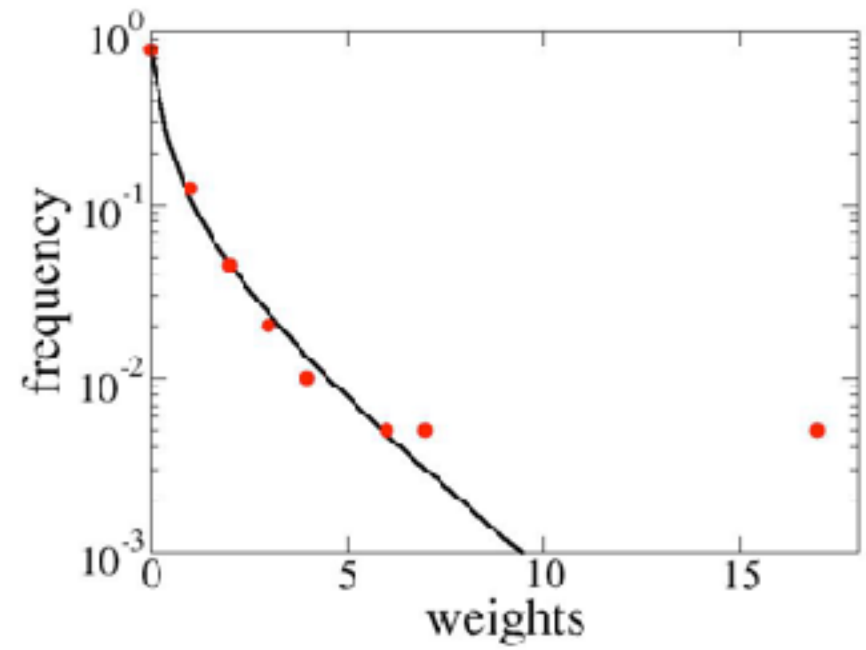
NB: useful for incomplete datasets

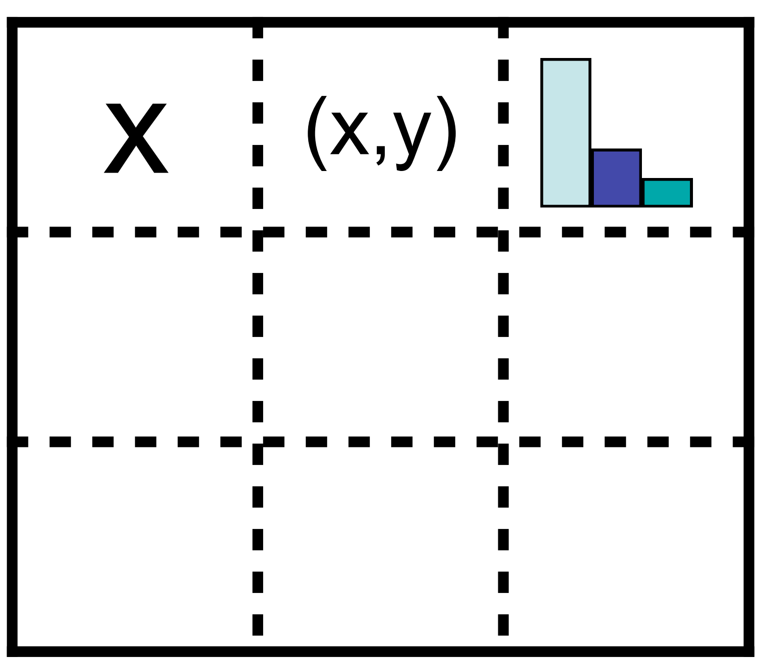
“synopsis” of dynamic network data

Role-based structure



Heterogeneities





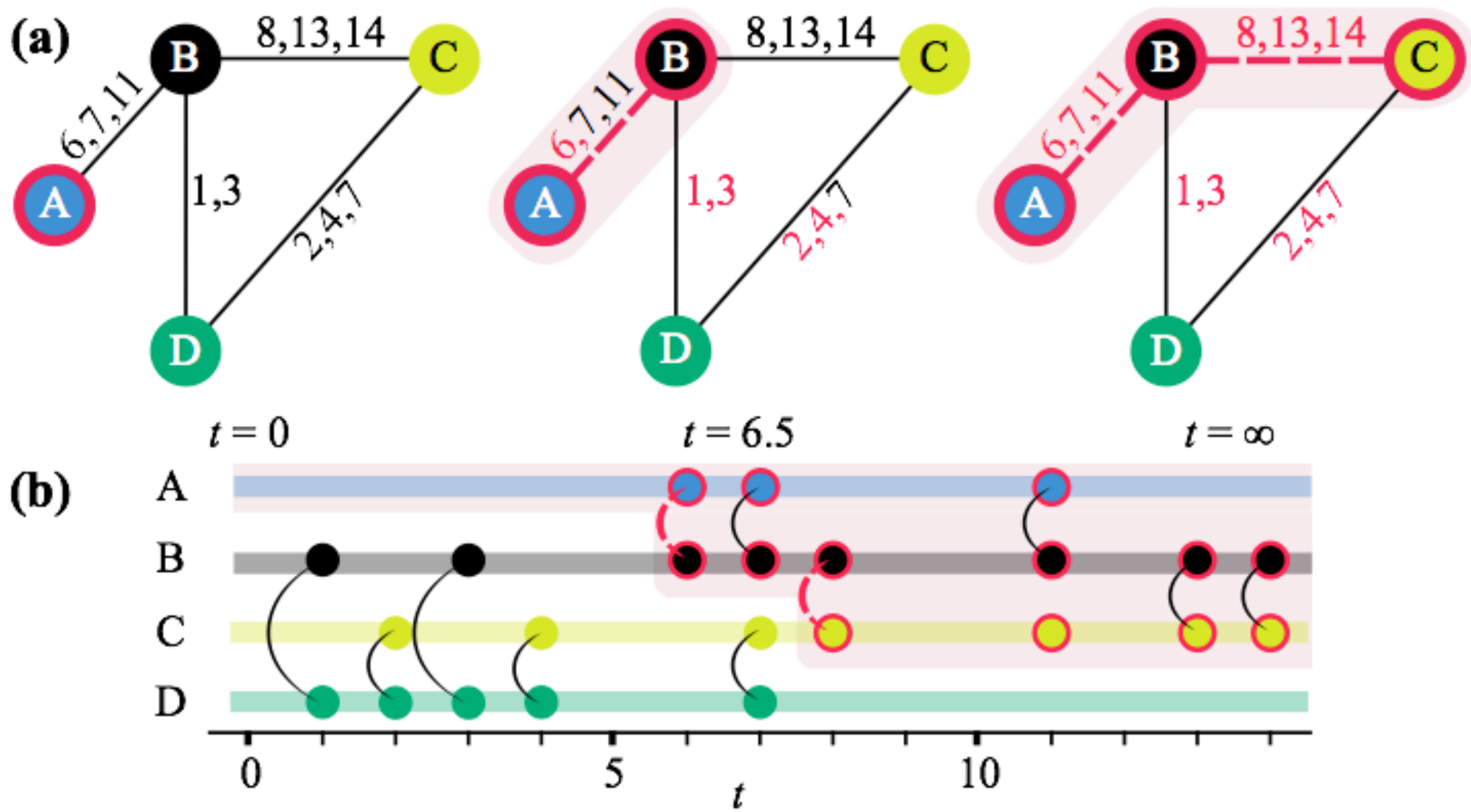
Contact matrix of distributions:

-role based

-takes heterogeneities into account

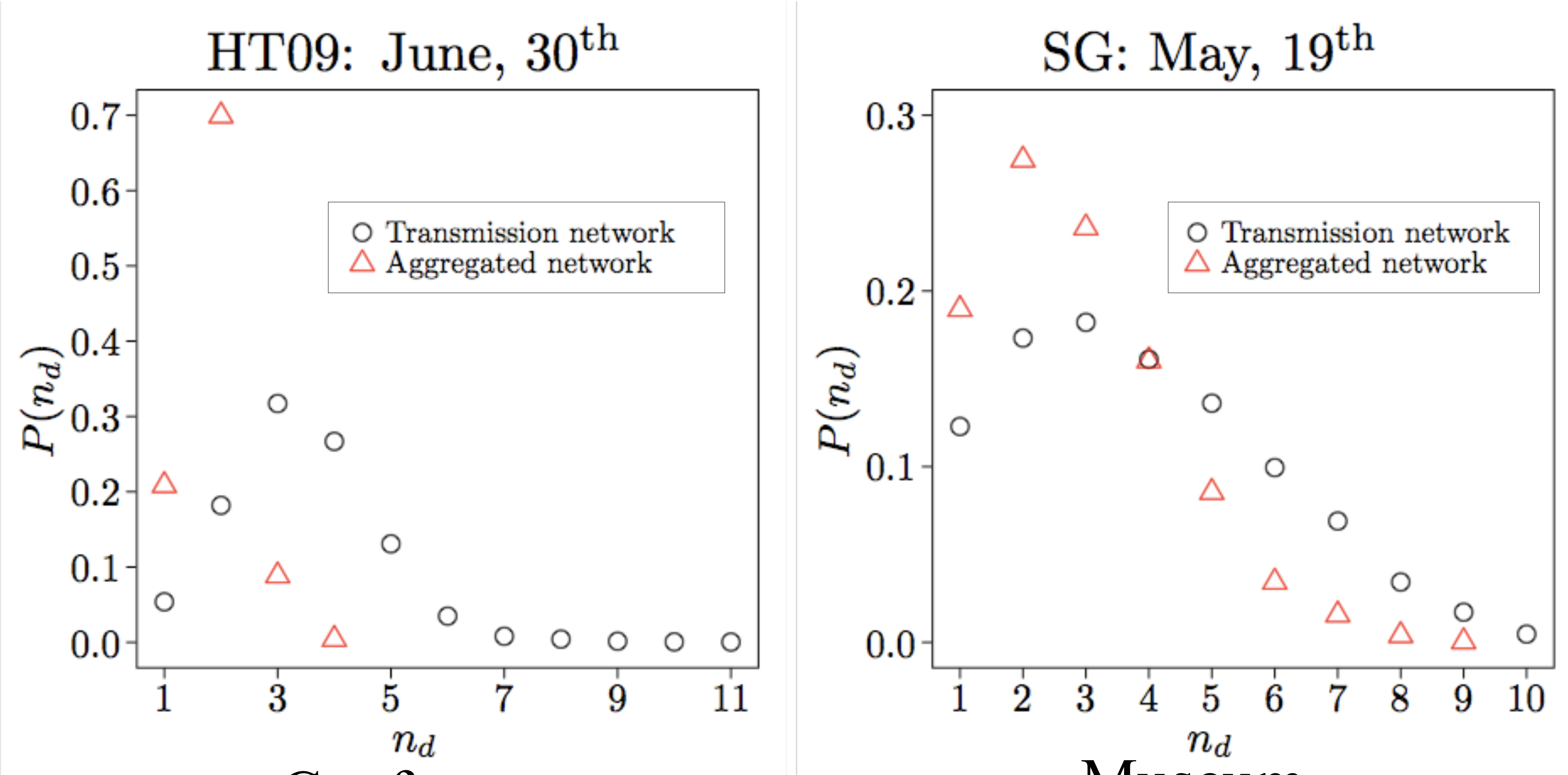
A. Machens et al., BMC Inf. Dis. (2013)
M. Génois et al., Nat Comms (2015)

Temporality matters: reachability issue



Review Holme-Saramaki, Phys. Rep. (2012), arXiv:1108.1780

Shortest paths on aggregated network vs fastest paths



Conference

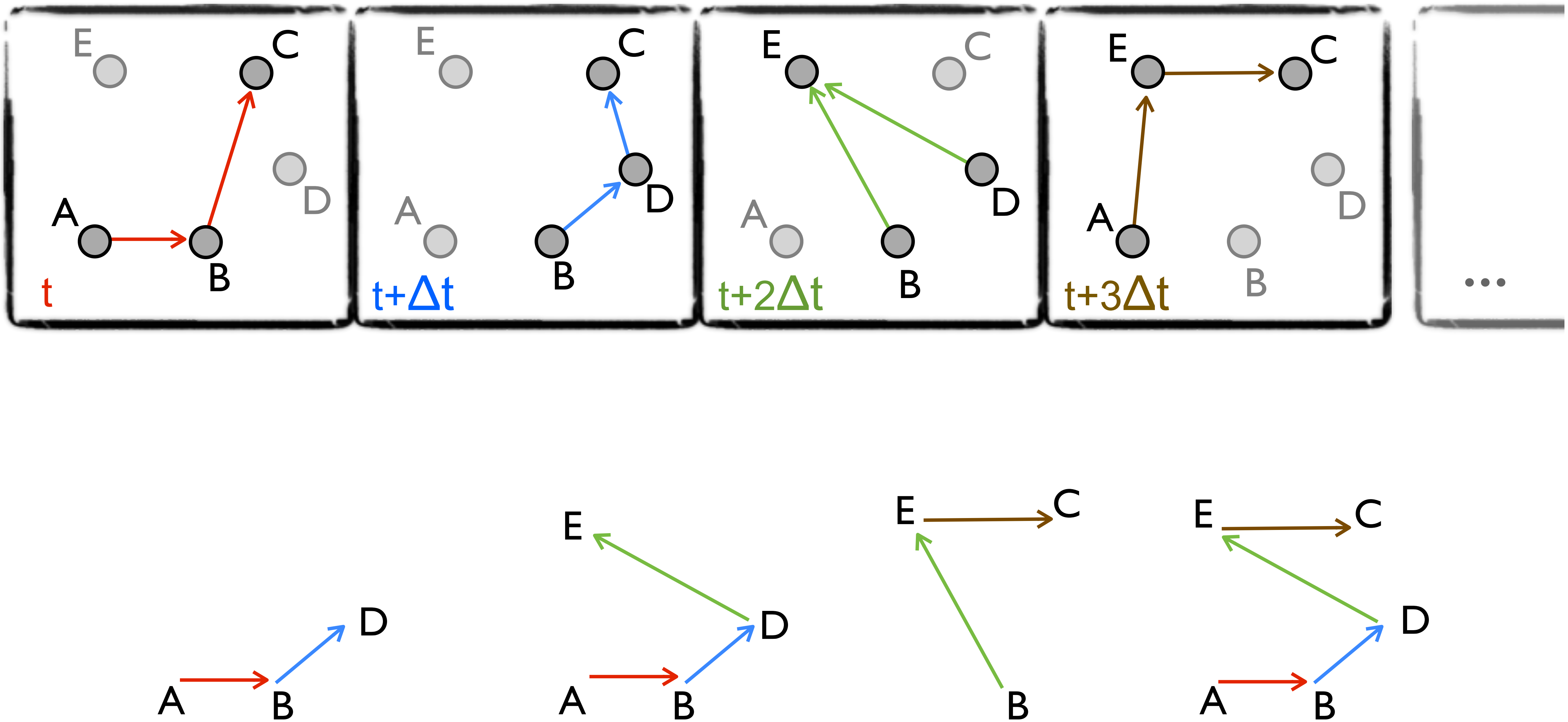
Museum

(face-to-face contact networks)

L. Isella *et al.*, Journal of Theoretical Biology 271, 166 (2011)

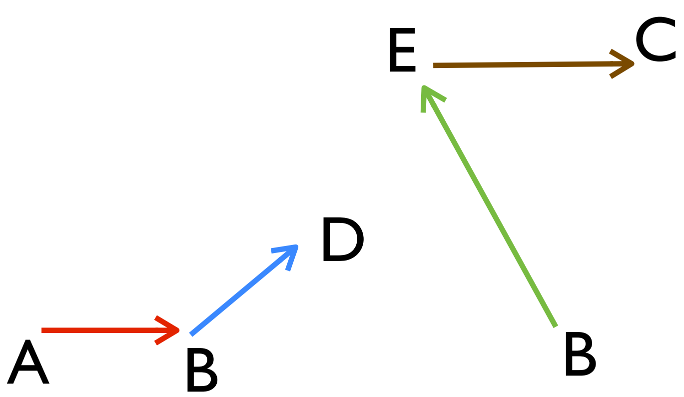
>Finding structures: Temporal motifs

Dynamical motifs

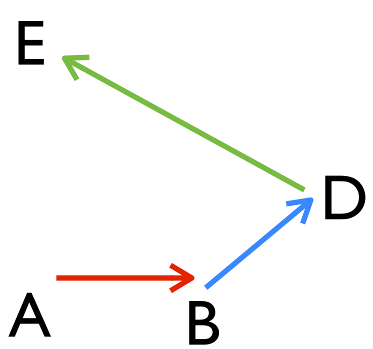


Bajardi et al, PLoS ONE (2011)

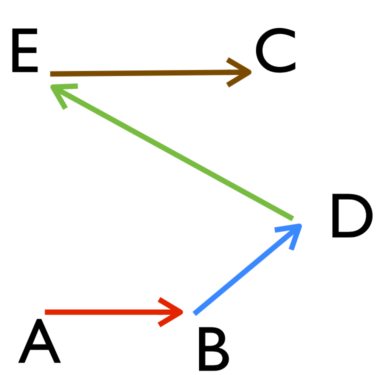
Dynamical motifs



length 2

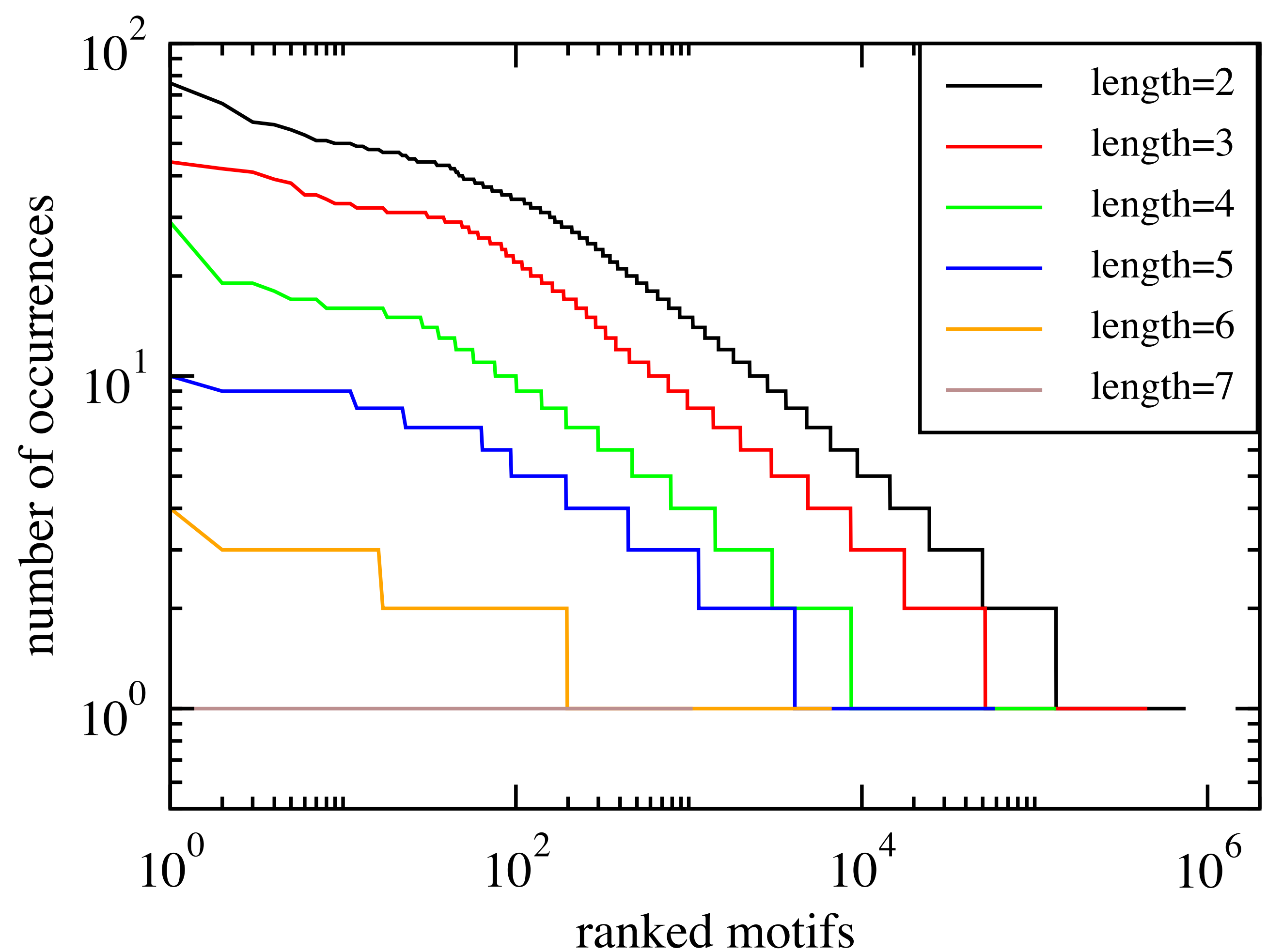


length 3



length 4

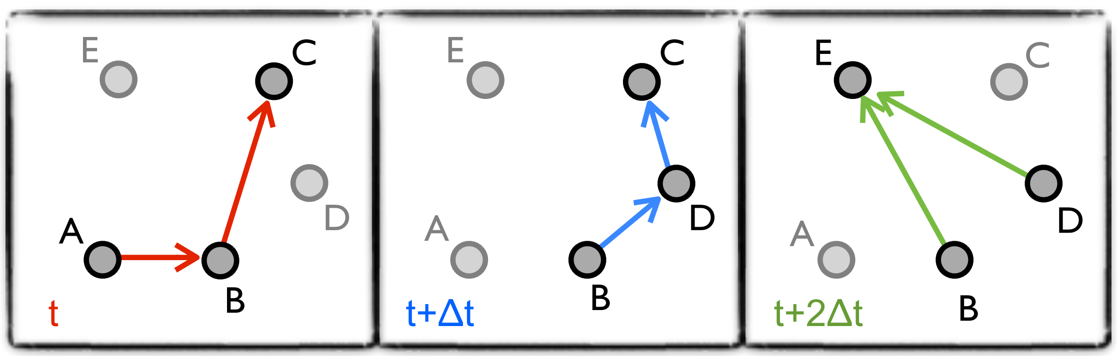
...



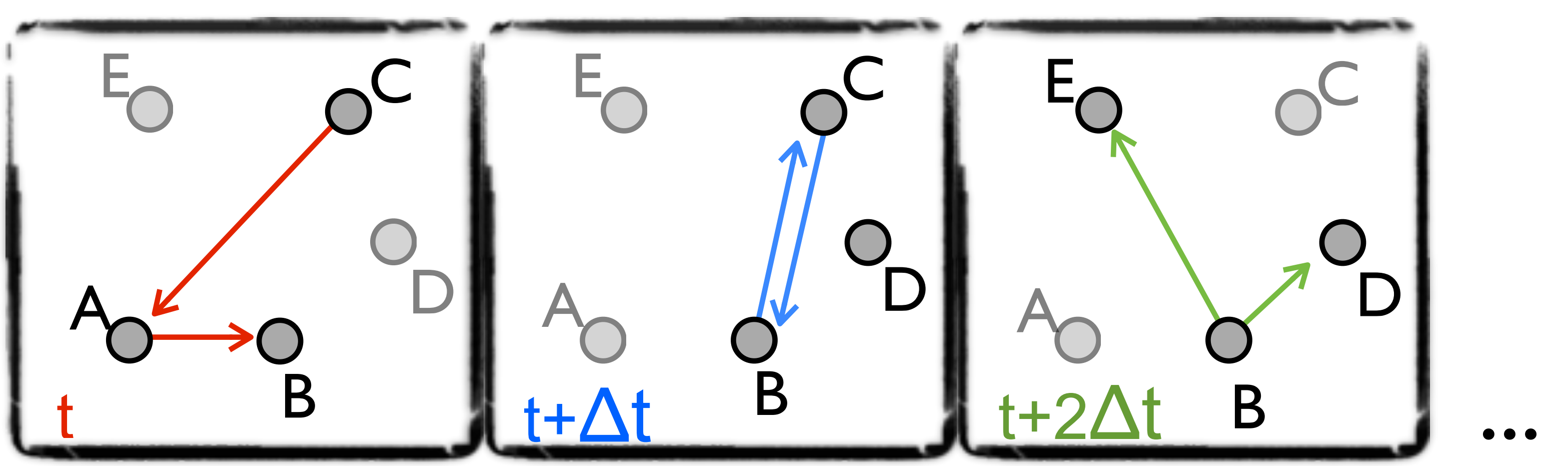
Bajardi et al, PLoS ONE (2011)

Dynamical motifs

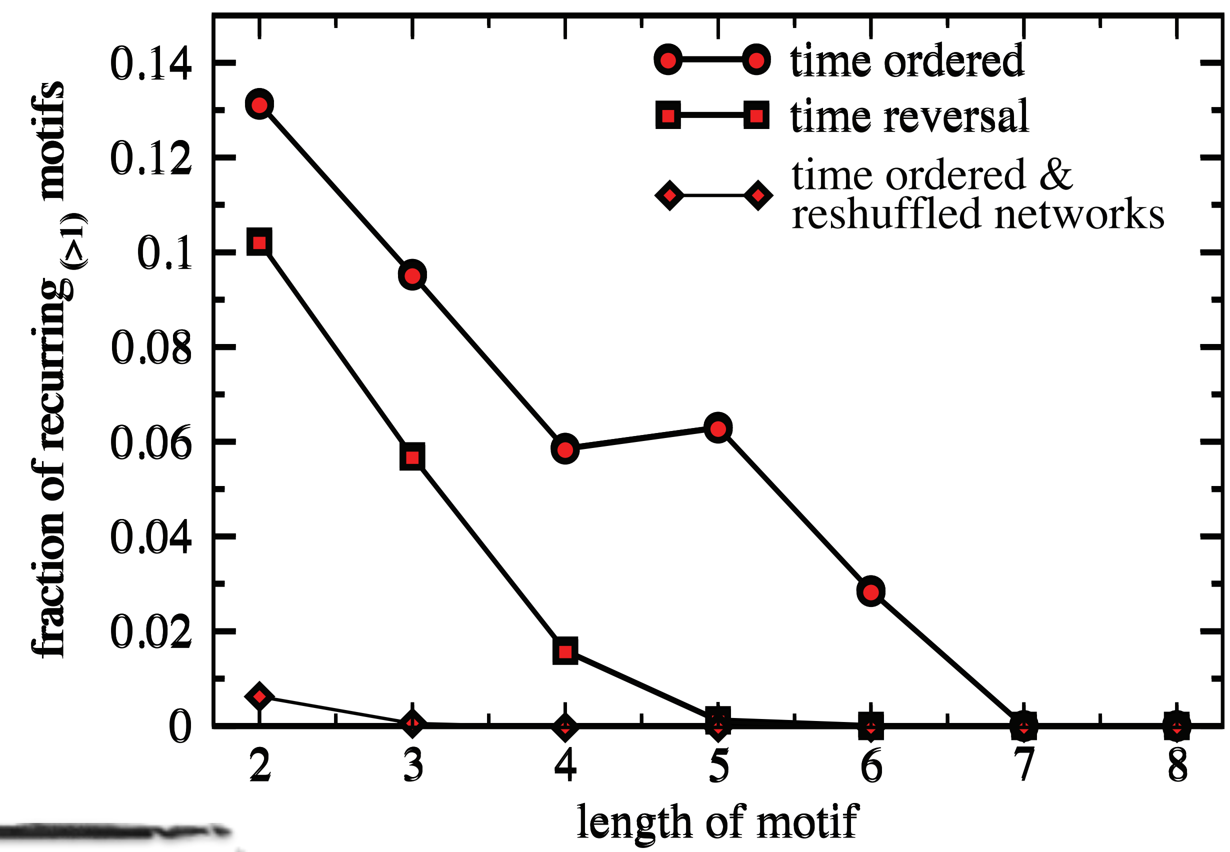
Original network:



Null model:
<u>reshuffled networks</u>



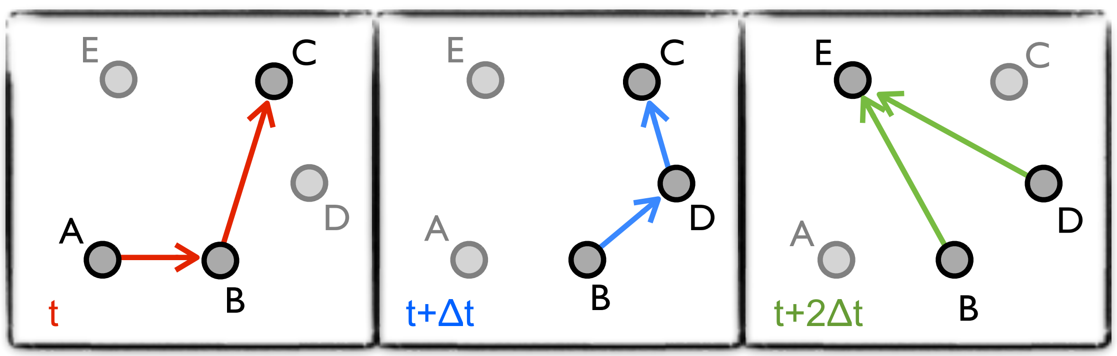
time

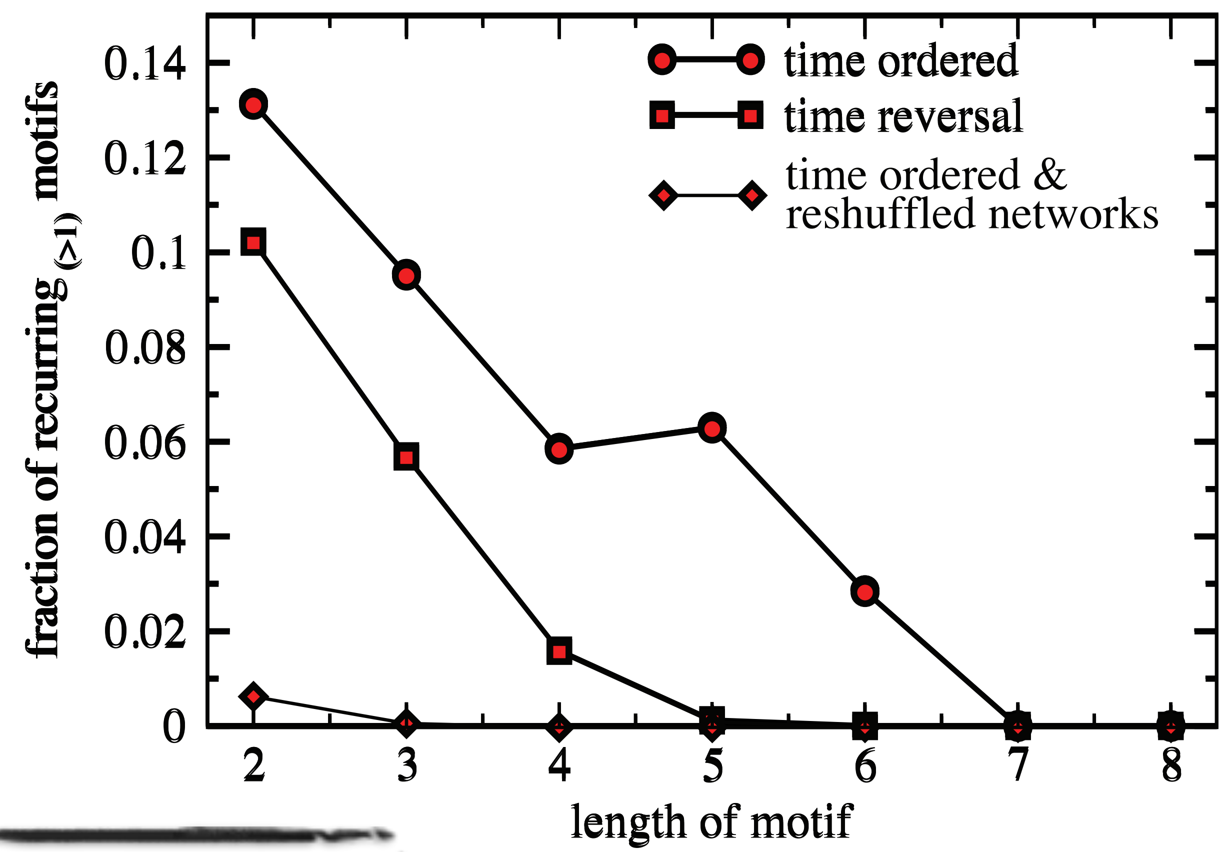


Bajardi et al, PLoS ONE (2011)

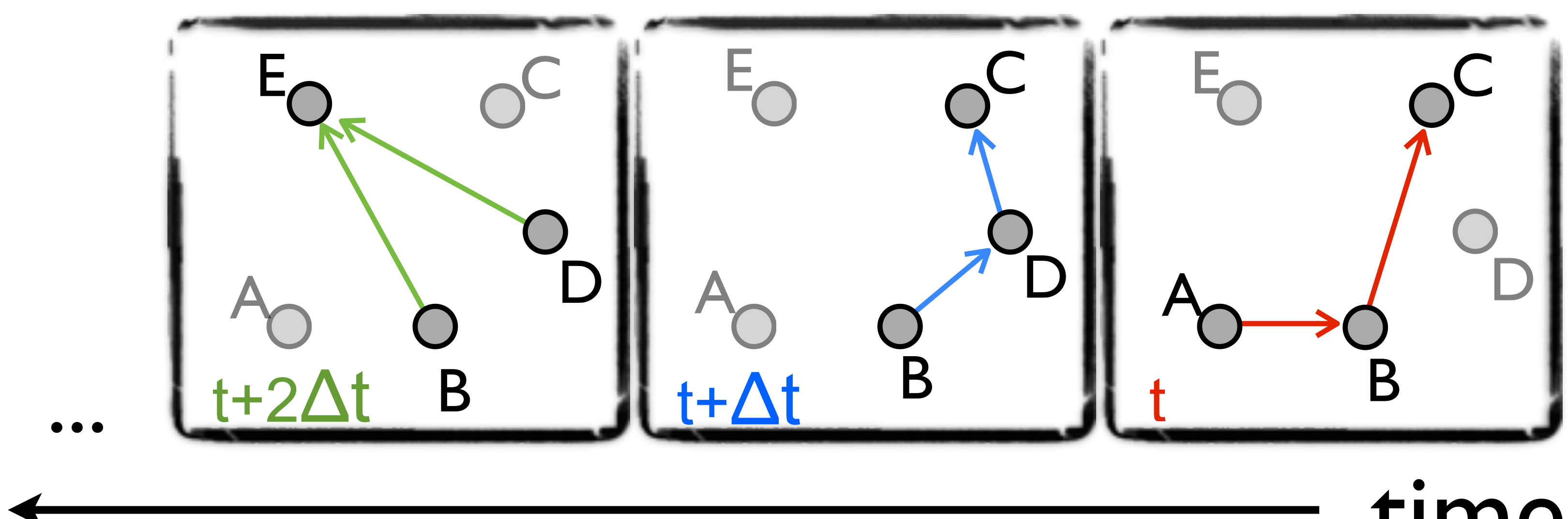
Dynamical motifs

Original network:





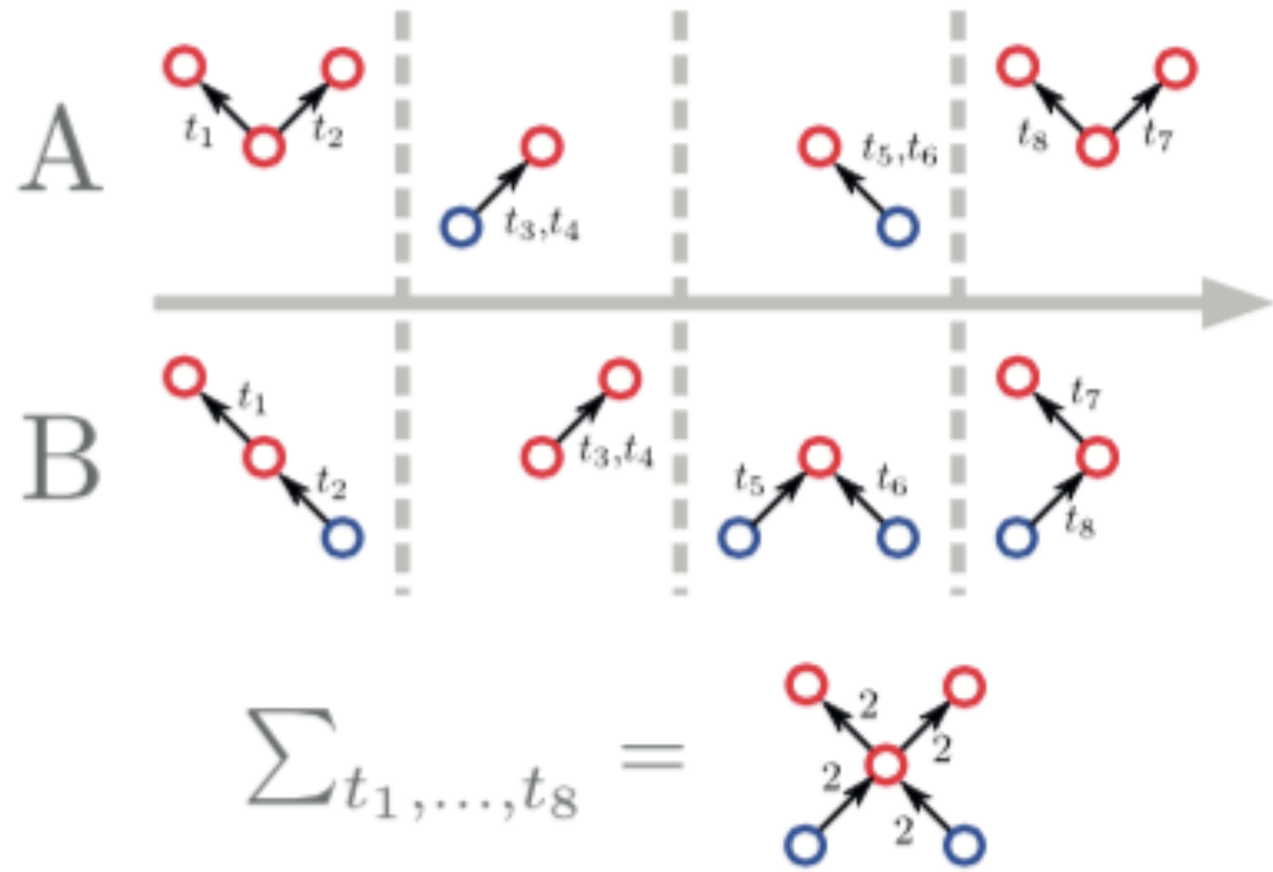
Null model:
<u>time reversal</u>



Bajardi et al, PLoS ONE (2011)

Temporal motifs

Same aggregated network, different temporal sequences



detection of temporal patterns?

L. Kovanen et al., J. Stat. Mech. (2011) P11005

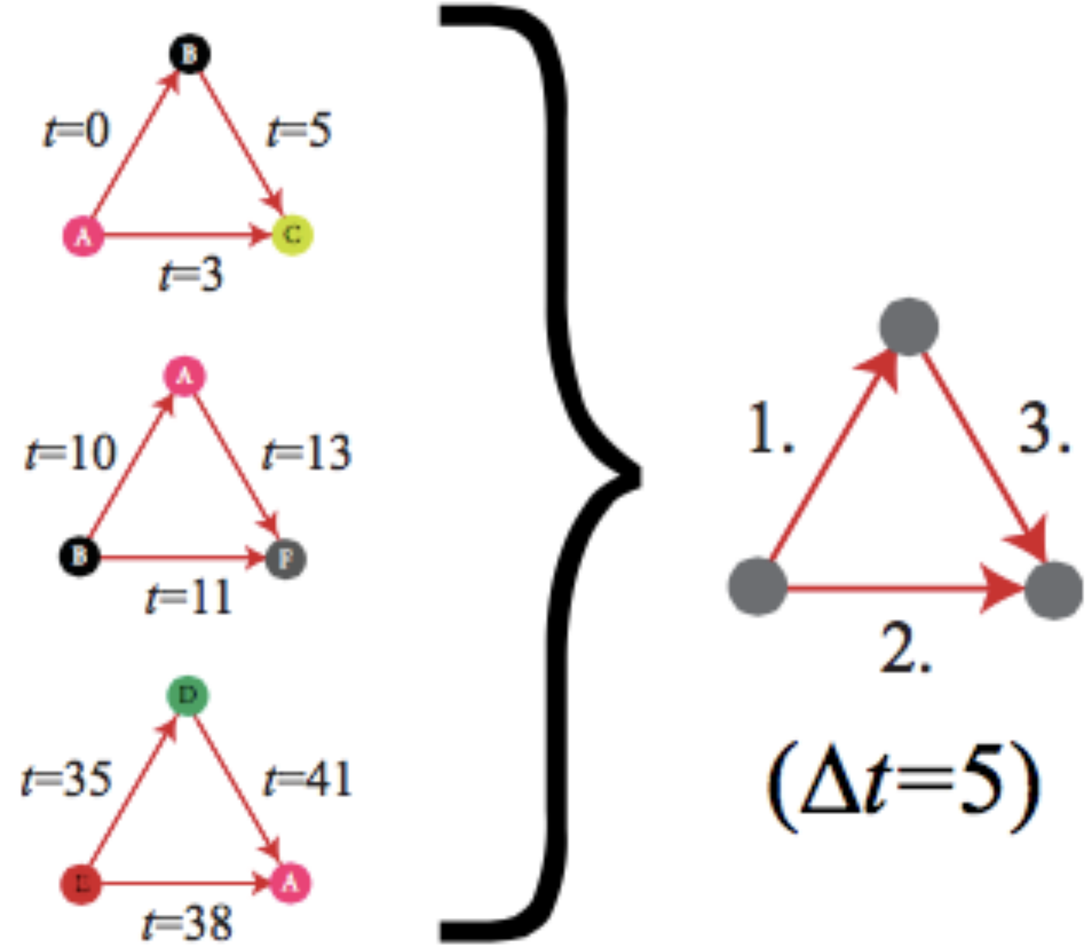
L. Kovanen et al., PNAS (2013)

Temporal motifs

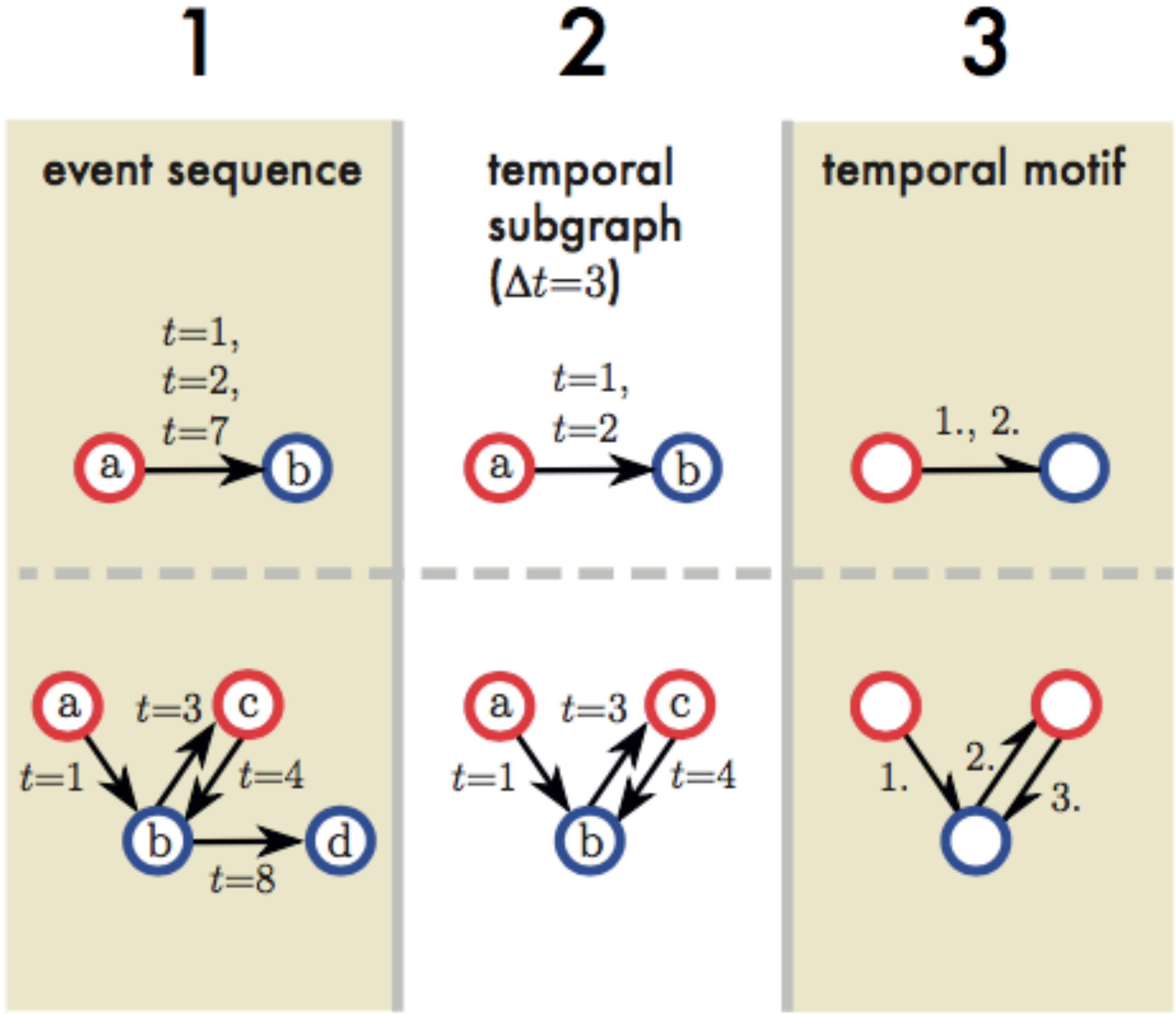
= Equivalence classes of valid temporal subgraphs

Valid: no events are skipped at each node

Equivalence classes: forget identities of nodes and exact timing



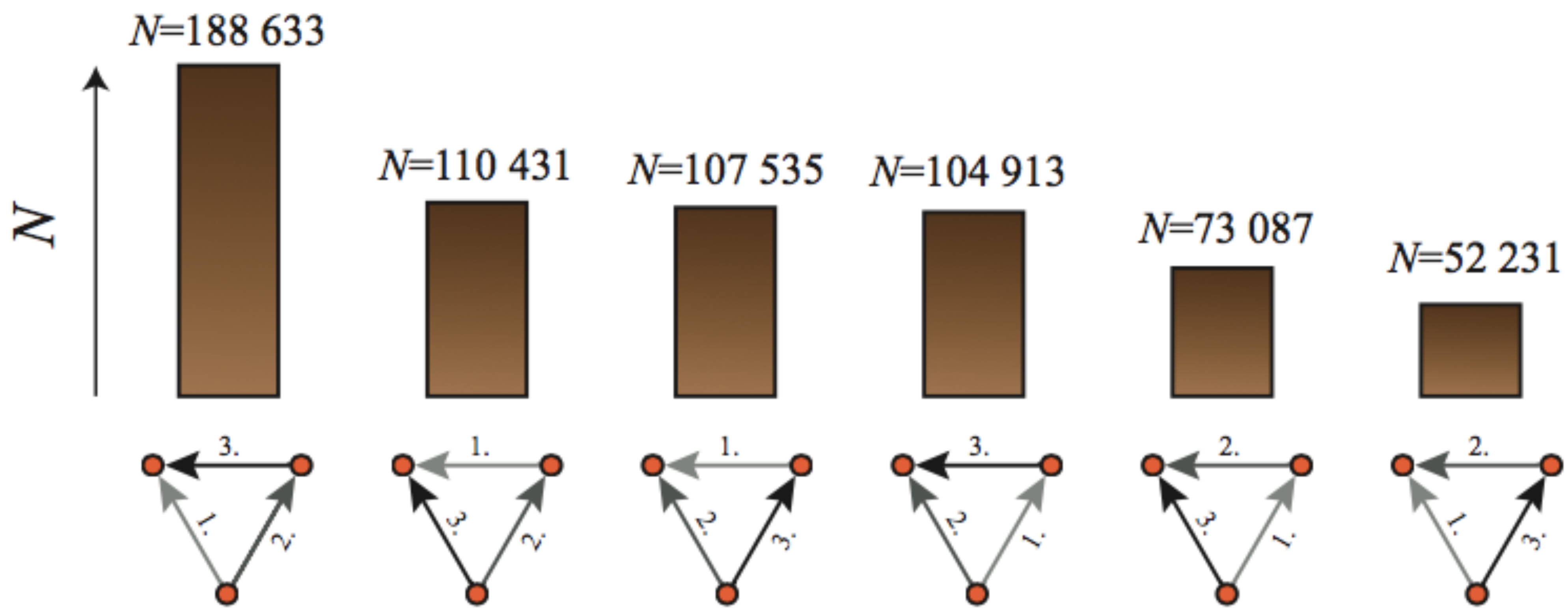
J. Saramäki



which motifs are most frequent?
metadata on nodes, compare with null models, …

Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences
L. Kovanen et al., PNAS (2013)

Structure: temporal motifs



Motifs in mobile phone call networks

Kovanen et al. (2012)

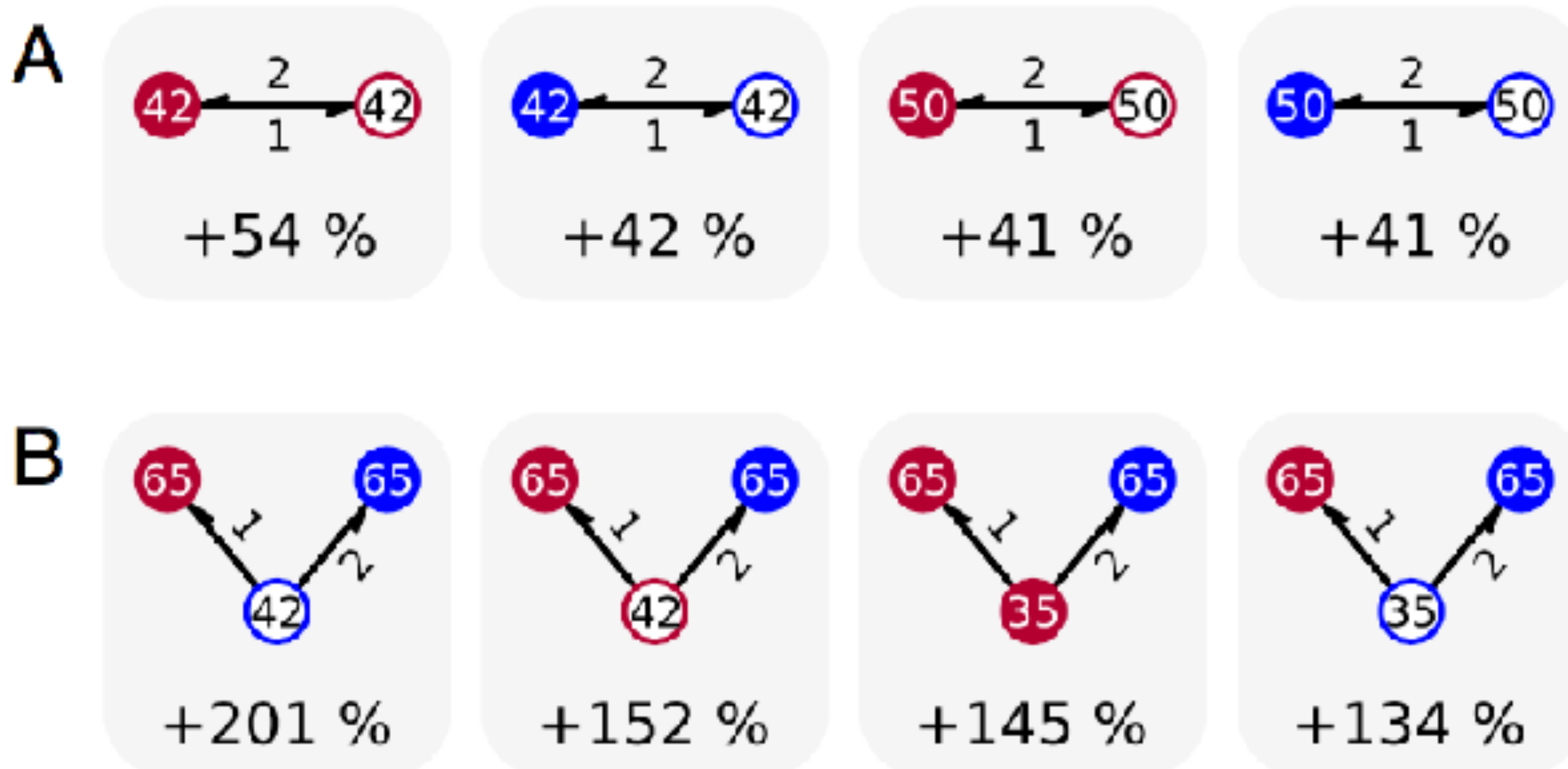


Fig. 4. The most common temporal motifs exhibit shared properties. (*A*) The four most common returned-call motifs. The numbers inside the nodes denote the age group (18–26, 27–32, 33–38, 39–45, 46–55, or 56–80; the value shown is the weighted average rounded to closest integer). The open nodes denote postpaid and filled prepaid customers; red denotes female, and blue, male. The arrows denote events, and the numbers next to them show their temporal order. In all four cases, the first call takes place from the prepaid (filled node) to the postpaid (open node) customer. The number below each motif shows the relative occurrence compared with the null model. (*B*) The four most common out-star motifs. In all four cases, the two receivers have the same age, a pattern that is typical for the most common out-stars.

Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences
L. Kovanen et al., PNAS (2013)

> Finding Structures:
decomposing a temporal network

Detection of structures: decomposition of temporal network

Data: temporal network with discrete time intervals

$\Updownarrow$

Time-dependent adjacency matrix A(i,j,t)

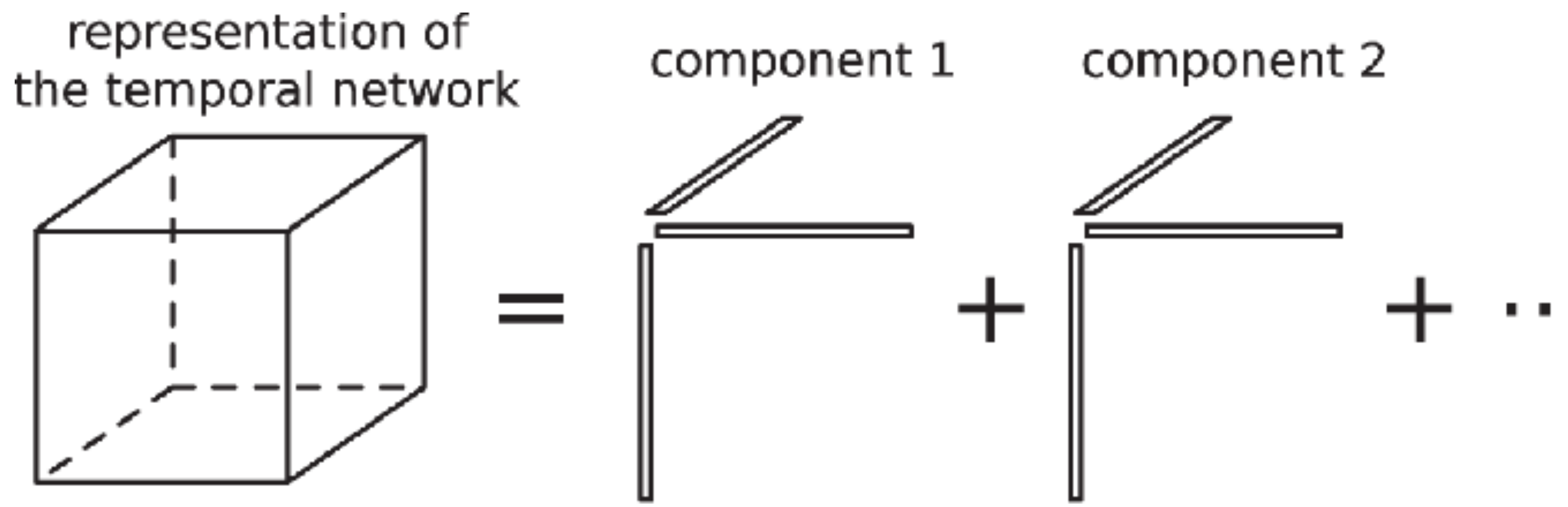
$\Updownarrow$

Three-way tensor T

L. Gauvin et al., PLoS ONE (2014)

Detection of structures: decomposition of temporal network

Kruskal decomposition



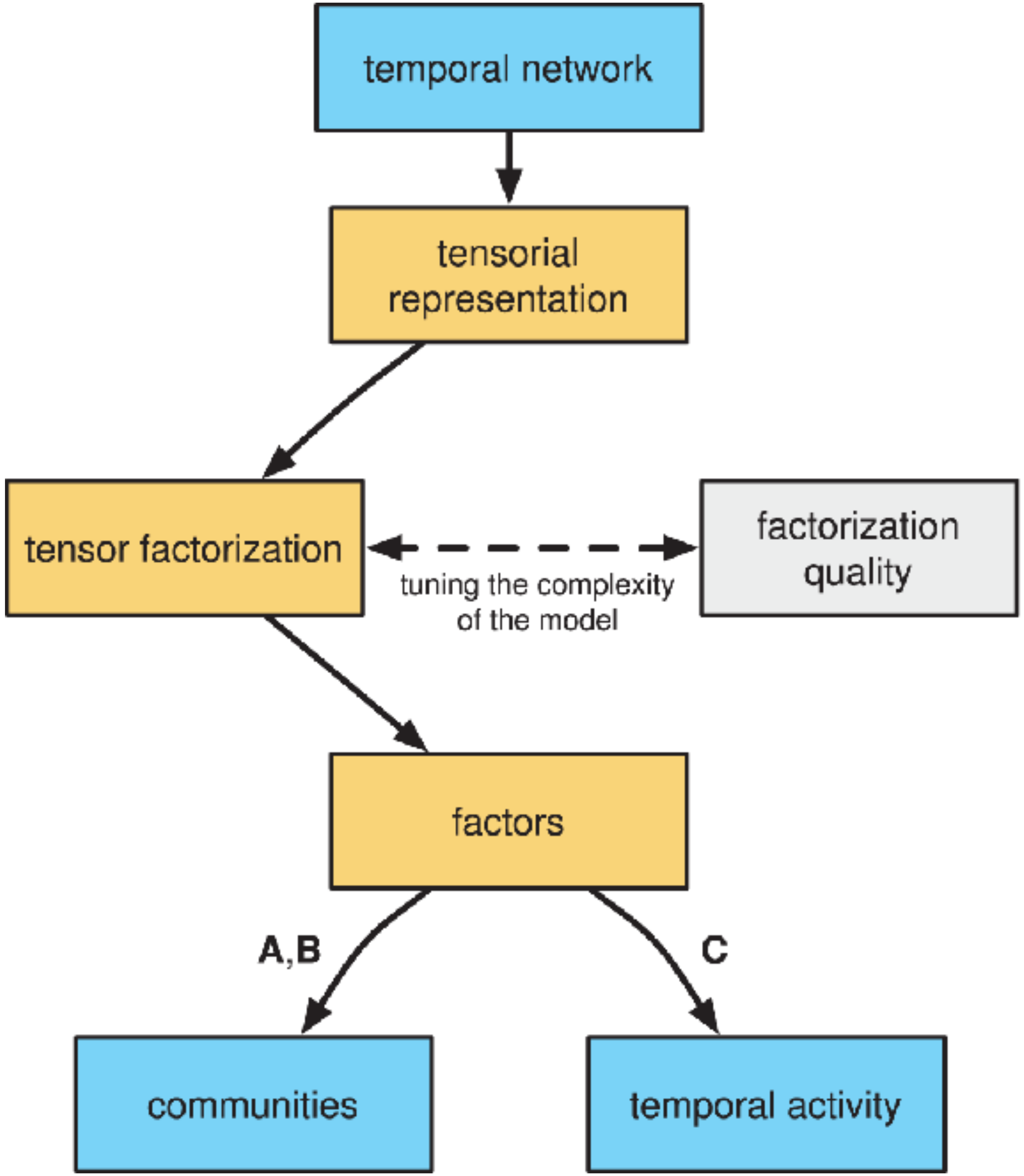
$$T \approx \tilde{\mathcal{T}} = \sum_{r=1}^{R} \mathcal{S}_r = \sum_{r=1}^{R} \mathbf{a_r} \circ \mathbf{b_r} \circ \mathbf{c_r}$$

$$t_{ijk} \approx \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}$$

undirected network: a=b

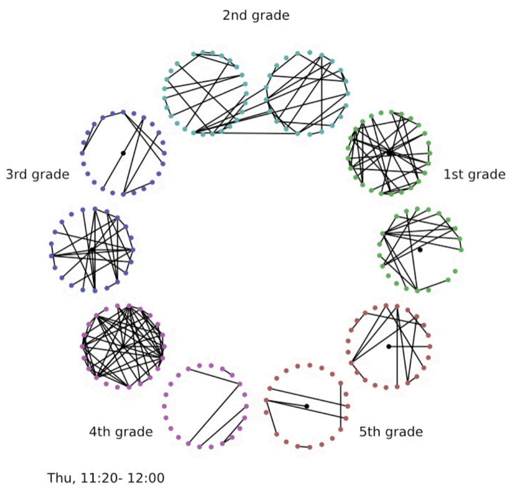
a_{ir}= membership of node i to component r
c_{kr}=timeline of component r

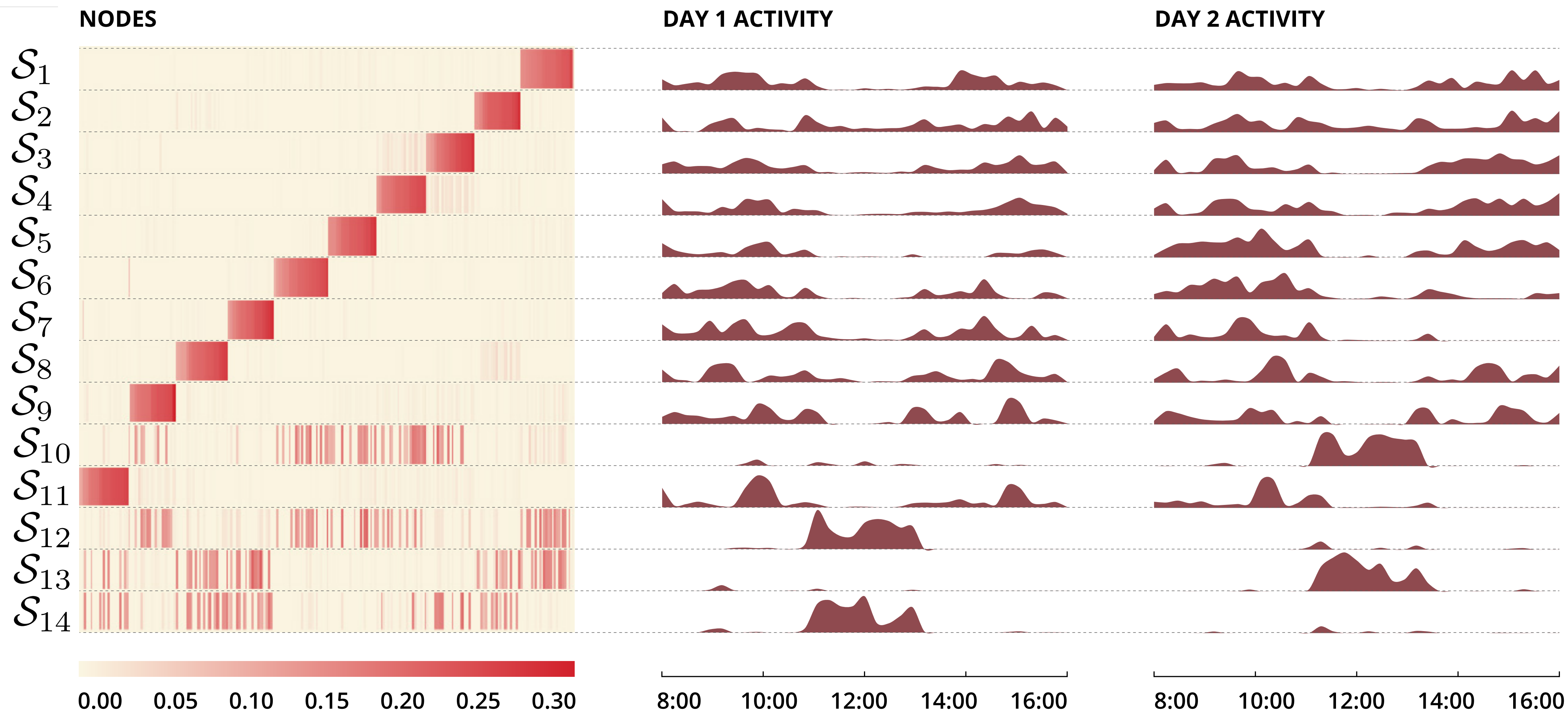
L. Gauvin et al., PLoS ONE (2014)



L. Gauvin et al., PLoS ONE (2014)

Primary school case study





10 classes
4 mixed-membership components = breaks

L. Gauvin et al., PLoS ONE (2014)

>Finding structures: Span cores

Reminder:
k-core decomposition for static networks

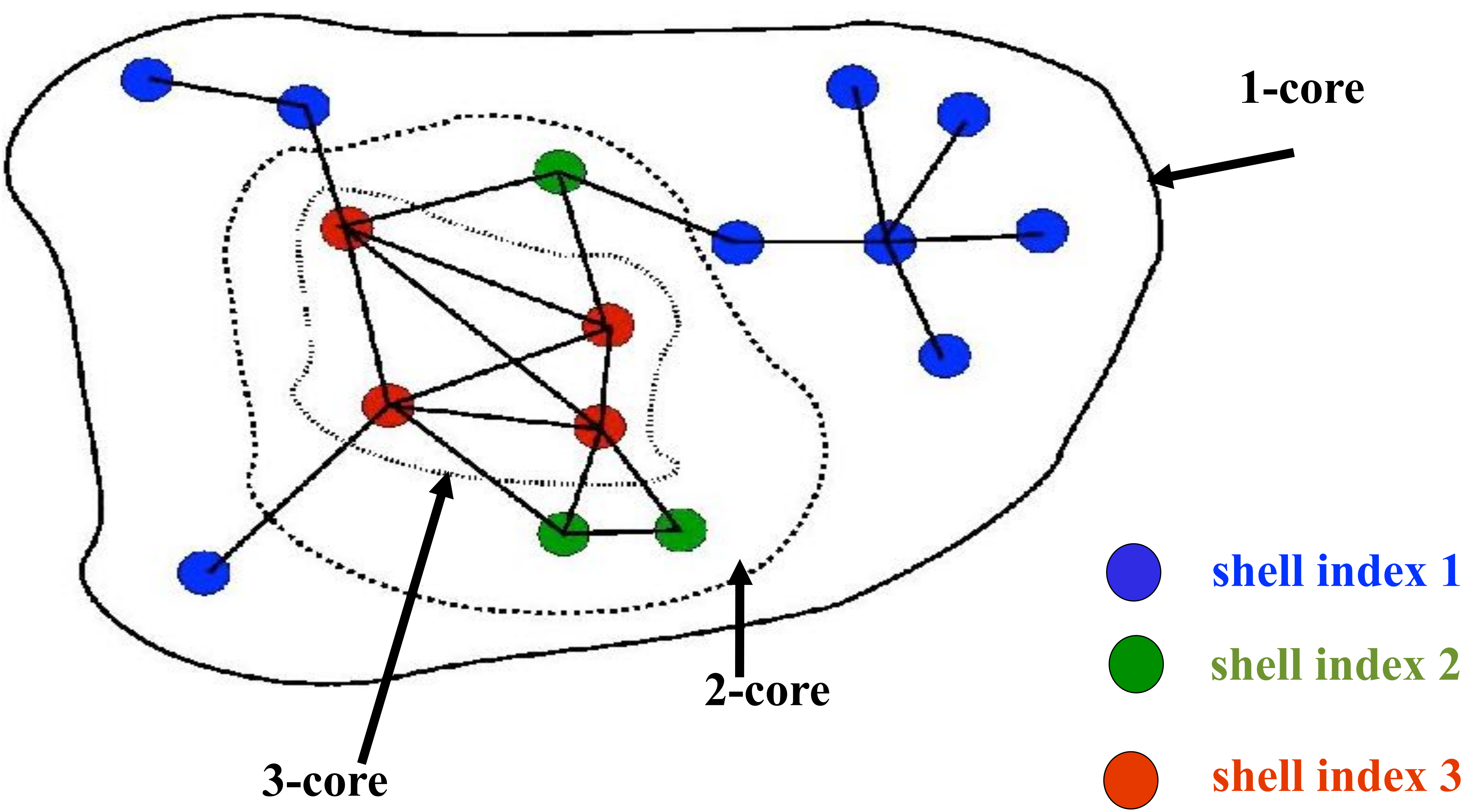
graph G=(V,E)

–k-core of graph G: maximal subgraph such that for all vertices in this subgraph have degree at least k

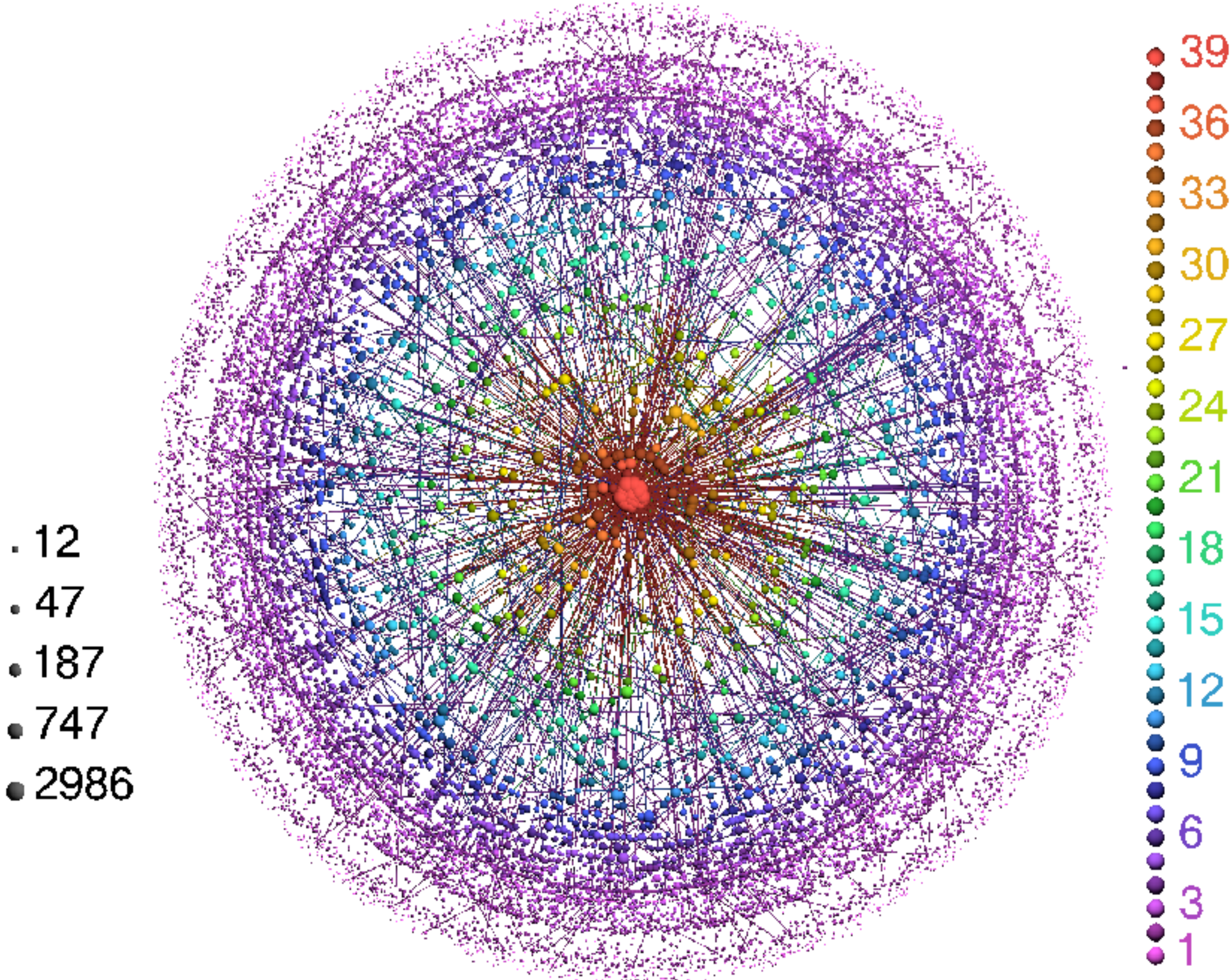
–vertex i has shell index k iff it belongs to the k-core but not to the (k+1)-core

–k-shell: ensemble of all nodes of shell index k

Example



http://lanet-vi.fi.uba.ar/



NB: role in spreading processes, cf Kitsak et al., Nat. Phys. (2010)

Span-core: definition

Temporal network G, set of vertices V,
temporal interval $T = [0,1,\ldots,t_{max}]$

Set of edges at time t: E_t

Set of edges active **at all times** t of an interval Δ : $E_\Delta = \bigcap_{t \in \Delta} E_t$

Temporal degree of a node within a subgraph S during Δ :

$$d_\Delta(S,u) = |\{v \in S \mid (u,v) \in E_\Delta[S]\}|$$

=number of nodes to which u is linked **at all times** during Δ

Galimberti, Barrat, Bonchi, Cattuto, Gullo, submitted (2018)

Span-core: definition

Definition 2 (((k, Δ)-core). *The* (k, Δ)-core *of a temporal graph* $G = (V, T, \tau)$ *is (when it exists) a maximal and non-empty set of vertices* $\emptyset \neq C_{k,\Delta} \subseteq V$, *such that* $\forall u \in C_{k,\Delta} : d_\Delta(C_{k,\Delta}, u) \geq k$, *where* $\Delta \sqsubseteq T$ *is a temporal interval and* $k \in \mathbb{N}^+$.

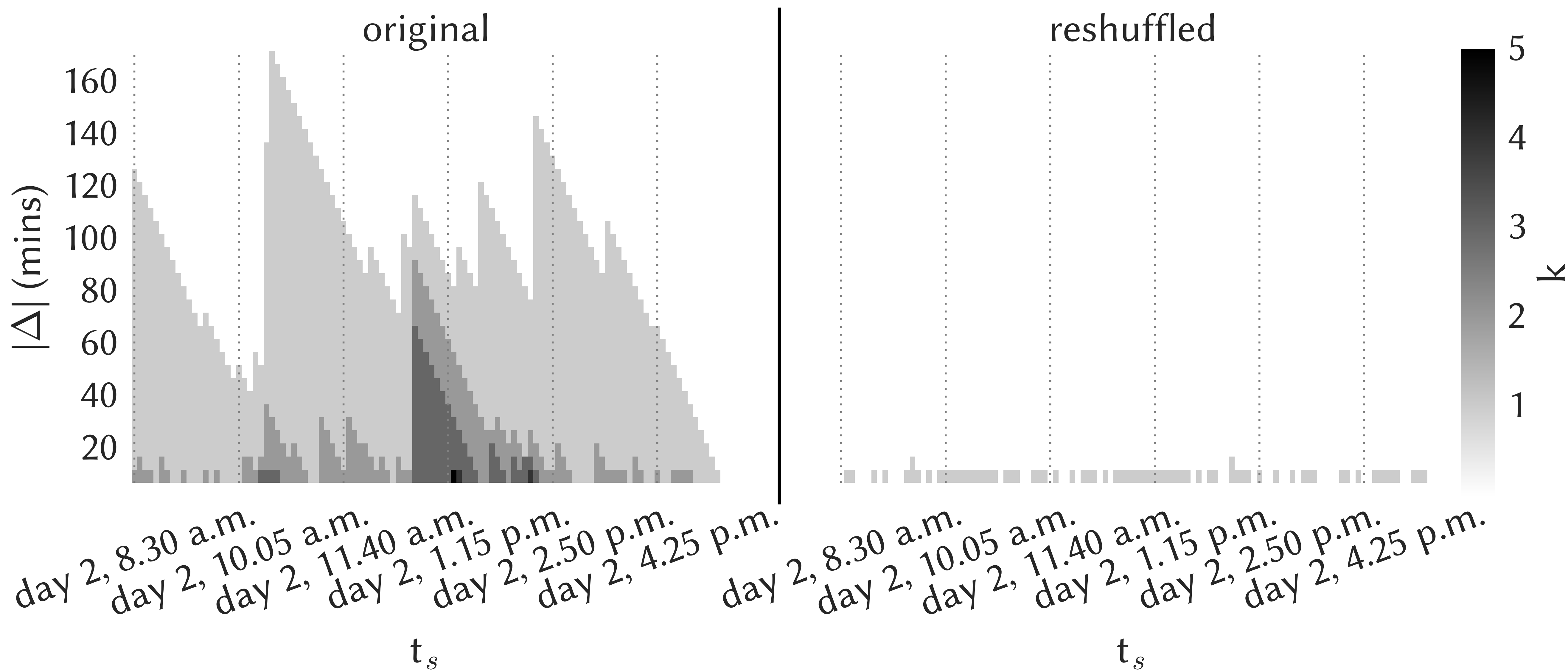
(all nodes of the core have degree at least k during the temporal interval, and have at least k "constant" neighbors during that interval)

Definition 3 (Maximal Span-core). *A span-core* $C_{k,\Delta}$ *of a temporal graph* G *is said* maximal *if there does not exist any other span-core* $C_{k',\Delta'}$ *of* G *such that* $k \leq k'$ *and* $\Delta \sqsubseteq \Delta'$.

Galimberti, Barrat, Bonchi, Cattuto, Gullo, submitted (2018)

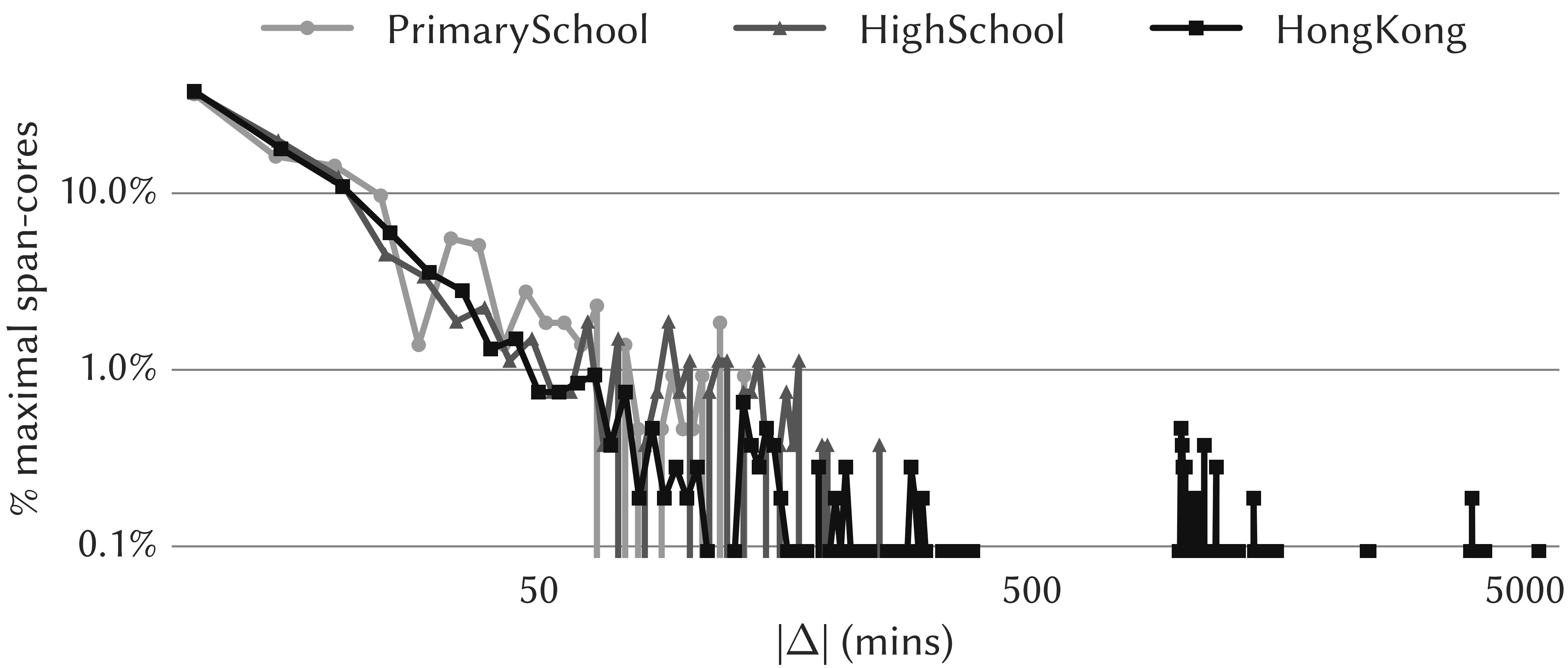
Span-core: examples

Primary school data:
Order of the span-cores as a function of their starting time and of the temporal span length



Galimberti, Barrat, Bonchi, Cattuto, Gullo, submitted (2018)

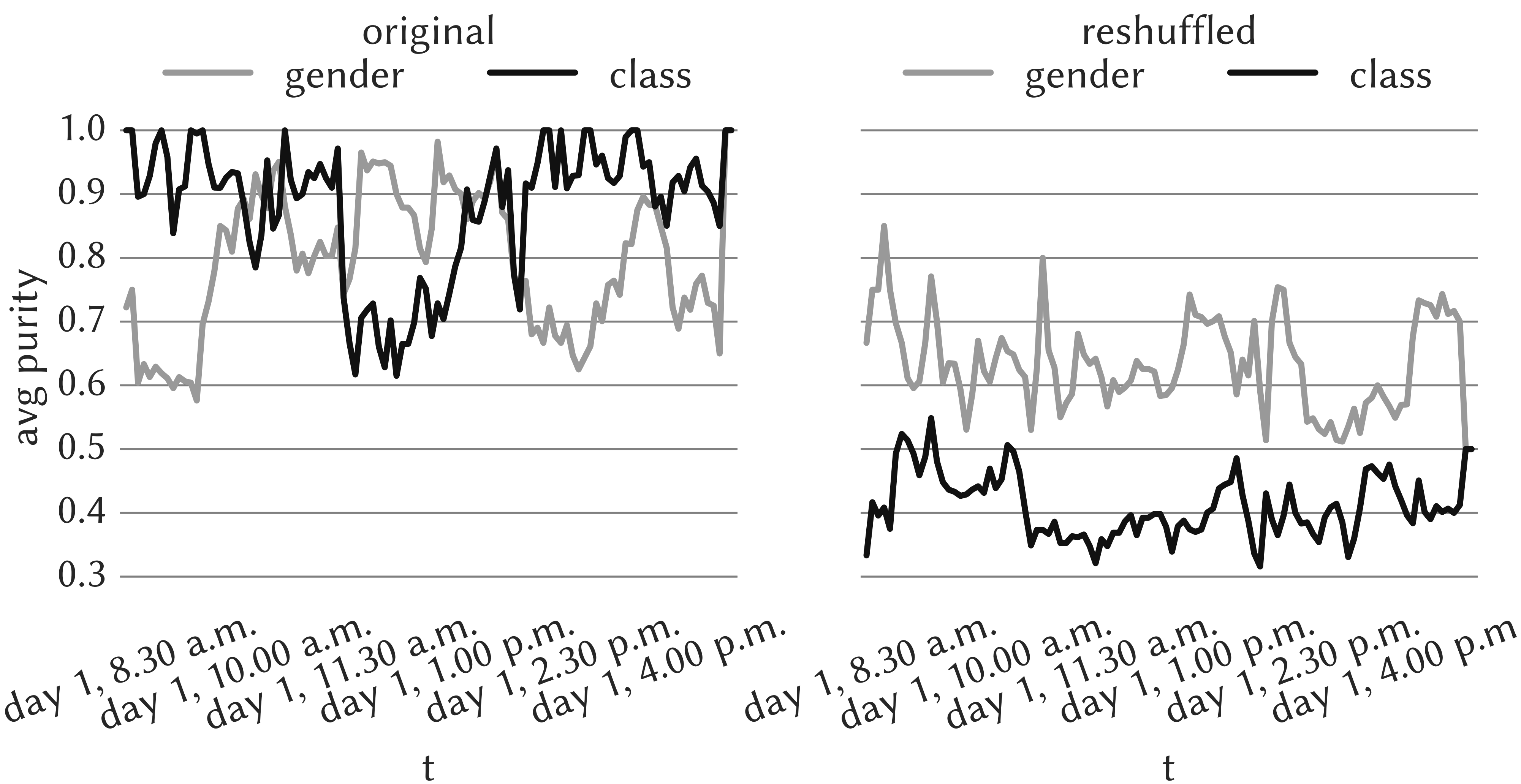
Span-core: examples



Distributions of maximal span-cores lengths

Galimberti, Barrat, Bonchi, Cattuto, Gullo, submitted (2018)

Span-core: examples

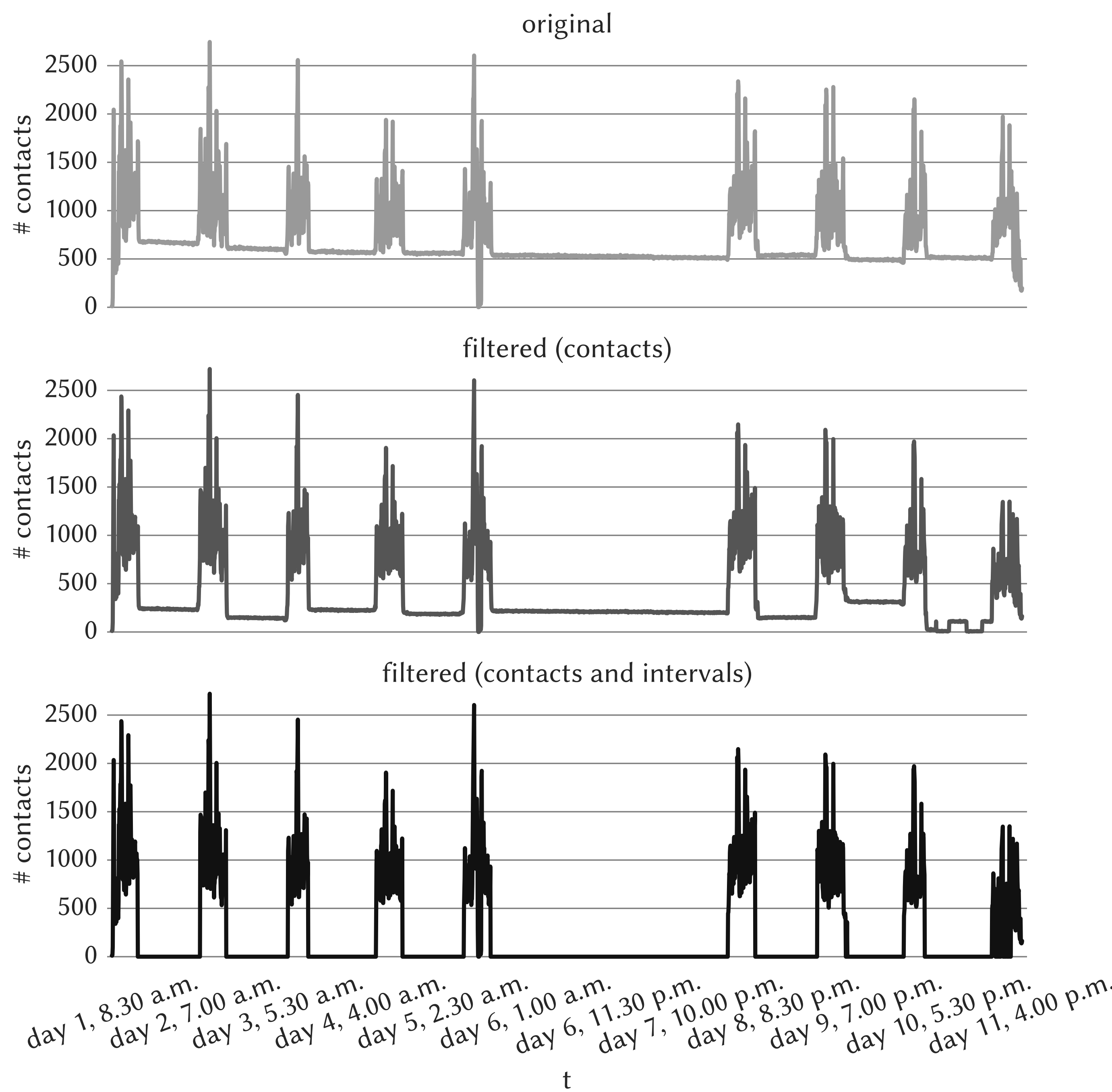


Primary school:
Average gender and class purity of the maximal span-cores

Galimberti, Barrat, Bonchi, Cattuto, Gullo, submitted (2018)

Span-core: anomaly detection

Hong-Kong primary school: filtering anomalous contacts



Galimberti, Barrat, Bonchi, Cattuto, Gullo, submitted (2018)

>Finding structures:
Backbones - significant ties

Filtering methods - static networks

Large scale data: mix of “important” and “random/noisy” links

How to extract the most relevant ties?

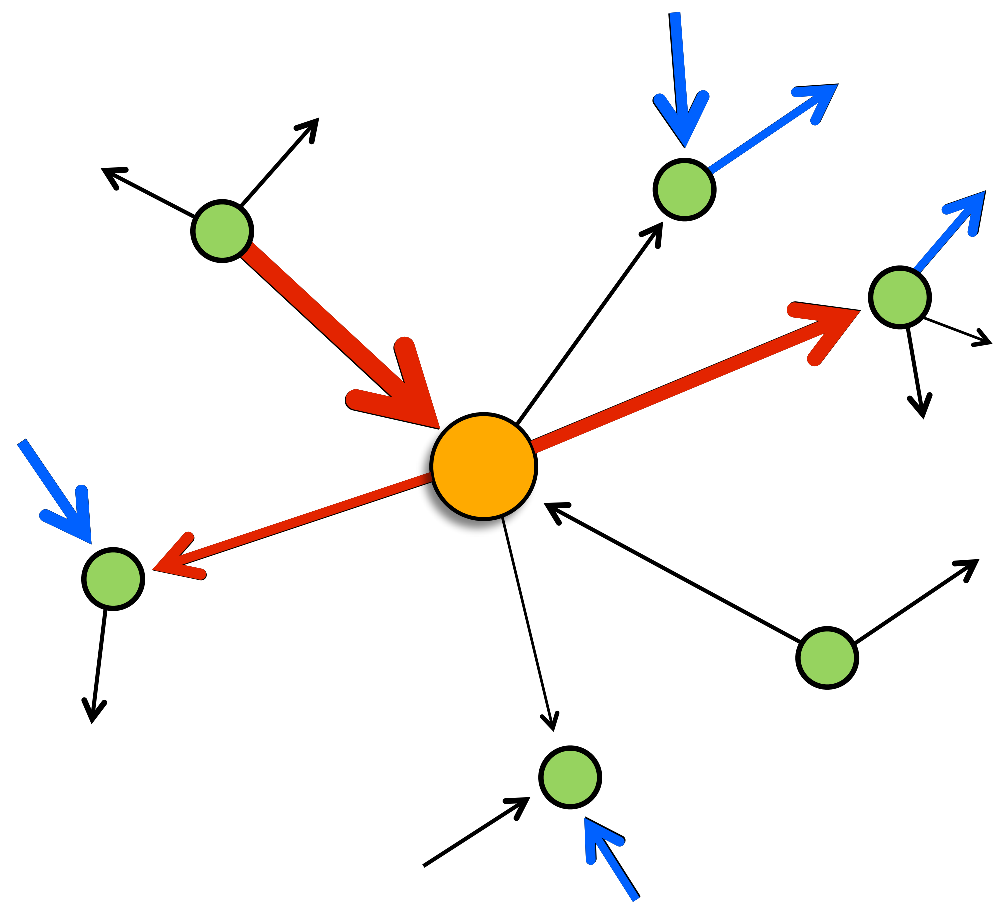
Filtering methods - static networks

- Retaining only large weights: links with weights larger than a (tunable) threshold => **global filter**

- Filters **based on a null model**: retain only weights larger than expected in a null model
    - **Disparity filter**: multiscale backbone (Serrano et al., PNAS 2009): for each node, retain edges that are “important for that node” => local filter

    -

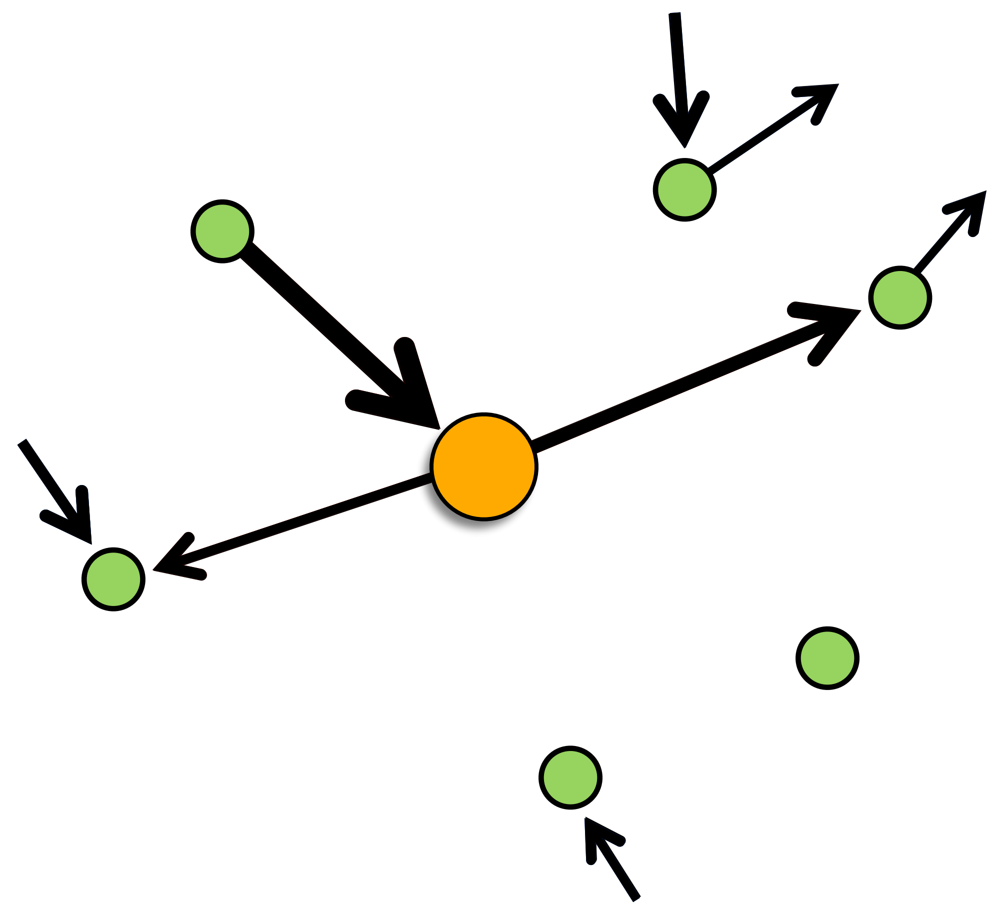
Disparity filter



The <u>disparity filter</u> selects the statistically significant links when the weights are heterogeneously distributed.

Serrano, Boguna, Vespignani - PNAS 2009

Disparity filter



The <u>disparity filter</u> selects the statistically significant links when the weights are heterogeneously distributed.

Serrano, Boguna, Vespignani - PNAS 2009

Filtering methods - static networks

- Retaining only large weights: links with weights larger than a (tunable) threshold => **global filter**

- Filters **based on a null model**: retain only weights larger than expected in a null model
    - **Disparity filter**: multiscale backbone (Serrano et al., PNAS 2009): for each node, retain edges that are “important for that node” => local filter

    - **Enhanced configuration model** (Gemmetto et al., arXiv: 1706.00230) = max entropy ensemble of networks with fixed degree and strength sequences => **p-value of observing a weight w_{ij} between i and j larger than or equal to the real weight** => retain only links with small p-value

Filtering method - temporal network

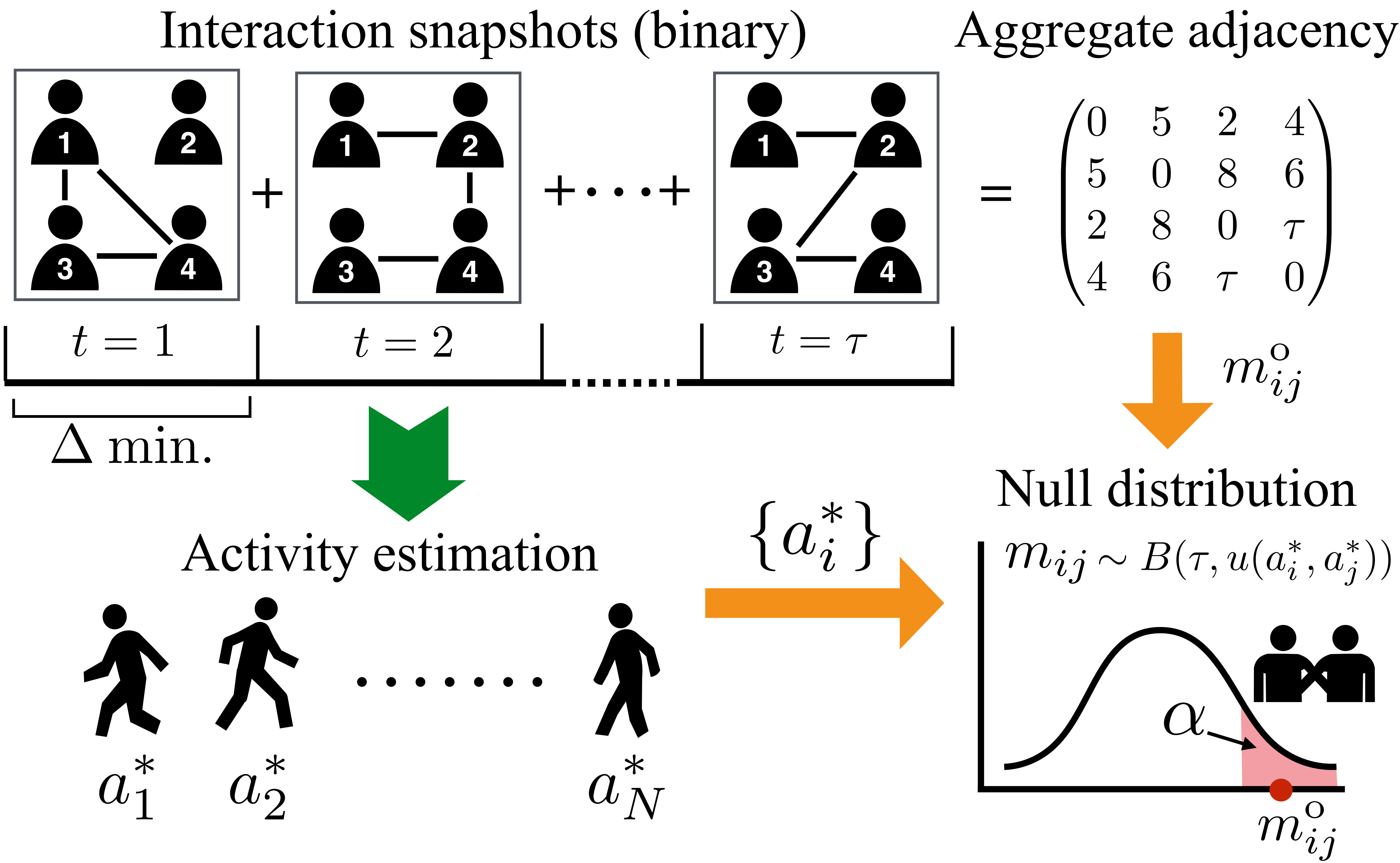
- Aggregation on temporal window and apply one of the static filter method (large weights, disparity filter, enhanced configuration model)

- Use a **temporal null model**:
    - define the temporal fitness model
        - node i, activity a_i
        - probability of interaction at any time: $u(a_i, a_j) = a_i a_j$
    - evaluate activities a_i^* of nodes (max likelihood)
    - compute the probability distribution of the number of interactions

$$g(m_{ij} | a_i^*, a_j^*) = \binom{\tau}{m_{ij}} u(a_i^*, a_j^*)^{m_{ij}} (1 - u(a_i^*, a_j^*))^{\tau - m_{ij}}.$$

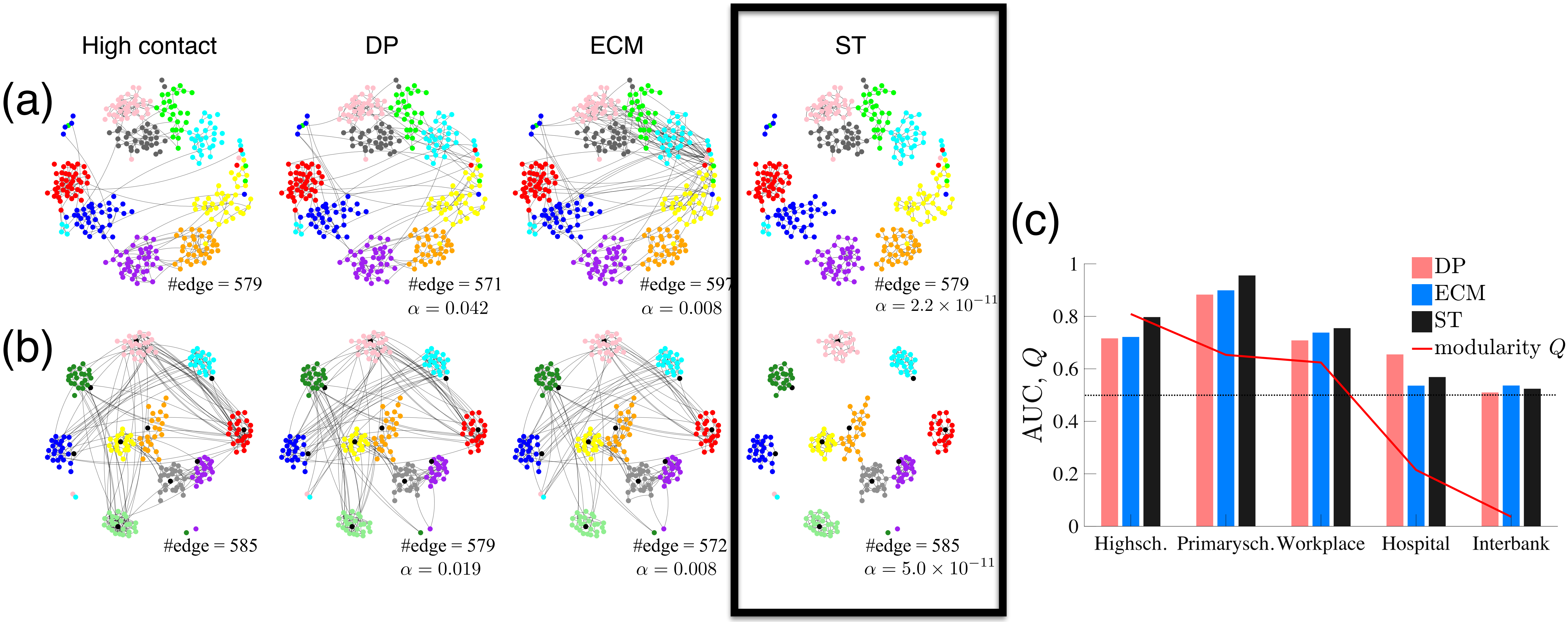
    - **ij is a significant tie if the observed number has a small p-value w.r.t.** g

T. Kobayashi, T. Takaguchi, A. Barrat, arXiv:1804.08828



T. Kobayashi, T. Takaguchi, A. Barrat, arXiv:1804.08828

Example: SocioPatterns datasets



- More ties detected than with other methods
- Significant ties seem to be mostly within groups, when a community structure is present

T. Kobayashi, T. Takaguchi, A. Barrat, arXiv:1804.08828

Beyond significant ties: significant structures

Probability that i,j,k form a triangle of **simultaneous** interactions at a given time

$$v(i,j,k) = u(a_i^*, a_j^*) \cdot u(a_j^*, a_k^*) \cdot u(a_k^*, a_i^*) \,,$$

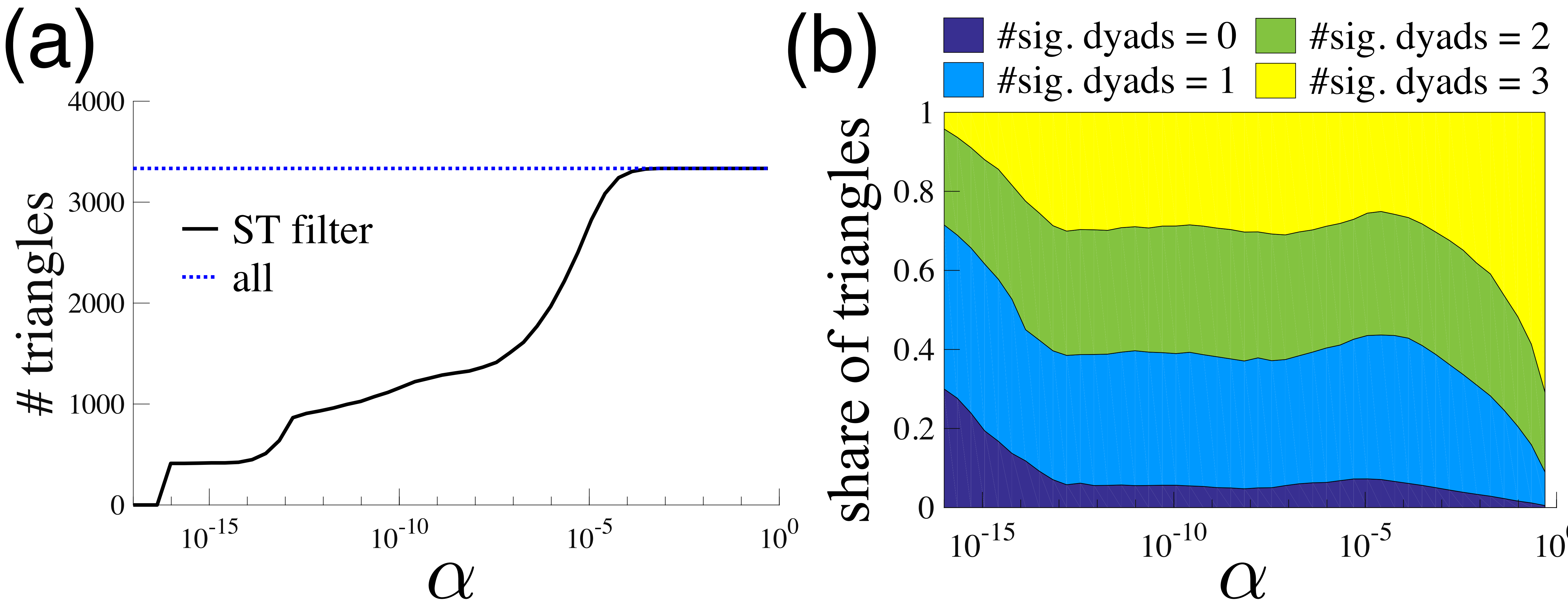
Probability that the triangle i,j,k is observed r_{ijk} times in the null model:

$$h(r_{ijk} \,|\, a_i^*, a_j^*, a_k^*) = \binom{\tau}{r_{ijk}} v(i,j,k)^{r_{ijk}} (1 - v(i,j,k))^{\tau - r_{ijk}}$$

=> possibility to **assign a significance to a triad of simultaneous interactions** or to any other temporal motif/structure
(not possible with static filters)

T. Kobayashi, T. Takaguchi, A. Barrat, arXiv:1804.08828

Beyond significant ties: significant structures



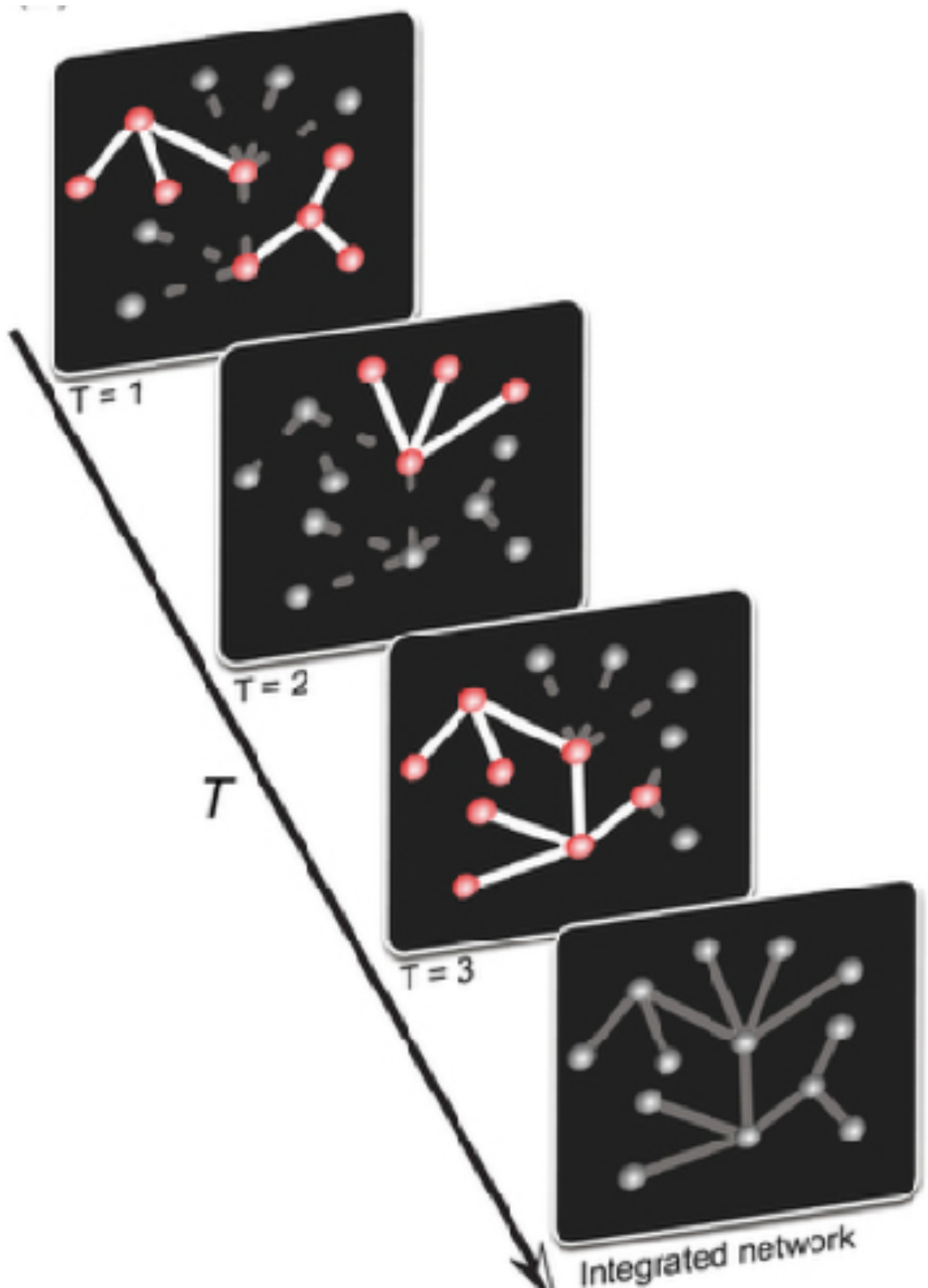
T. Kobayashi, T. Takaguchi, A. Barrat, arXiv:1804.08828

>Models of temporal networks

> activity-driven model

Activity-driven network

Model: N nodes, each with an "**activity**" a, taken from a distribution F(a)



At each time step:

- node i active with probability a(i)
- each active node generate m links to other randomly chosen nodes
- iterate with no memory

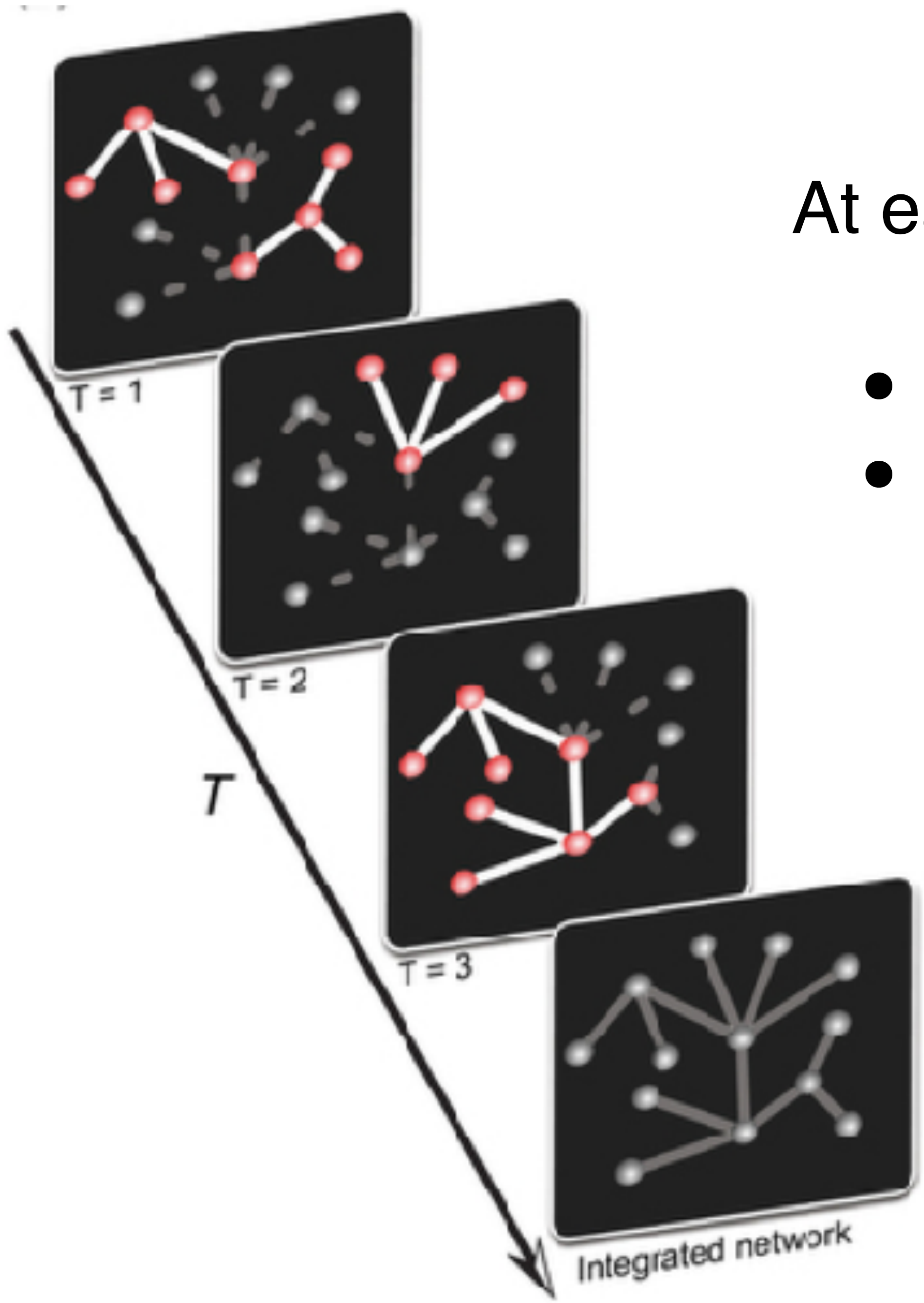
Aggregate degree distribution ~ F

No memory, no correlations…
"Toy" model allowing for analytical computations

Perra et al., Sci. Rep. (2012)

Activity-driven network **with attractiveness**

Model: N nodes, each with an "**activity**" a, taken from a distribution F(a) and an "**attractiveness**" b, from another distribution G(b)



At each time step:

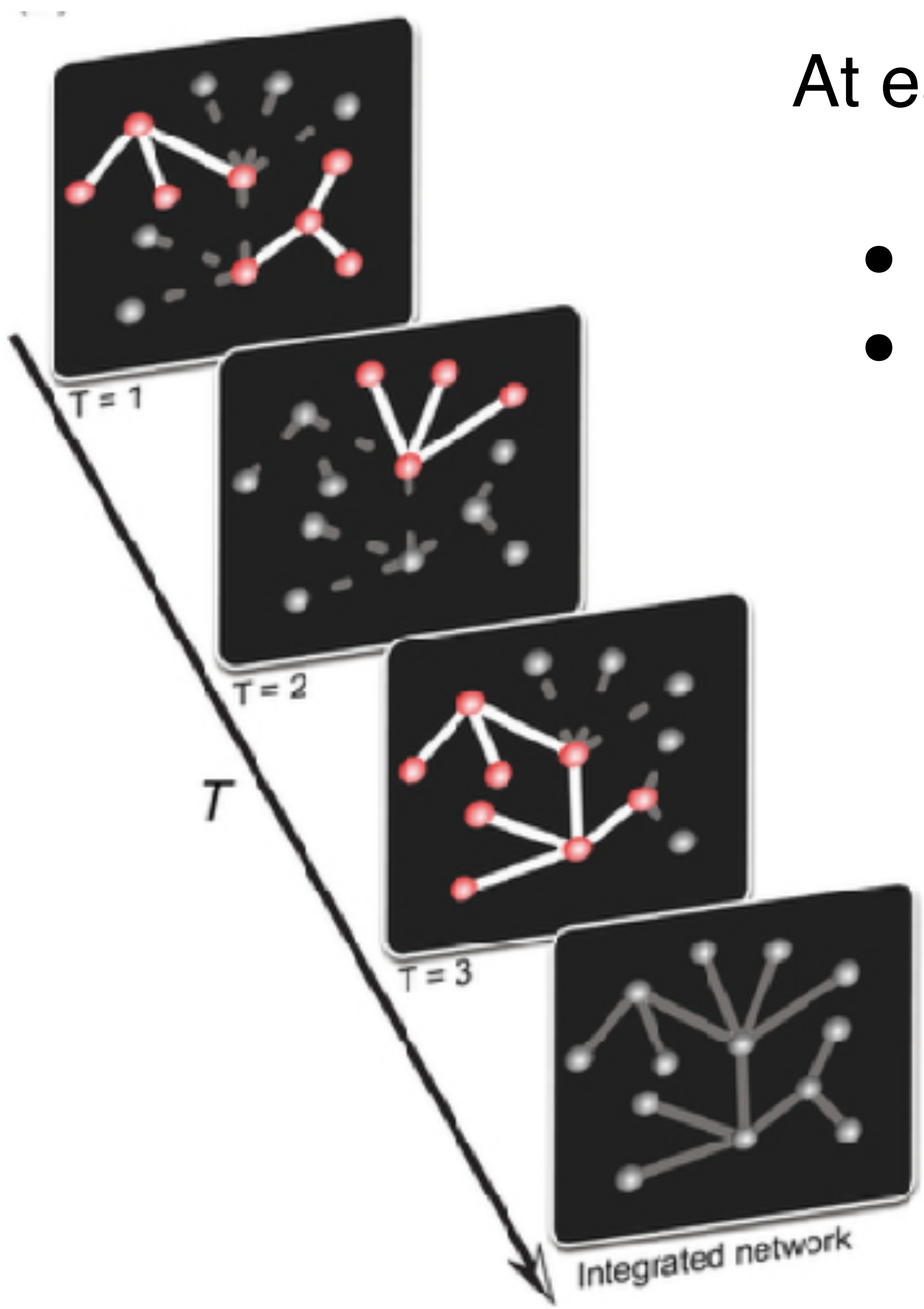
- each node i becomes active with probability a(i)
- each active node generate m links to other nodes chosen with probability proportional to their "attractiveness" b

Possibility to tune F, G and potential correlations between a and b

Alessandretti et al., Phys. Rev. (2017)

Activity-driven network **with memory**

Model: N nodes, each with an “activity” a, taken from a distribution F(a)



At each time step:

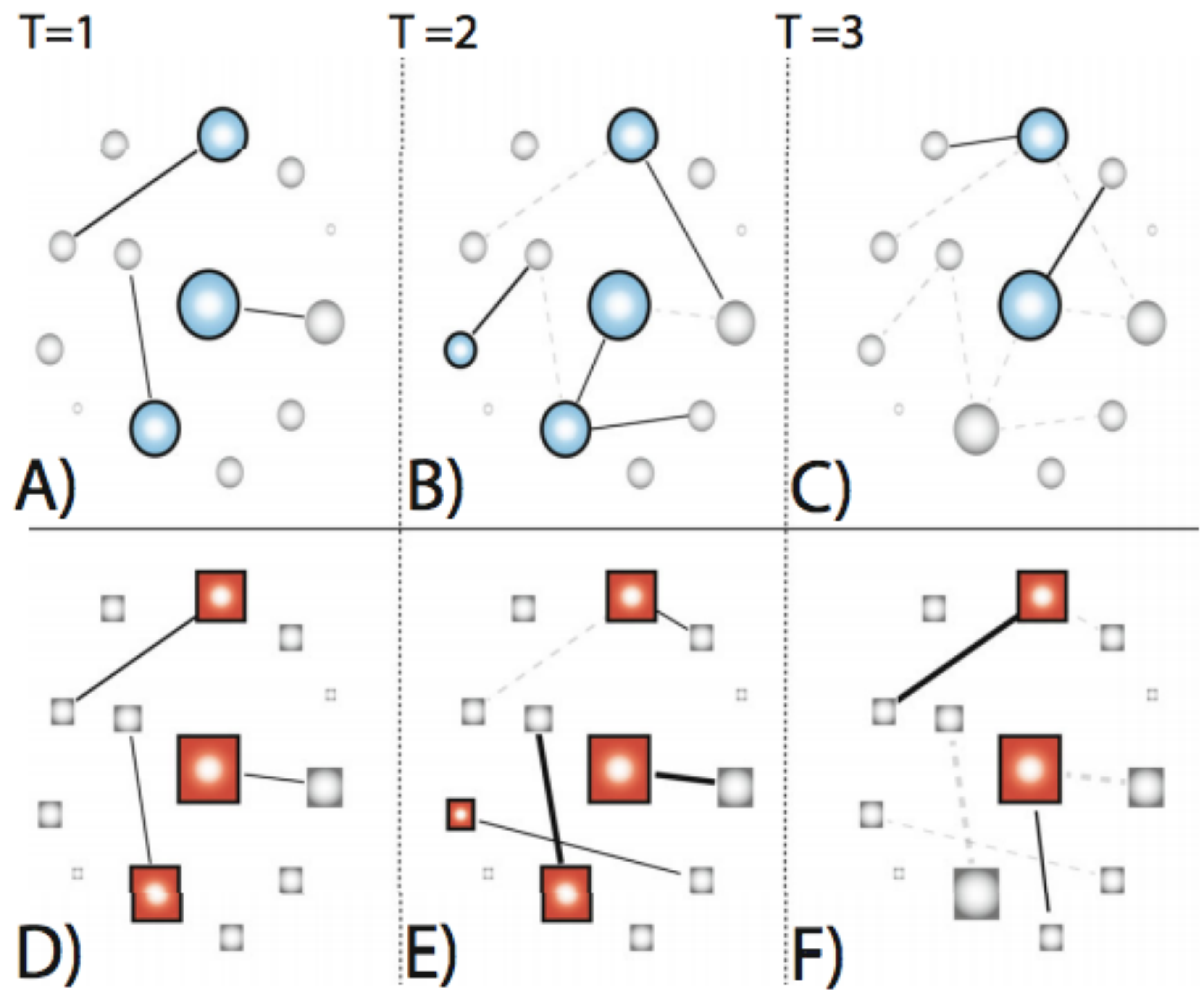
- each node i becomes active with probability a(i)
- each active node generate m links to other nodes; if it has already been in contact with n distinct nodes, the new link is
  - towards a new node with proba p(n)=c/(c+n)
  - towards an already contacted node with proba 1-p(n)

(reinforcement mechanism seen in data)

Karsai et al., Sci. Rep. (2014)

Memory-less

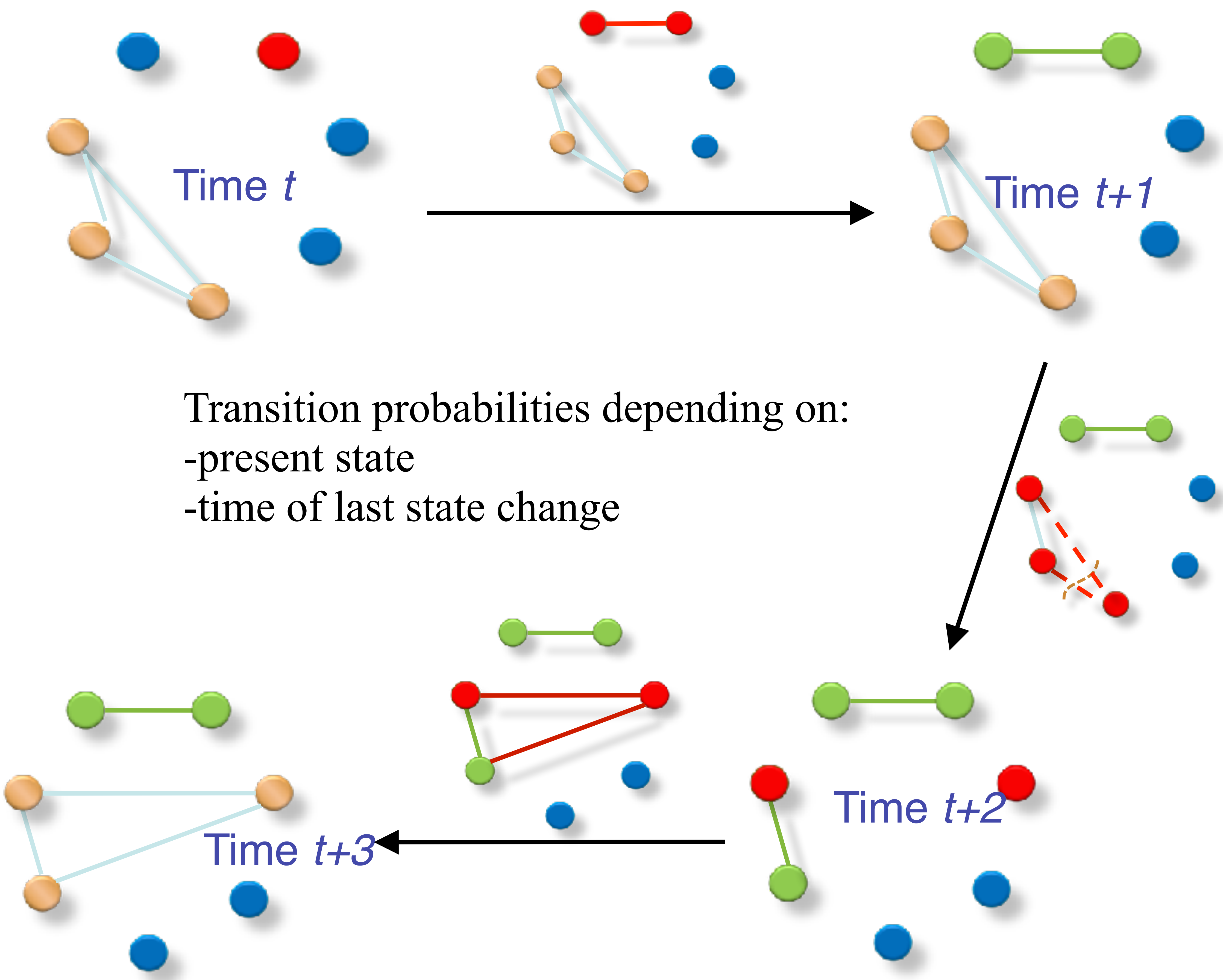
With Memory



Sun et al., EPJB (2015)

>a model of interacting agents

Interacting agents



Transition probabilities depending on:
-present state
-time of last state change

A simple model of interacting agents

At each timestep: choose an agent i at random:

- i isolated: with proba $b_0\, f(\,t,t_i\,)$, agent *i* changes its state, and chooses an agent *j* with probability $\Pi(\,t,t_j\,)$
- i in a group: with probability $b_1 f(\,t,t_i\,)$, agent i changes its state :
  – with probability λ, agent i leaves the group
  – with probability $1\text{-}\lambda$, it introduces an isolated agent *n* chosen with probability $\Pi(\,t,t_n\,)$ to the group

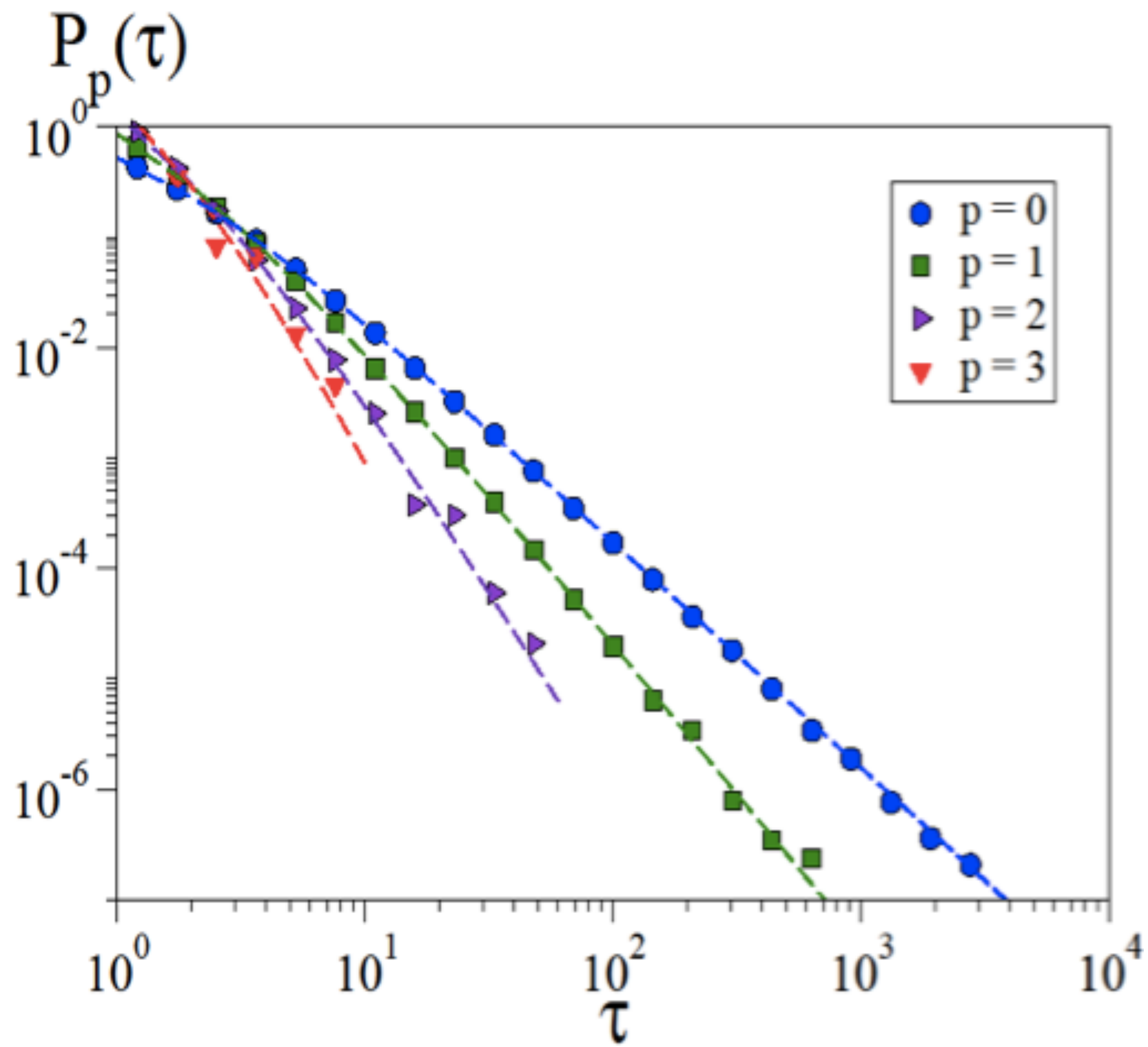
Parameters: b_0, b_1, λ

J. Stehlé et al., Physical Review E 81, 035101 (2010)
K. Zhao et al., Physical Review E 83, 056109 (2011)

Analytical and numerical results

$$\begin{cases} P_0(\tau) \propto (\tau+1)^{-1-b_0\frac{3\lambda-1}{2\lambda-1}} \\ P_p(\tau) \propto (\tau+1)^{-1-b_1(p+1)} \\ \text{where} \quad \tau = \dfrac{t-t_p}{N} \end{cases}$$

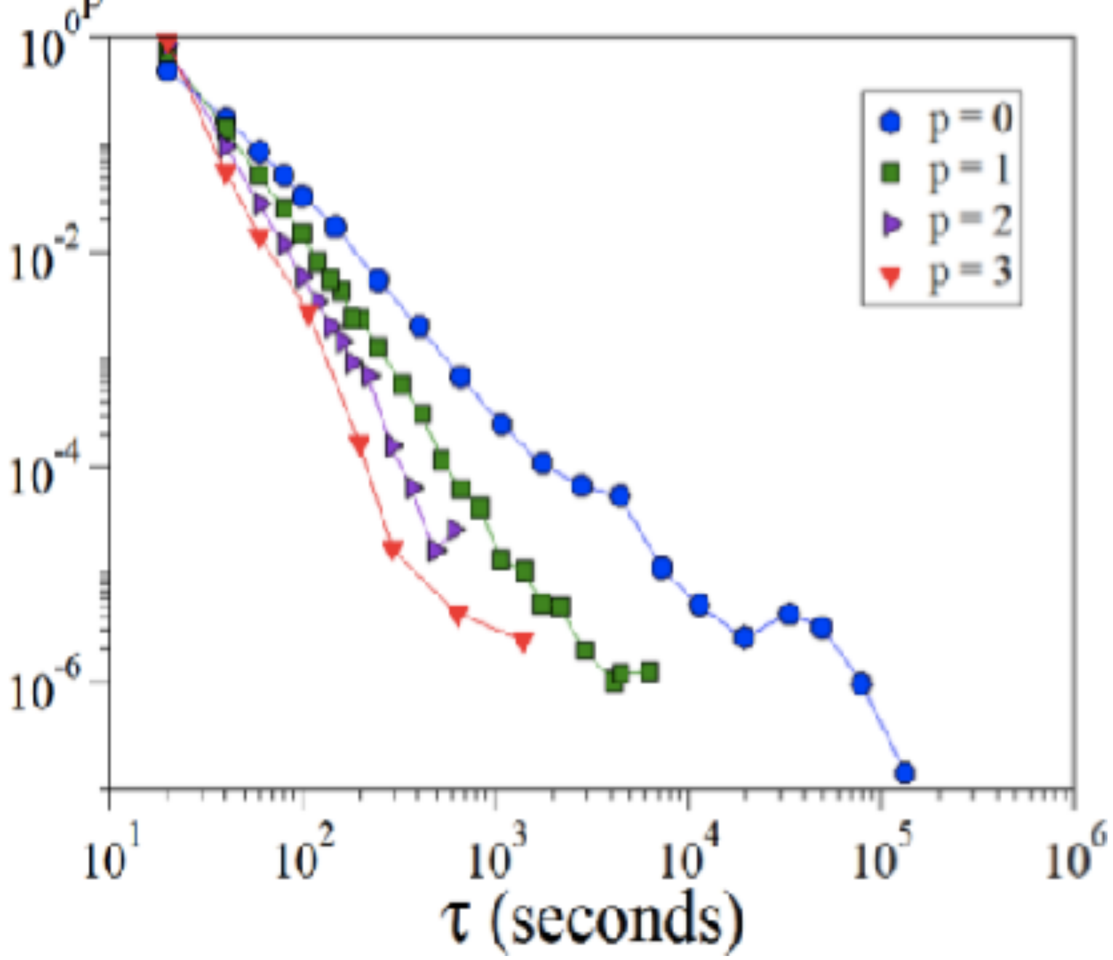
Duration in a given state ***p***



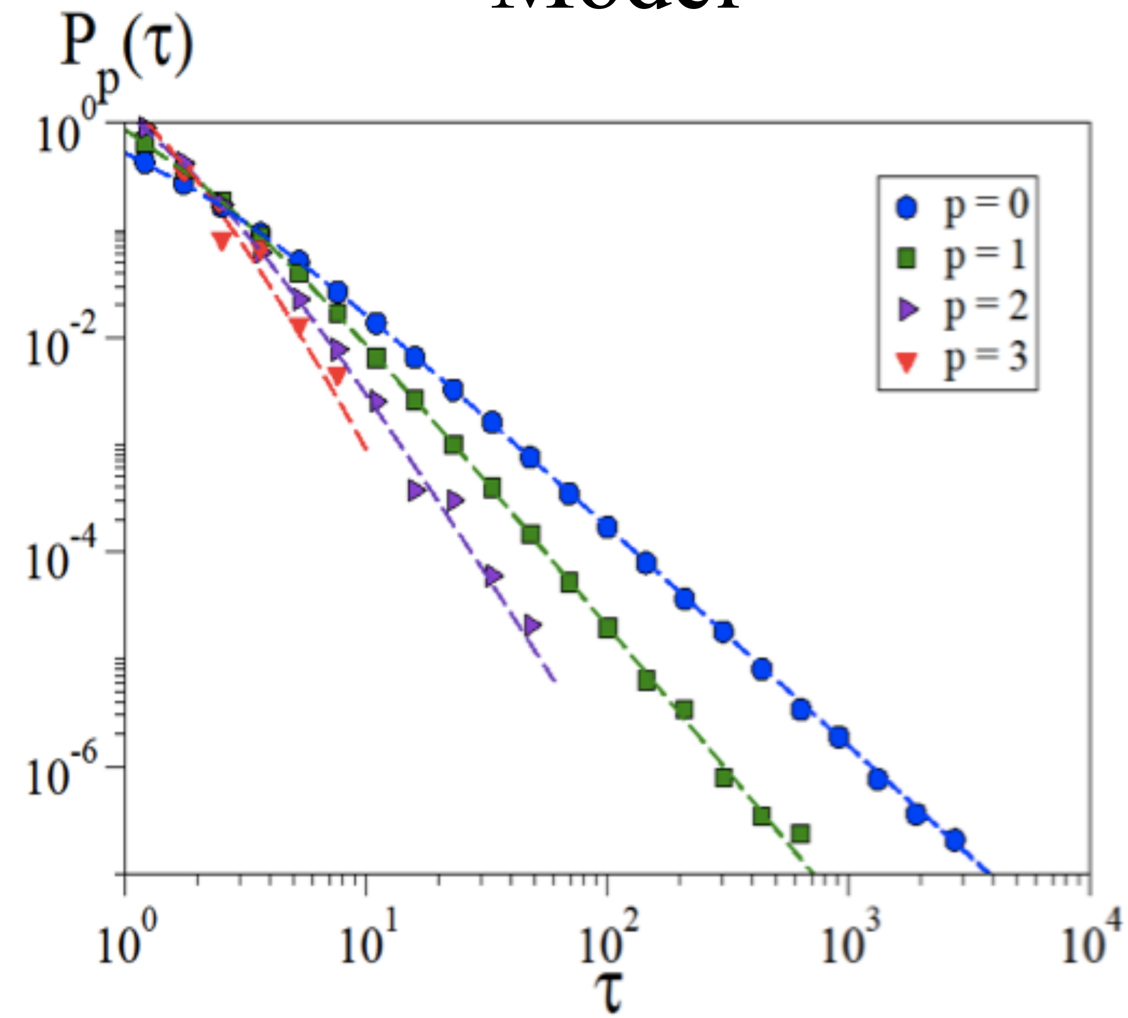
J. Stehlé et al., Physical Review E 81, 035101 (2010)
K. Zhao et al., Physical Review E 83, 056109 (2011)

Distributions of times spent with p neighbours

SocioPatterns data
(face-to-face contacts)



Model



J. Stehlé et al., Physical Review E 81, 035101 (2010)
K. Zhao et al., Physical Review E 83, 056109 (2011)

>another model of interacting agents

N agents;

t_i= last time i changed state

t_{ij} = last time (i,j) changed state

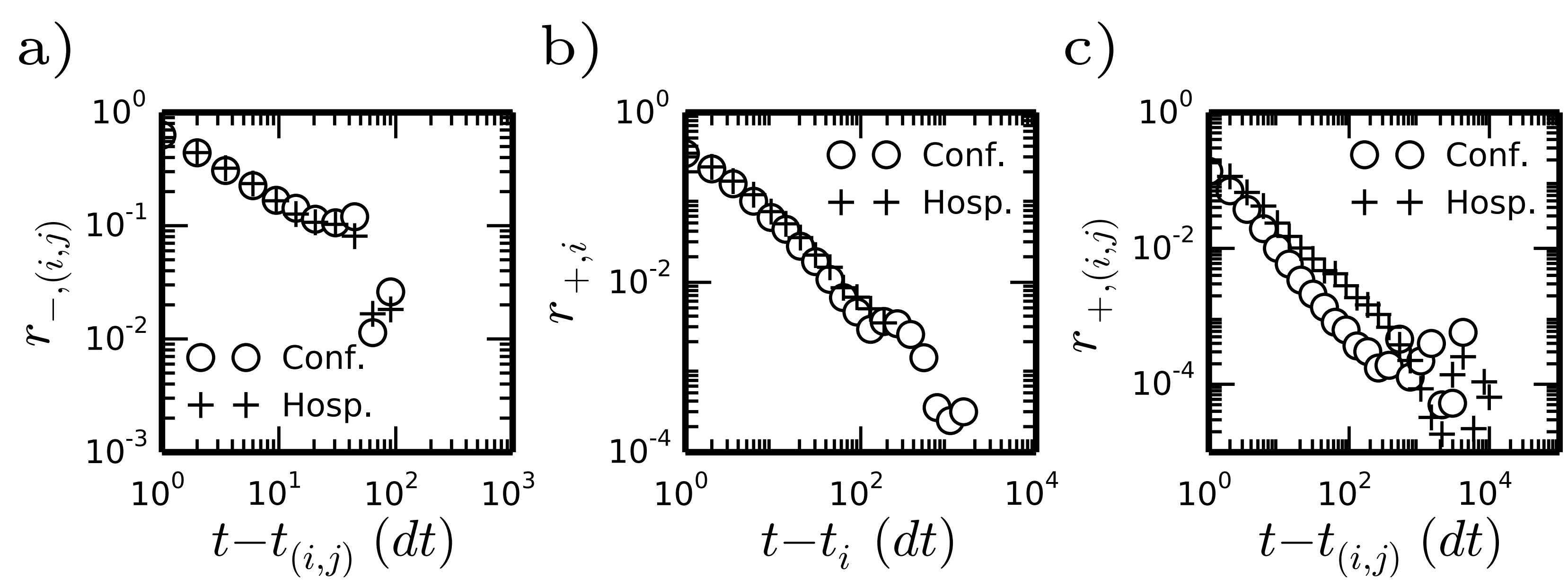
at each time step dt:

(i) Each active link (i, j) is inactivated with proba dt z $f(t - t_{ij})$

(ii) Each agent i initiates a contact with another agent j with probability dt b $f_a(t - t_i)$, j chosen among agents that are not in contact with i with probability $\Pi_a(t-t_j)\Pi(t- t_{ij})$

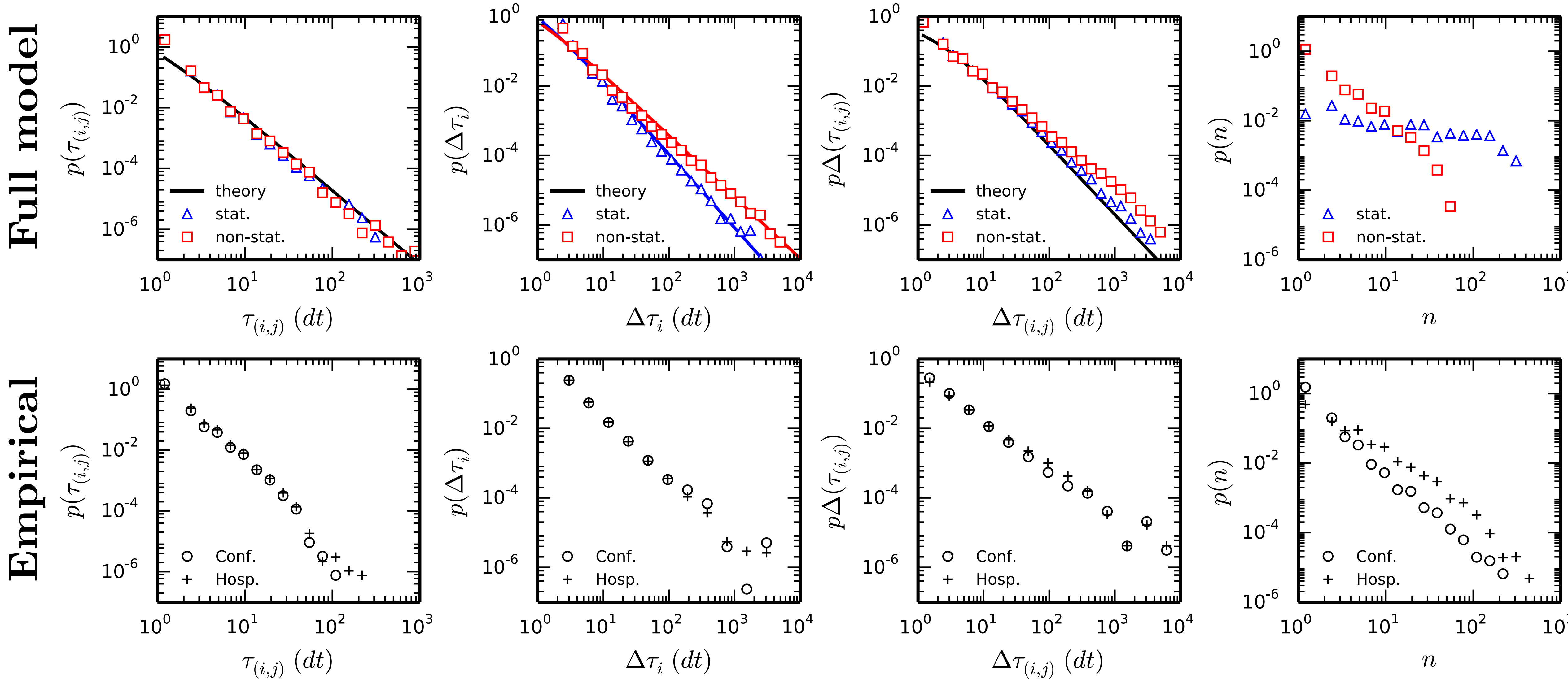
memory kernels f, f_a, Π_a, Π:

- constant: Poissonian dynamics
- empirics: decrease as 1/x



C. Vestergaard et al., Phys. Rev. E 90, 042805 (2014)

Memory mechanisms can be incorporated separately
Need all to reproduce empirical data

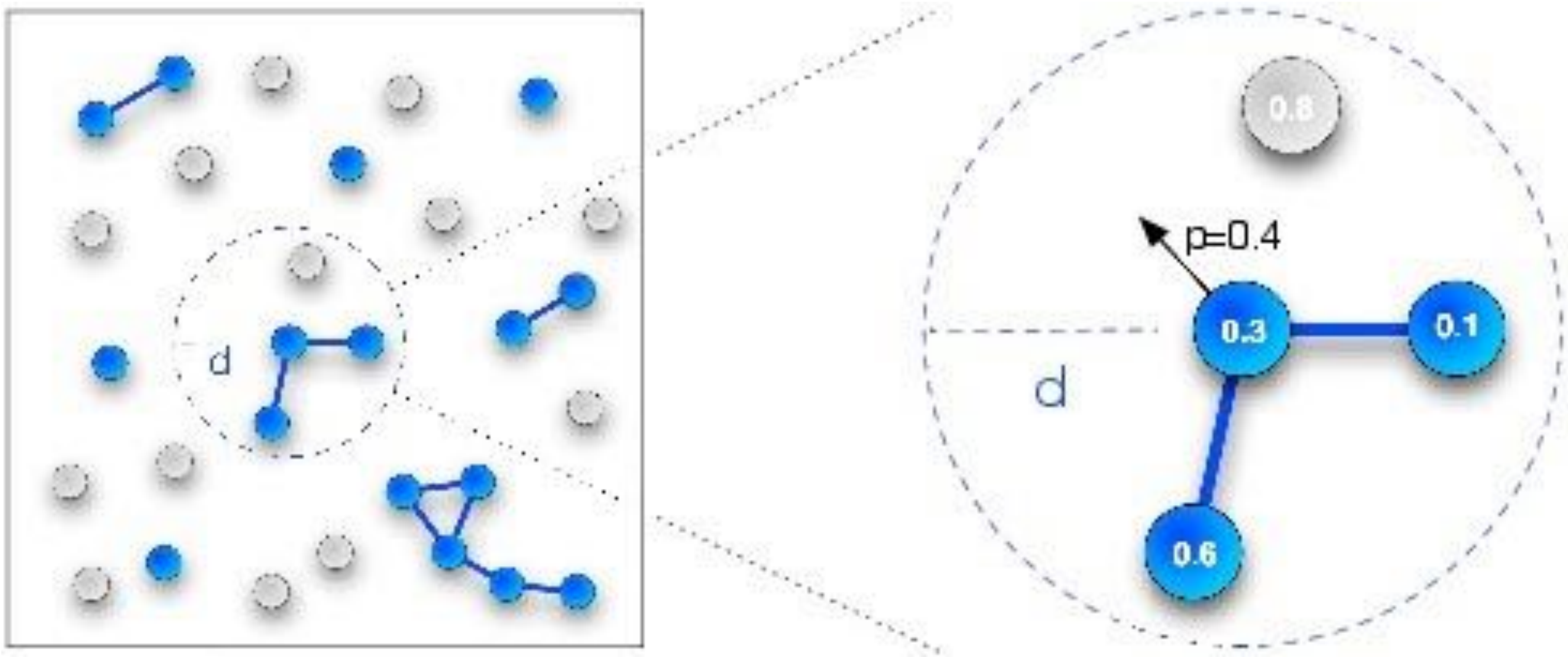


C. Vestergaard et al., Phys. Rev. E 90, 042805 (2014)

>still another model
of interacting agents

Agents with different "attractiveness" performing random walks

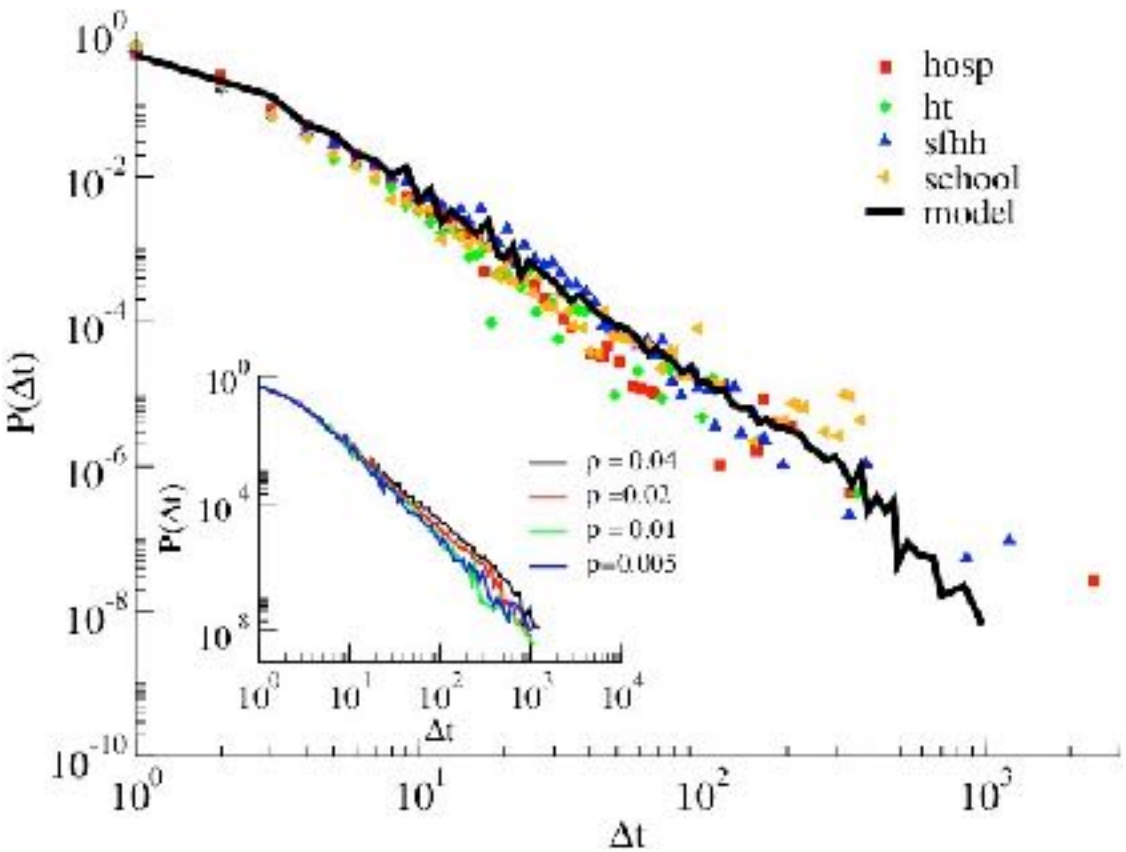
Agent i -> attractiveness a_i

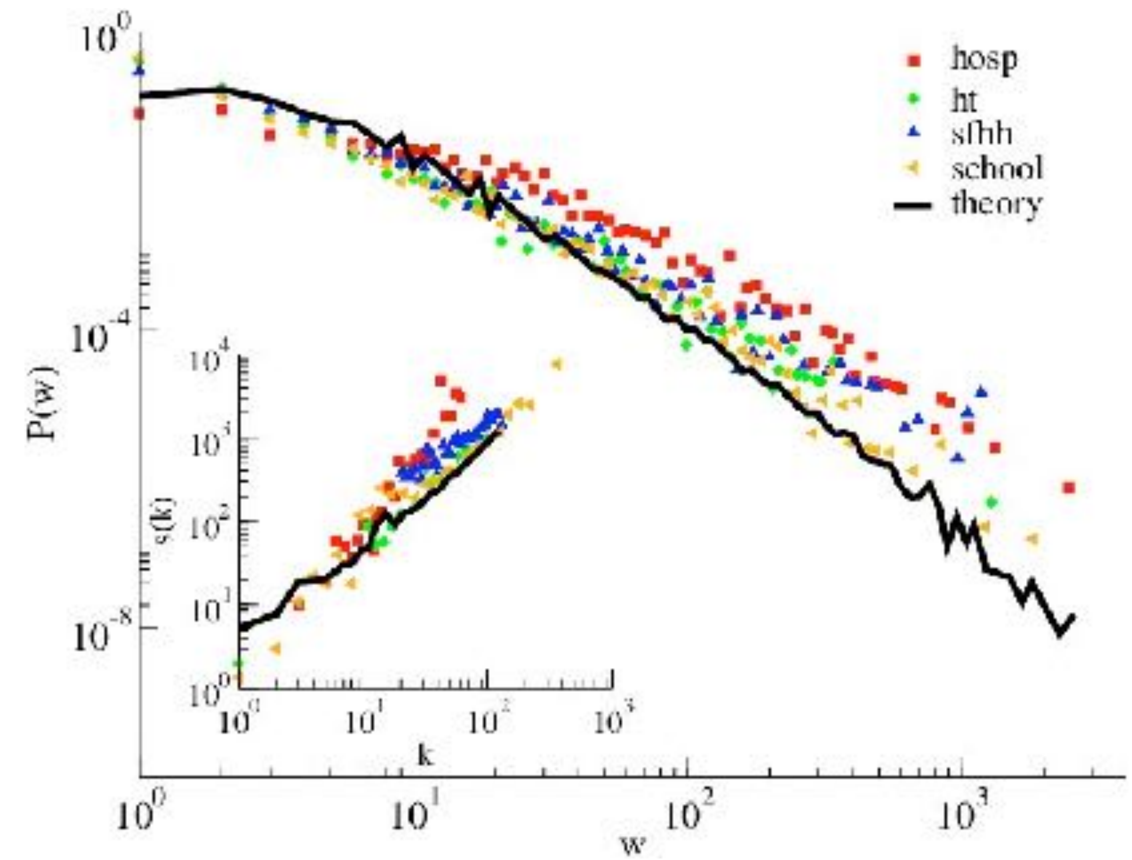


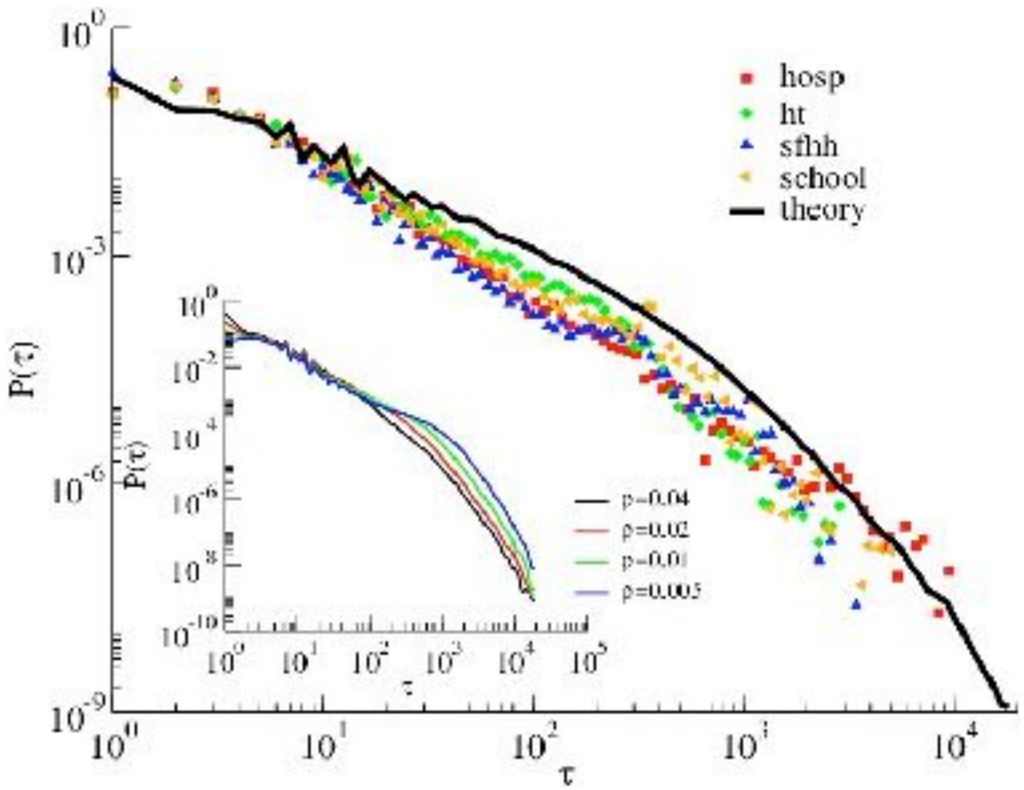
Walking probability: $p_i(t) = 1 - \max_{j \in \mathcal{N}_i(t)} \{a_j\},$

M. Starnini, A. Baronchelli, R. Pastor-Satorras
Modeling human dynamics of face-to-face interaction networks, Phys. Rev. Lett. 110, 168701 (2013)

Agents with different "attractiveness" performing random walks







M. Starnini, A. Baronchelli, R. Pastor-Satorras
Modeling human dynamics of face-to-face interaction networks, Phys. Rev. Lett. 110, 168701 (2013)

>Generative models from data

>Surrogate temporal networks from known empirical statistics

- Generate static underlying structure
- Generate timelines of events
  - known $P(\Delta t)$: generate successive events with inter-event times taken from $P(\Delta t)$
  - known $P(\Delta t)$ and $P(n)$: assign a number of events to each link from $P(n)$, and inter-event times from $P(\Delta t)$
  - known $P(\Delta t)$, $P(n)$, $P(\tau)$
  - known intervals of activity or overall activity timeline
  - …

L. Gauvin et al., Sci Rep 3:3099 (2013)
M. Génois et al., Nat Comm 6:8860 (2015)

NB: no temporal correlations, some can be introduced (through correlated intervals of activities)

>Null Models of temporal networks

What are null models?

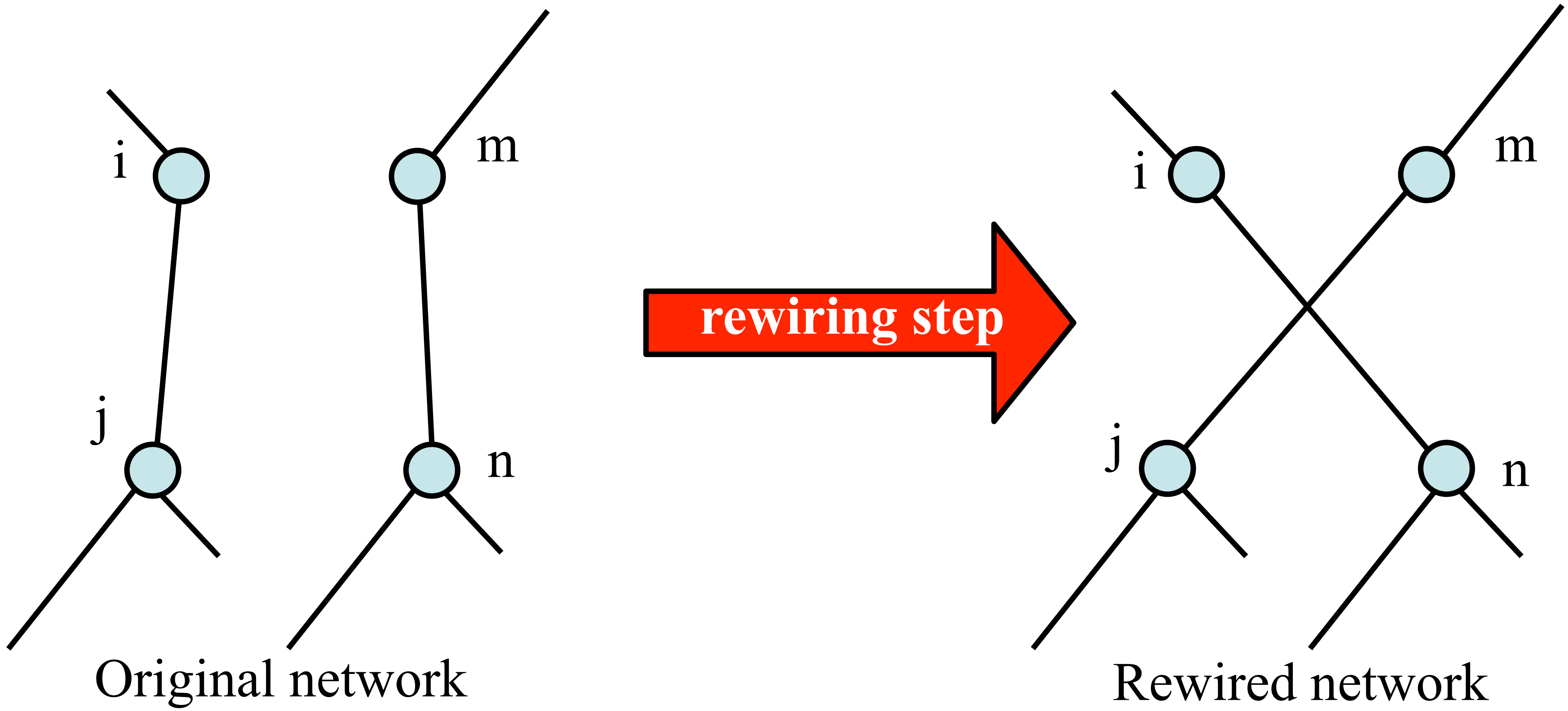
- ensemble of instances of randomly built systems
- that preserve some properties of the studied systems

Aim:

- understand which properties of the studied system are simply random, and which ones denote an underlying mechanism or organizational principle
- compare measures with the known values of a random case

Static null models

- Fixing size (N, E): random (Erdös-Renyi) graph
- Fixing degree sequence: **reshuffling/rewiring** methods



- preserves the degree of each node
- destroys topological correlations

Null models for temporal networks

Many properties and correlations

many possible null models

Random times

for each link event: pick time at random

keeps:
- link structure
- number of events per link
- corresponding static correlations

destroys:
- global time ordering
- activity timeline
- burstiness
- all temporal correlations

Time shuffling

shuffle times of events

ID1	ID2	time
1	4	2
2	3	8
1	5	12
3	4	15
…..		

ID1	ID2	time
1	4	8
2	3	15
1	5	2
3	4	12
…..		

keeps:

- link structure
- number of events per link
- corresponding static correlations
- global time ordering and activity timeline

destroys:

- interevent times
- all temporal correlations

Interval shuffling

for each link: randomize sequence of inter-event durations

for each link: randomize sequence of contacts

keeps:

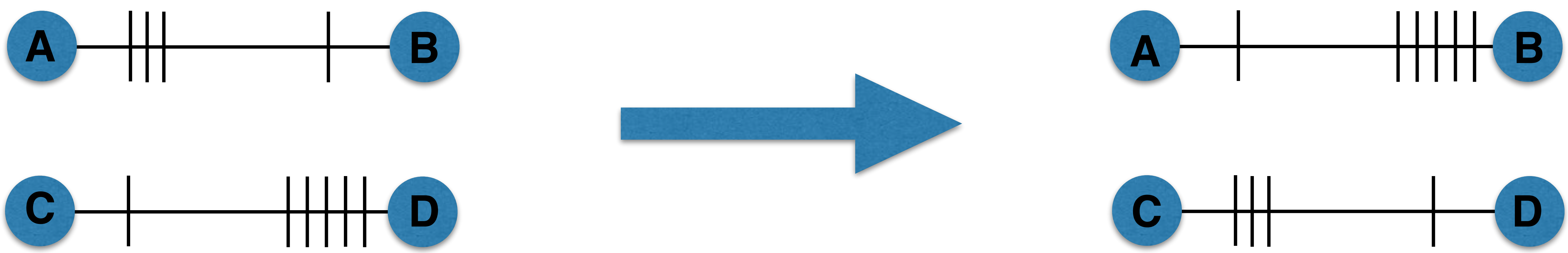
- link structure
- number of events per link
- corresponding static correlations
- link burstiness

destroys:

- all temporal correlations
- activity timeline

Link-sequence shuffling

Randomly exchange sequence of events of different links



keeps:

- link structure
- distribution of link weights
- global activity timeline
- link burstiness

destroys:

- number of events per link & corresponding correlations
- correlations between structure and activity
- all temporal correlations

NB: weight-conserving link-sequence shuffling

To go further

Gauvin et al., Randomized reference models for temporal networks

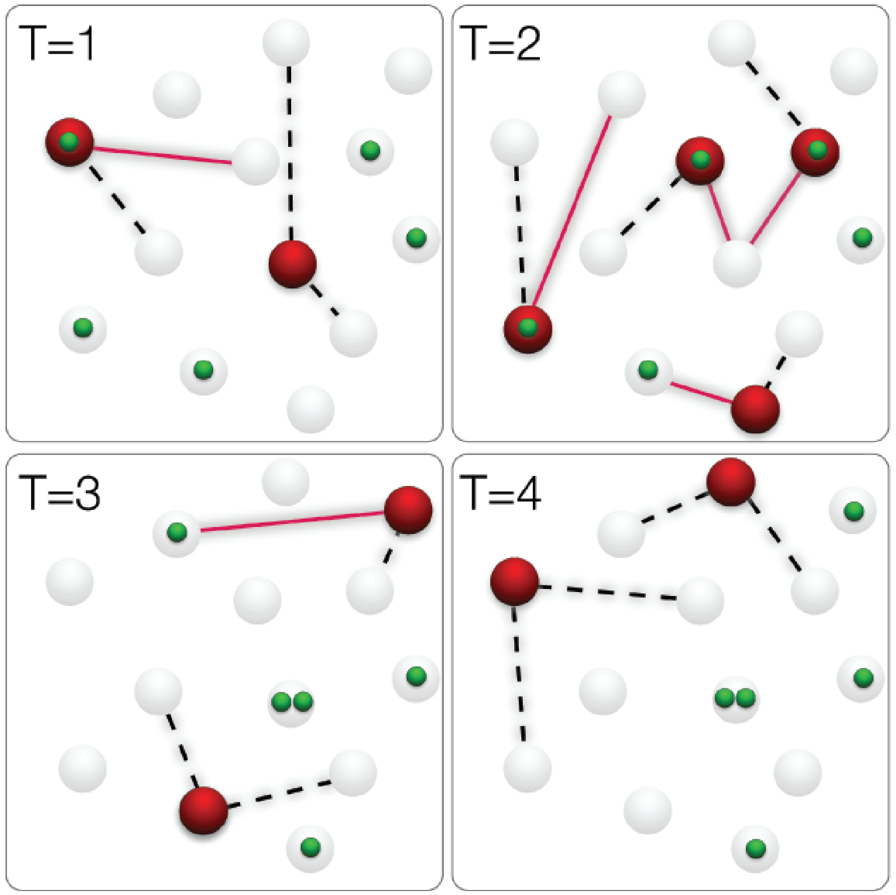
arXiv:1806.04032

Canonical name	Common name	Aggregated								Temporal-topological									
		topological			weighted				temp.		node				link				
		G^{stat}	k_i	L	a_i	s_i	$n_{(i,j)}$	$w_{(i,j)}$	A^t	Γ^t	Φ_i	α_i^m	$\Delta\alpha_i^m$	d_i^t	$\Theta_{(i,j)}$	$\tau_{(i,j)}^m$	$\Delta\tau_{(i,j)}^m$	$t_{(i,j)}^1$	$t_{(i,j)}^w$
Event shufflings:																			
$\mathrm{P}[p(\boldsymbol{\tau})]$	*Event shuffling*	–	–	–	–	μ	–	–	μ	–	–	–	–	μ	–	μ	–	–	–
$\mathrm{P}[p(\mathbf{t},\boldsymbol{\tau})]$		–	–	–	–	μ	–	–	×	–	–	–	–	$\mu_\mathcal{T}$	–	$\pi_\mathcal{T}$	–	–	–
Link shufflings:																			
$\mathrm{P}[p_\mathcal{L}(\boldsymbol{\Theta})]$	*Link shuffling*	–	μ	×	μ	μ	p	p	×	–	–	–	–	$\mu_\mathcal{T}$	$p_\mathcal{L}$	$p_\mathcal{L}$	$p_\mathcal{L}$	p	p
$\mathrm{P}[\mathrm{I}_\lambda,p_\mathcal{L}(\boldsymbol{\Theta})]$		I_λ	μ	×	μ	μ	p	p	×	–	–	–	–	$\mu_\mathcal{T}$	$p_\mathcal{L}$	$p_\mathcal{L}$	$p_\mathcal{L}$	p	p
$\mathrm{P}[\mathbf{k},p_\mathcal{L}(\boldsymbol{\Theta})]$	*Maslov-Sneppen*	–	×	×	μ	μ	p	p	×	–	–	–	–	$\mu_\mathcal{T}$	$p_\mathcal{L}$	$p_\mathcal{L}$	$p_\mathcal{L}$	p	p
$\mathrm{P}[\mathbf{k},\mathrm{I}_\lambda,p_\mathcal{L}(\boldsymbol{\Theta})]$		I_λ	×	×	μ	μ	p	p	×	–	–	–	–	$\mu_\mathcal{T}$	$p_\mathcal{L}$	$p_\mathcal{L}$	$p_\mathcal{L}$	p	p
Timeline shufflings:																			
$\mathrm{P}[\mathcal{L},p(\boldsymbol{\tau})]$	*Timeline shuffling*	×	×	×	–	μ	μ	μ	μ	–	–	–	–	μ	–	μ	–	–	–
$\mathrm{P}[\mathcal{L},p(\mathbf{t},\boldsymbol{\tau})]$		×	×	×	–	μ	μ	μ	×	–	–	–	–	$\mu_\mathcal{T}$	–	$\pi_\mathcal{T}$	–	–	–
$\mathrm{P}[\mathbf{n},p(\mathbf{t},\boldsymbol{\tau})]$		×	×	×	×	μ	×	μ	×	–	–	–	–	$\mu_\mathcal{T}$	–	$\pi_\mathcal{T}$	–	–	–
$\mathrm{P}[\boldsymbol{\pi}_\mathcal{L}(\boldsymbol{\tau})]$		×	×	×	×	×	×	×	μ	–	–	–	–	μ	–	$\pi_\mathcal{L}$	–	–	–
$\mathrm{P}[\boldsymbol{\pi}_\mathcal{L}(\boldsymbol{\tau}),\mathbf{t}^1,\mathbf{t}^w]$		×	×	×	×	×	×	×	μ	–	–	–	–	μ	–	$\pi_\mathcal{L}$	$\mu_\mathcal{L}$	×	×
$\mathrm{P}[\boldsymbol{\pi}_\mathcal{L}(\boldsymbol{\tau}),\boldsymbol{\pi}_\mathcal{L}(\boldsymbol{\Delta\tau})]$		×	×	×	×	×	×	×	μ	–	–	–	–	μ	–	$\pi_\mathcal{L}$	$\pi_\mathcal{L}$	–	–
$\mathrm{P}[\boldsymbol{\pi}_\mathcal{L}(\boldsymbol{\tau}),\boldsymbol{\pi}_\mathcal{L}(\boldsymbol{\Delta\tau}),\mathbf{t}^1]$		×	×	×	×	×	×	×	μ	–	–	–	–	μ	–	$\pi_\mathcal{L}$	$\pi_\mathcal{L}$	×	×
$\mathrm{P}[\mathbf{per}(\boldsymbol{\Theta})]$		×	×	×	×	×	×	×	μ	–	–	–	–	μ	–	×	×	–	–
$\mathrm{P}[\boldsymbol{\tau},\boldsymbol{\Delta\tau}]$		×	×	×	×	×	×	×	μ	–	–	–	–	μ	–	×	×	–	–
Intersections:																			
$\mathrm{P}[\mathcal{L},p_\mathcal{L}(\boldsymbol{\Theta})]$		×	×	×	μ	μ	p	p	×	–	–	–	–	$\mu_\mathcal{T}$	$p_\mathcal{L}$	$p_\mathcal{L}$	$p_\mathcal{L}$	p	p
$\mathrm{P}[\mathbf{w},p_\mathcal{L}(\boldsymbol{\Theta})]$		×	×	×	μ	×	p	×	×	–	–	–	–	$\mu_\mathcal{T}$	$p_\mathcal{L}$	$p_\mathcal{L}$	$p_\mathcal{L}$	p	p
$\mathrm{P}[\mathbf{n},p_\mathcal{L}(\boldsymbol{\Theta})]$		×	×	×	×	μ	×	p	×	–	–	–	–	$\mu_\mathcal{T}$	$p_\mathcal{L}$	$p_\mathcal{L}$	$p_\mathcal{L}$	p	p
Instant-event shufflings:																			
$\mathrm{P}[E]$	*Instant-event shuffling*	–	–	–	–	μ	–	–	μ	–	–	–	–	μ	–	–	–	–	–
Timeline shufflings:																			
$\mathrm{P}[\mathcal{L},E]$	*Timeline shuffling*	×	×	×	–	μ	–	μ	μ	–	–	–	–	μ	–	–	–	–	–
$\mathrm{P}[\mathbf{w}]$		×	×	×	–	×	–	×	μ	–	–	–	–	μ	–	–	–	–	–
$\mathrm{P}[\mathbf{w},\mathbf{t}^1,\mathbf{t}^w]$		×	×	×	–	×	–	×	μ	–	–	–	–	μ	–	–	–	×	×
$\mathrm{P}[\mathbf{w},\boldsymbol{\pi}_\mathcal{L}(\boldsymbol{\Delta\tau})]$		×	×	×	×	×	×	×	μ	–	–	–	–	μ	–	–	$\pi_\mathcal{L}$	–	–
$\mathrm{P}[\mathbf{w},\boldsymbol{\pi}_\mathcal{L}(\boldsymbol{\Delta\tau}),\mathbf{t}^1,\mathbf{t}^w]$		×	×	×	×	×	×	×	μ	–	–	–	–	μ	–	$\mu_\mathcal{L}$	$\pi_\mathcal{L}$	×	×
Sequence shufflings:																			
$\mathrm{P}[p_\mathcal{T}(\boldsymbol{\Gamma})]$	*Sequence shuffling*	×	×	×	–	×	–	×	p	$p_\mathcal{T}$	–	–	–	$p_\mathcal{T}$	–	–	–	–	–
$\mathrm{P}[p_\mathcal{T}(\boldsymbol{\Gamma}),\mathbf{sgn}(\mathbf{A})]$		×	×	×	–	×	–	×	p, sgn	$p_\mathcal{T}$	–	–	–	$p_\mathcal{T}$	–	–	–	–	–
Snapshot shufflings:																			
$\mathrm{P}[\mathbf{t}]$	*Snapshot shuffling*	–	–	–	–	μ	–	–	×	–	–	–	–	$\mu_\mathcal{T}$	–	–	–	–	–
$\mathrm{P}[\mathbf{t},\boldsymbol{\Phi}]$		–	–	–	–	μ	–	–	×	–	×	×	×	$\mu_\mathcal{T}$	–	–	–	–	–
$\mathrm{P}[\mathbf{d}]$		–	–	–	–	μ	–	–	×	–	×	×	×	×	–	–	–	–	–
$\mathrm{P}[\mathbf{iso}(\boldsymbol{\Gamma})]$		–	–	–	–	μ	–	–	×	$\simeq$	–	–	–	$\pi_\mathcal{T}$	–	–	–	–	–
$\mathrm{P}[\mathbf{iso}(\boldsymbol{\Gamma}),\boldsymbol{\Phi}]$		–	–	–	–	μ	–	–	×	$\simeq$	×	×	×	$\pi_\mathcal{T}$	–	–	–	–	–
Intersections:																			
$\mathrm{P}[\mathcal{L},\mathbf{t}]$		×	×	×	–	μ	–	μ	×	–	–	–	–	$\mu_\mathcal{T}$	–	–	–	–	–
$\mathrm{P}[\mathbf{w},\mathbf{t}]$	*Timestamp shuffling*	×	×	×	–	×	–	×	×	–	–	–	–	$\mu_\mathcal{T}$	–	–	–	–	–

>Dynamical processes on temporal networks

>Random walks

Random walks on activity-driven network



N nodes
W walkers
W_a average number of walkers in a node of activity a
w = W/N

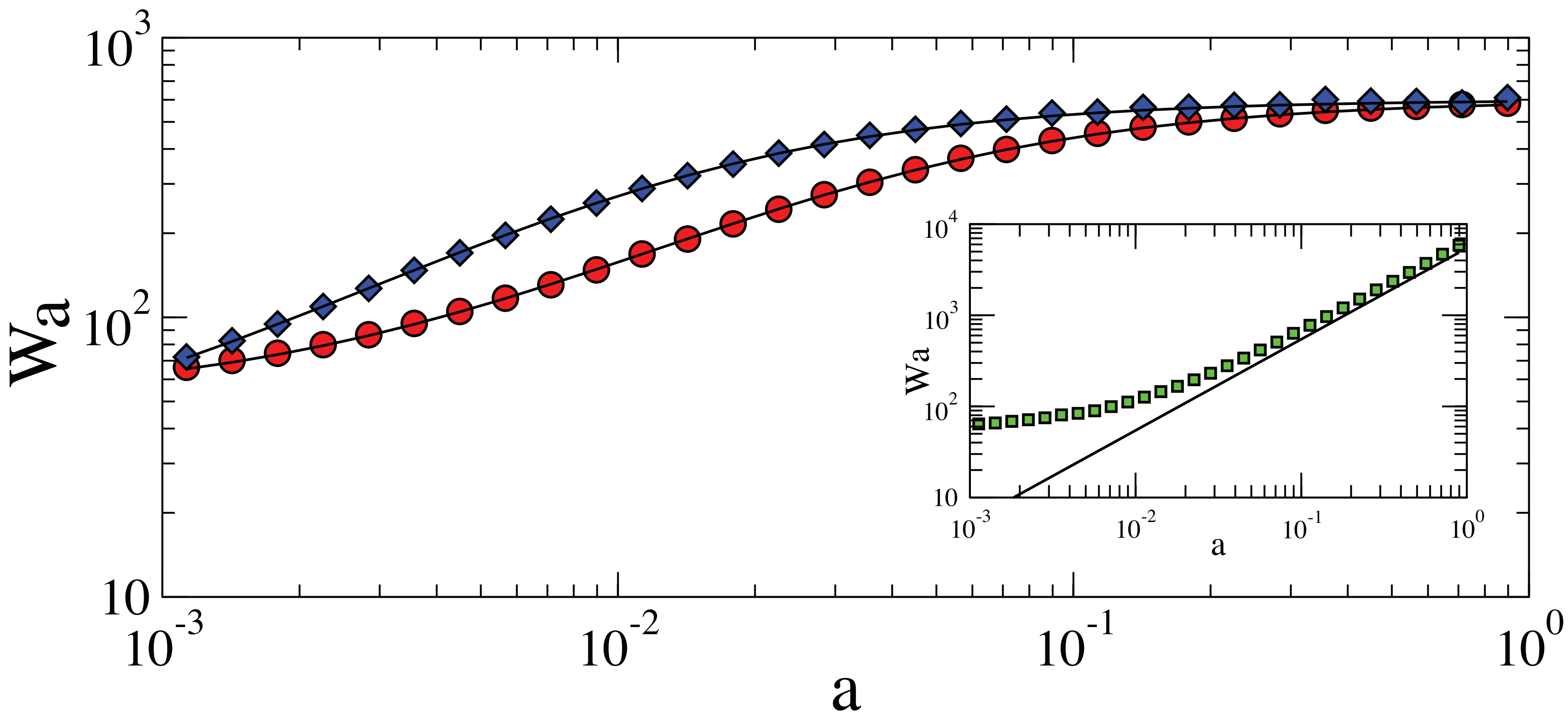
$$\frac{\partial W_a(t)}{\partial t} = -aW_a(t) + amw - m\langle a\rangle W_a(t) + \int a' W_{a'}(t) F(a')\,\mathrm{d}a'$$

Perra et al., Phys. Rev. E (2012)

Random walks on activity-driven network

Stationary state:

$$W_a = \frac{amw + \phi}{a + m\langle a\rangle} \qquad \phi = \int aF(a)\,\frac{amw+\phi}{a+m\langle a\rangle}\mathrm{d}a$$



More walkers on more active nodes

Perra et al., Phys. Rev. E (2012)

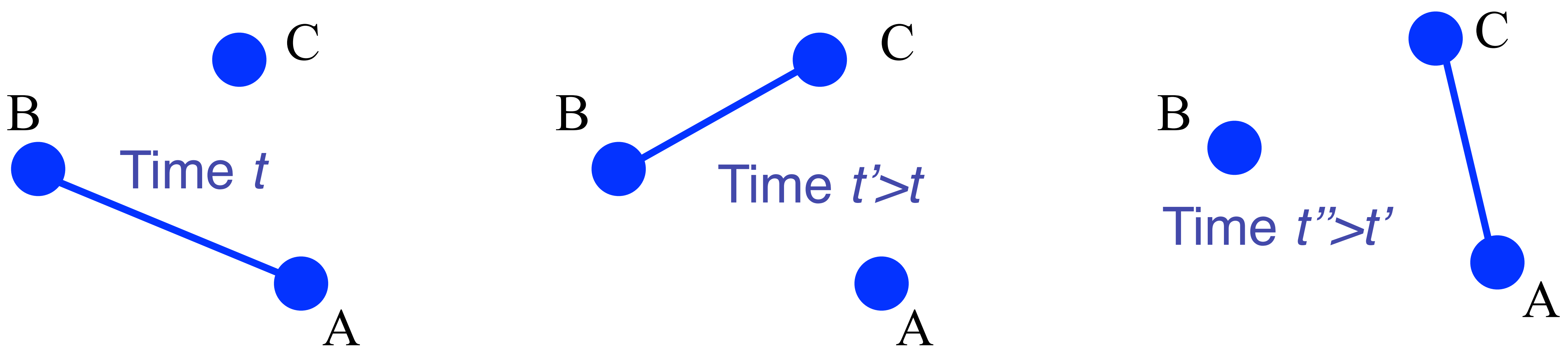
Random walks on empirical temporal networks
Starnini et al., Phys Rev E (2012)
arXiv:1203.2477

>Toy spreading processes

Toy spreading processes on temporal networks

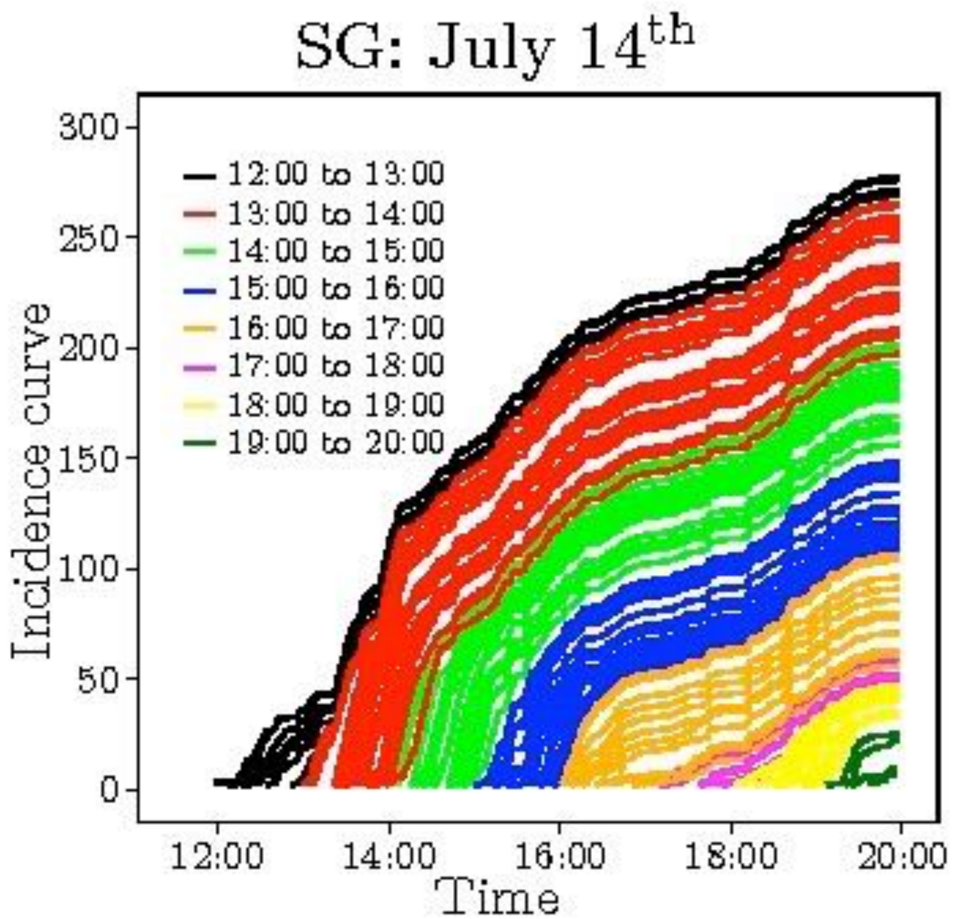
- deterministic SI process to probe the causal structure of the dynamical network

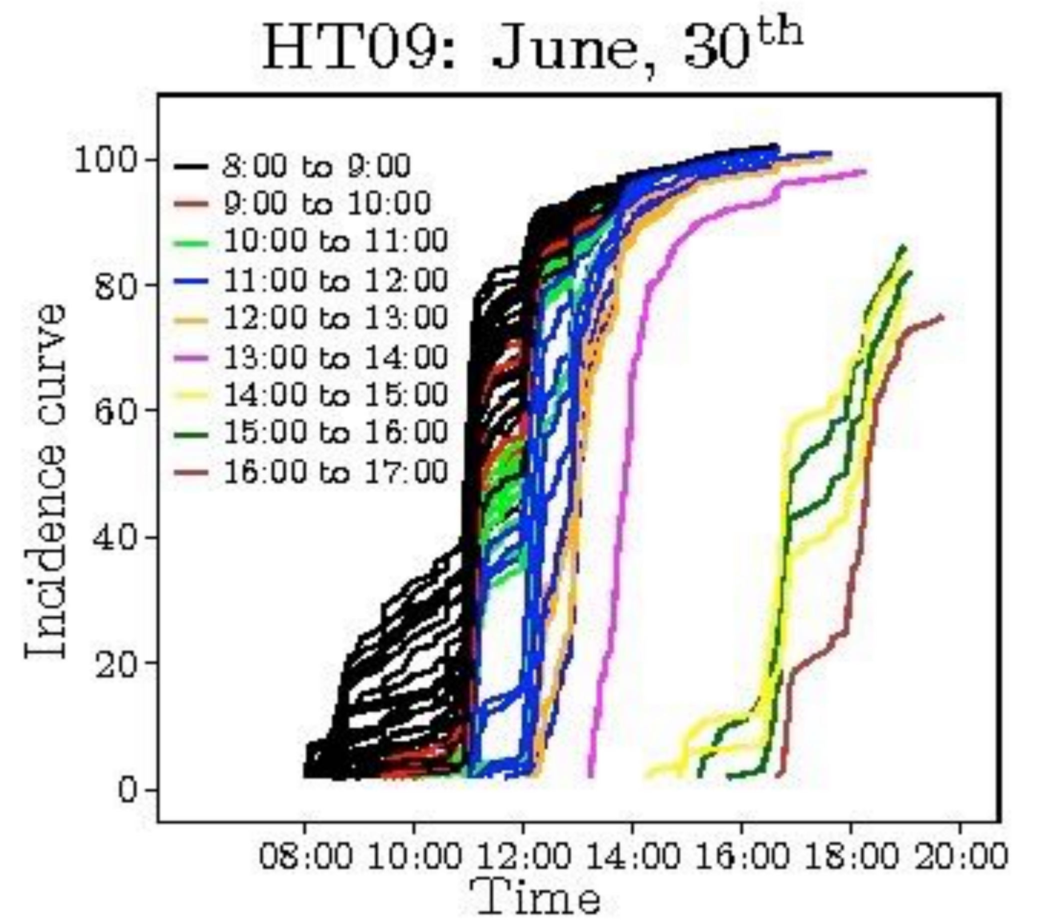
- **fastest paths ≠ shortest paths**



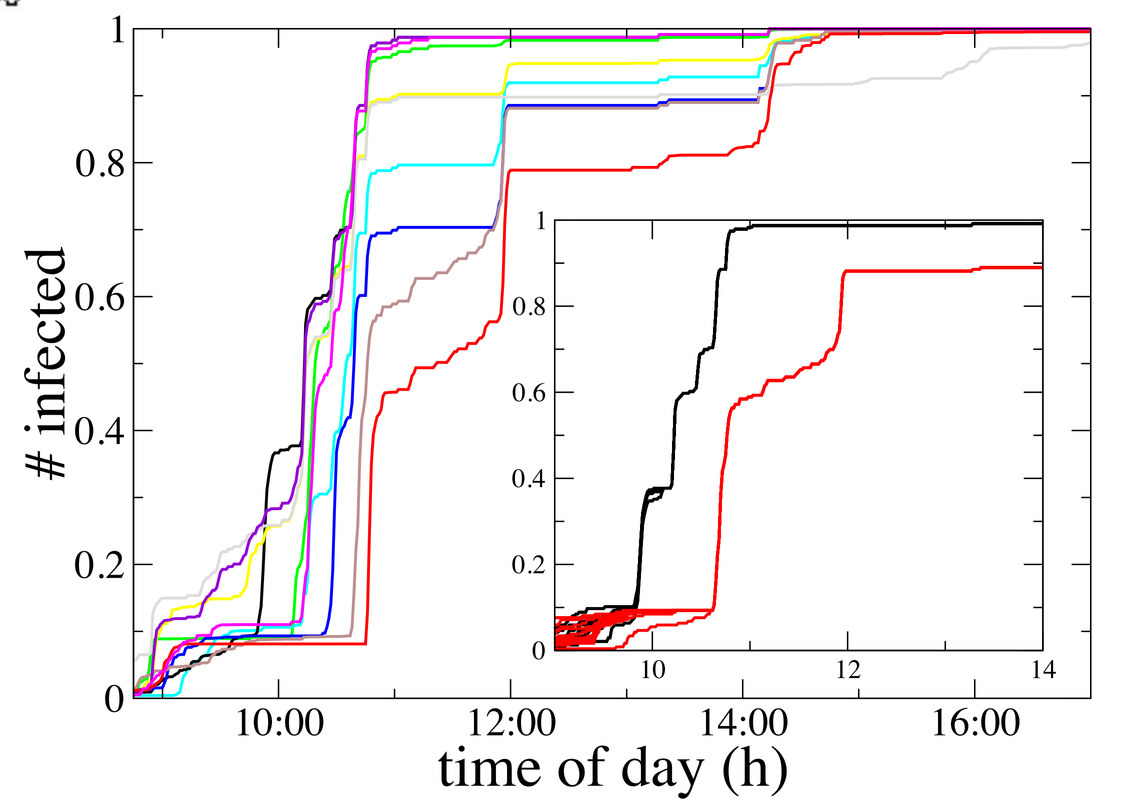
Fastest path= A->B->C
Shortest path= A-C

Example: contact networks





SI deterministic
spreading process



Mobile phone data:

- community structure (C)
- weight-topology correlations (W)
- burstiness on single links (B)
- daily patterns (D)
- event-event correlations between links (E)

Effects of the different ingredients?

↓

Use series of null models!

M. Karsai et al., Small But Slow World: How Network Topology and Burstiness Slow Down Spreading, Phys. Rev. E (2011).

- community structure (C)
- weight-topology correlations (W)
- burstiness on single links (B)
- daily patterns (D)
- event-event correlations between links (E)

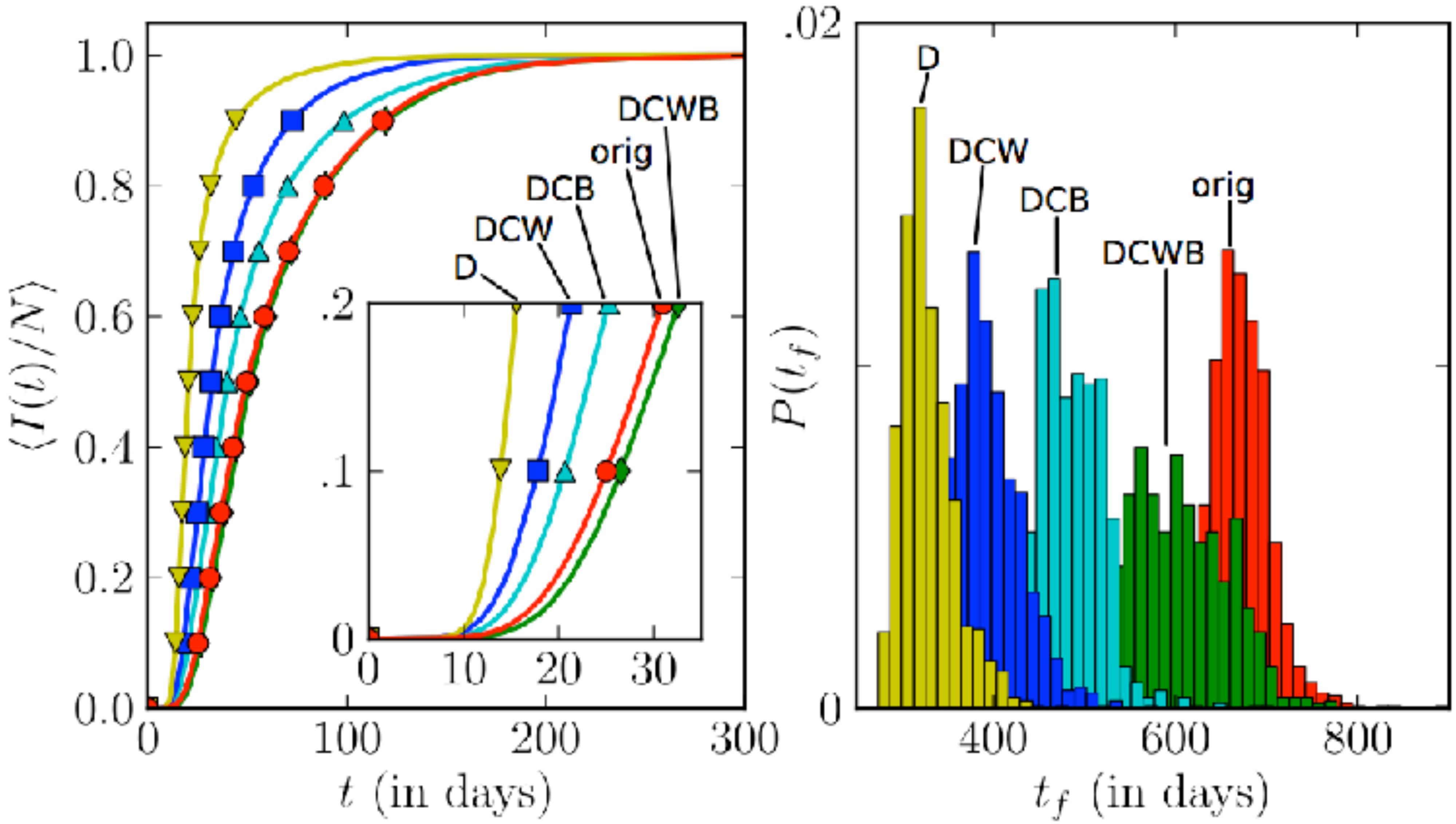
Null models

EVENT SEQUENCE	D	C	W	B	E
Original	✓	✓	✓	✓	✓
Equal-weight link-sequence shuffled	✓	✓	✓	✓	
Link-sequence shuffled	✓	✓		✓	
Time shuffled	✓	✓	✓		
Configuration model	✓				

M. Karsai et al., Small But Slow World: How Network Topology and Burstiness Slow Down Spreading, Phys. Rev. E (2011).

Mobile phone data

- community structure (C)
- weight-topology correlations (W)
- burstiness on single links (B)
- daily patterns (D)
- event-event correlations between links (E)



Burstiness slows down spread
Correlations slightly favour spread

M. Karsai et al., Small But Slow World: How Network Topology and Burstiness Slow Down Spreading, Phys. Rev. E (2011).
Kivela et al, Multiscale Analysis of Spreading in a Large Communication Network, JSTAT (2012)

More results

Rocha et al., PLOS Comp Biol (2011)
- data: temporal network of sexual contacts
- temporal correlations accelerate outbreaks

Pan & Saramaki, PRE (2011)
- data: mobile phone call network
- slower spread when correlations removed

Miritello et al., PRE (2011)
- data: mobile phone call network
- burstiness decreases transmissibility

Takaguchi et al., PLOS ONE (2013)
- data: contacts in a conference; email
- threshold-based spreading model
- burstiness accelerate spreading

Rocha & Blondel, PLOS Comp Biol (2013)
- model with tuneable distribution of inter-event times (no correlations)
- burstiness => initial speedup, long time slowing down

Still somewhat unclear picture

Results

- depend on data set
- depend on spreading model
- generally
  - burstiness slows down spreading
  - correlations (e.g., temporal motifs) favor spreading
  - role of turnover
  - +: effect of static patterns

>SIS and SIR models on temporal networks

SIS model on activity-driven network

Activity-based mean-field theory:

I_a^t Number of infectious nodes of activity a

$$I_a^{t+\Delta t} = I_a^t - \mu I_a^t \Delta t + \beta a m \Delta t (N_a - I_a^t) \int \frac{I_{a'}^t}{N} da' + \beta \Delta t (N_a - I_a^t) m \int a' \frac{I_{a'}^t}{N} da'$$

Write equations for $\theta^t = \int a I_a^t da \ , \quad I^t = \int I_a^t da$

$$\partial_t I = -\mu I + \beta m \langle a \rangle I + \beta m \theta$$
$$\partial_t \theta = -\mu \theta + \beta m \langle a \rangle \theta + \beta m \langle a^2 \rangle I$$

(neglecting 2nd order terms)

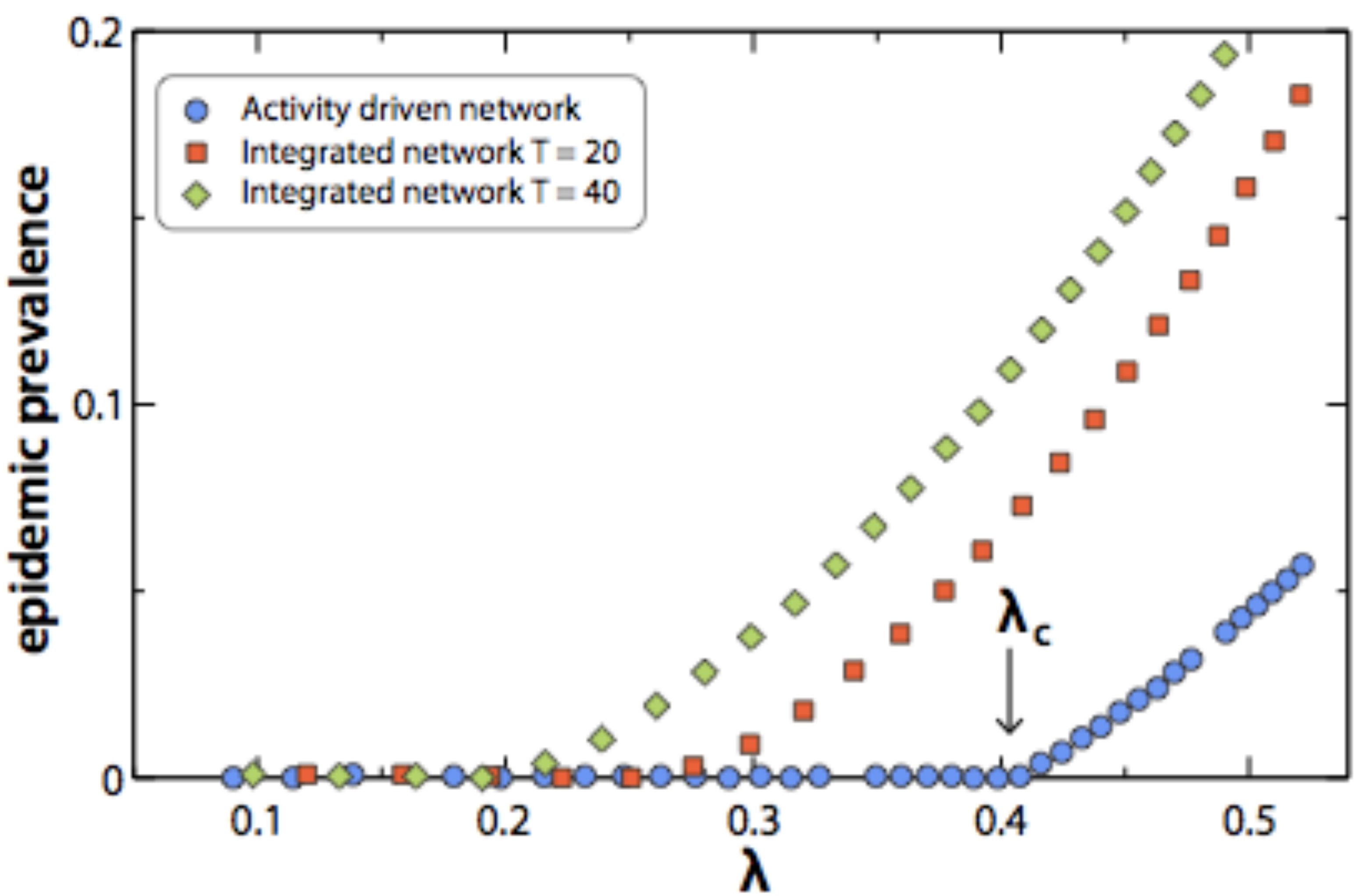
Condition: largest eigenvalue larger than 0

Epidemic threshold: $\dfrac{\beta}{\mu} > \dfrac{1}{m\left(\langle a \rangle + \sqrt{\langle a^2 \rangle}\right)}$

Perra et al., Sci. Rep. (2012)

SIS model on activity-driven network

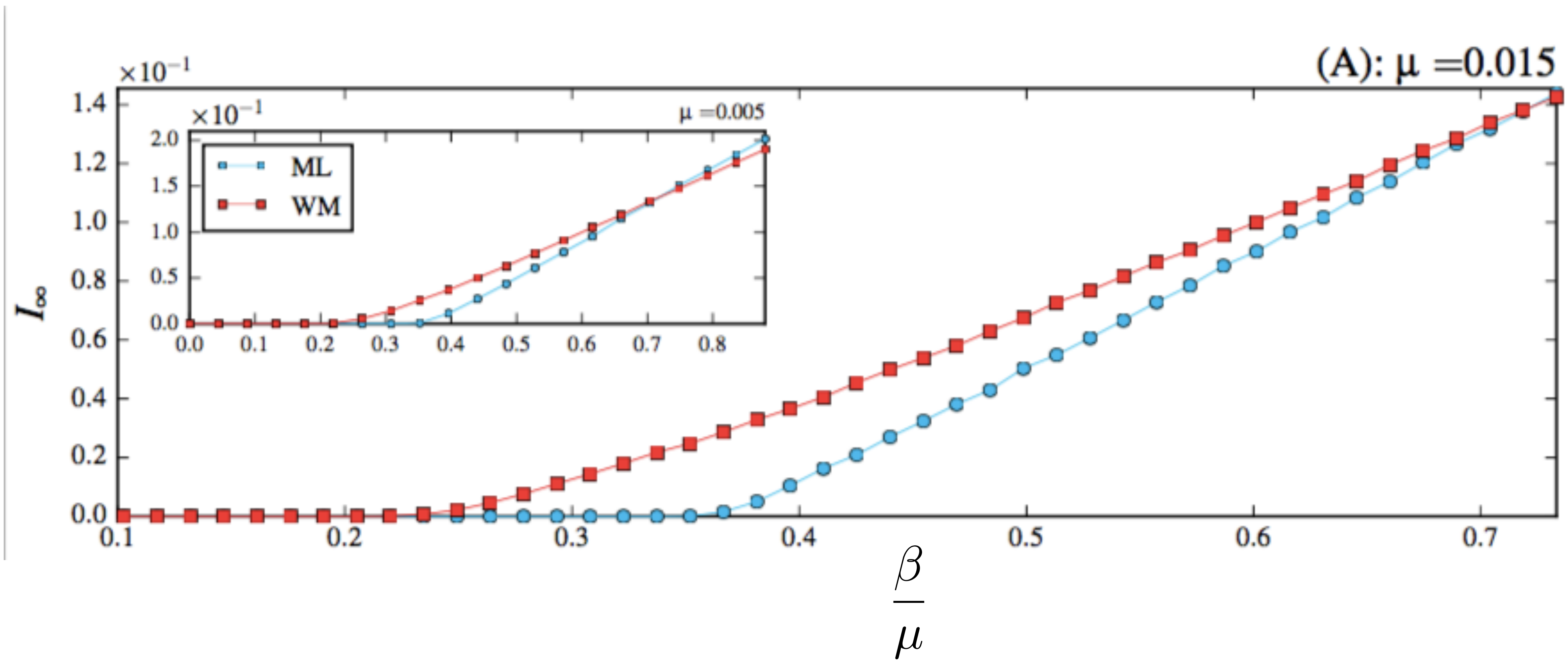
Epidemic threshold: $$\lambda_c = \frac{1}{m\left(\langle a \rangle + \sqrt{\langle a^2 \rangle}\right)}$$



Perra et al., Sci. Rep. (2012)

SIS and SIR models
on activity-driven networks with memory

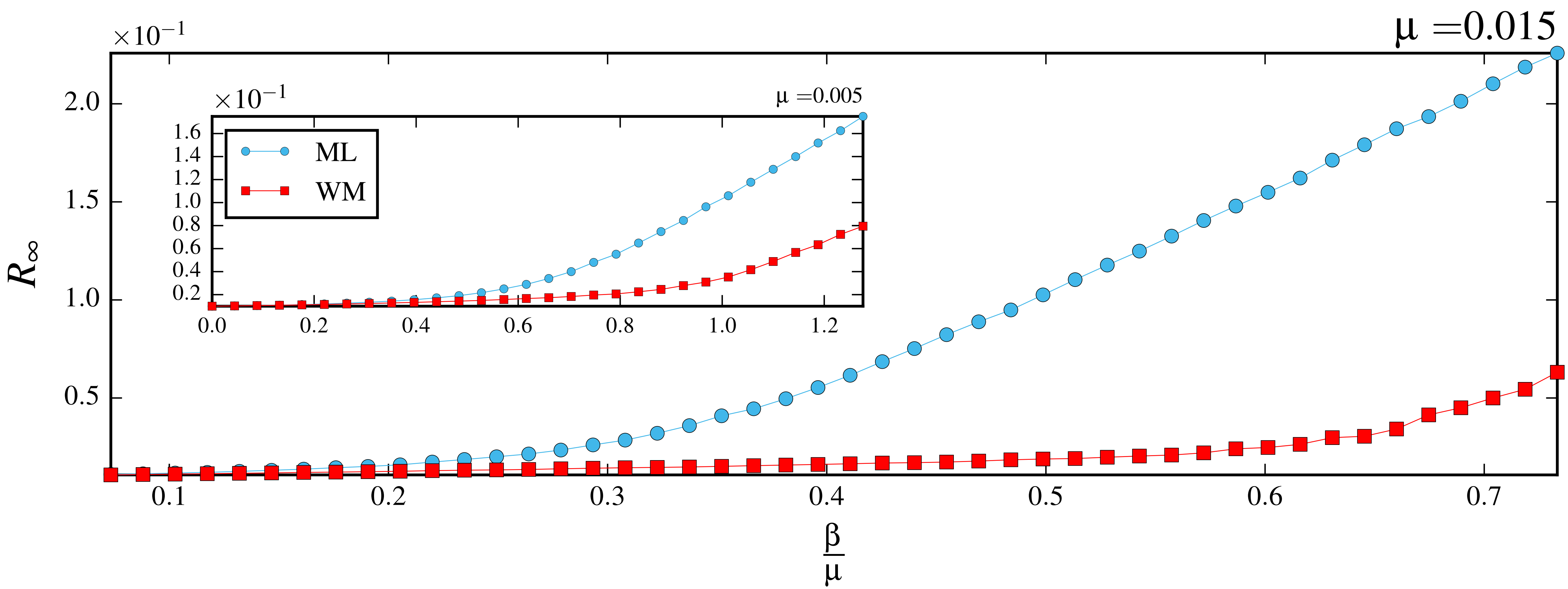
SIS model: memory favors spread



Sun et al., EPJB (2015)

SIS and SIR models
on activity-driven networks with memory

SIR model: memory hinders spread



Sun et al., EPJB (2015)

>Voter model on temporal networks

A. Moinet, A. Barrat, R. Pastor-Satorras, arXiv:1804.07476

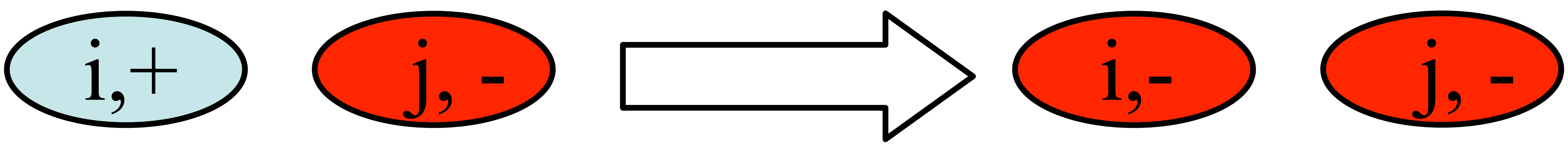
Voter model

N agents i=1,..N

Opinion s_i = 1 or -1

At each time step:

- Choose one agent i
- i chooses at random one of his neighbors j
- Agent i adopts the opinion of agent j

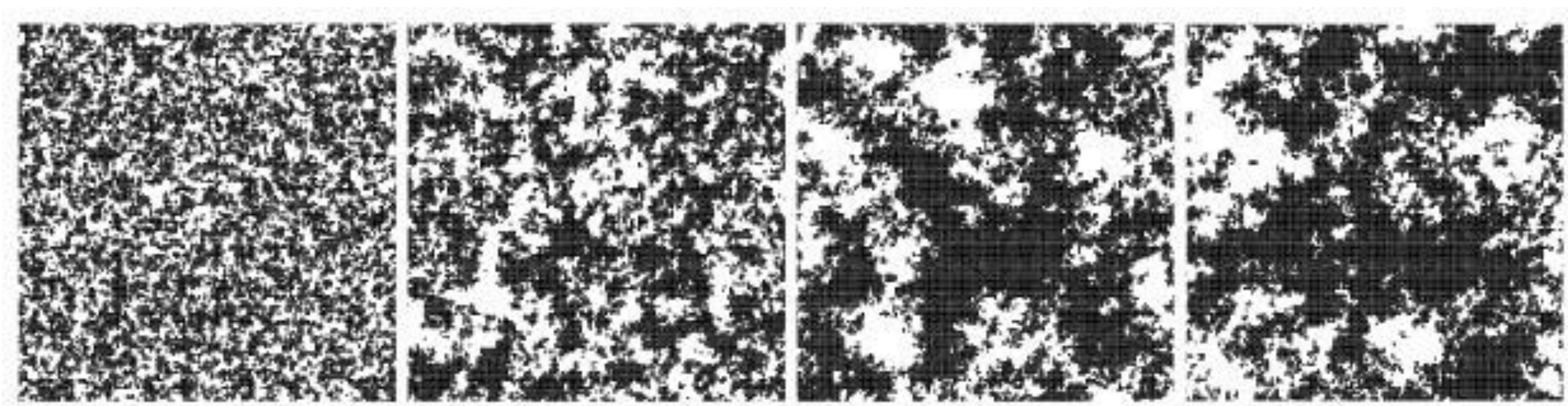


Consensus time?

Exit probability (proba that a single agent imposes its opinion to the rest of the population)?

Voter model

Example: agents sitting on a 2-dimensional lattice



"Coarsening" of domains of similar opinion agents

Voter model on static networks

$$T_N \sim \begin{cases} N & \nu > 3, \\ N/\ln N & \nu = 3, \\ N^{(2\nu-4)/(\nu-1)} & 2 < \nu < 3, \\ (\ln N)^2 & \nu = 2, \\ \mathcal{O}(1) & \nu < 2. \end{cases}$$

on scale-free networks, exponent ν
(no correlations)

Exit probability for a node of degree k: ~ k
=> **hubs have higher exit probability**, as they are chosen by another node with probability proportional to their degree

Sood, Redner, PRL 2005
Sood, Antal, Redner, PRE 2008

Voter model on activity-driven networks

Voter + Moran processes on activity-driven with attractiveness
Heterogeneous mean-field analysis

At each time step: i becomes active (depending on its activity)
and chooses j to interact with (depending on its attractiveness)

1. Voter dynamics: $s_i := s_j$ (i.e., i adopts j's state).
2. Moran dynamics: $s_j := s_i$.
3. Mixed dynamics: With probability p, $s_i := s_j$; with the complementary probability $1 - p$, $s_j := s_i$.

Moinet, Barrat, Pastor-Satorras, arXiv:1804.07476

Voter model on activity-driven networks

Voter + Moran processes on activity-driven with attractiveness
Heterogeneous mean-field analysis

Consensus time for the Voter model: $\tau = \langle a \rangle \dfrac{\langle b\, a^{-1} \rangle^2}{\langle b^2 a^{-1} \rangle}$

Exit probability for the Voter model $E_{a,b} = \dfrac{b\, a^{-1}}{N\, \langle b\, a^{-1} \rangle}$

=> nodes with large activity have small exit probability!
(but large degree in aggregated network!)

=> **opposite** result w.r.t. using an aggregated network view!

Moinet, Barrat, Pastor-Satorras, arXiv:1804.07476

> Making use of the data:
studying human behaviour

> Gender homophily from contact behavior in a primary school

J. Stehlé, F. Charbonnier, T. Picard, C. Cattuto, A. Barrat
Gender homophily from spatial behavior in a primary school: a sociometric study.
Social Networks 35(4): 604-613 (2013)

Gender homophily: a well-studied problem

Field initiated by Moreno (1953)

- From early childhood to pre-adolescence: increasing separation of boys and girls
- During adolescence: reversed tendency (romantic relationships)
- Structural differences between ego-centered networks of boys and girls
- Time shift in the evolution of same gender preference between boys and girls

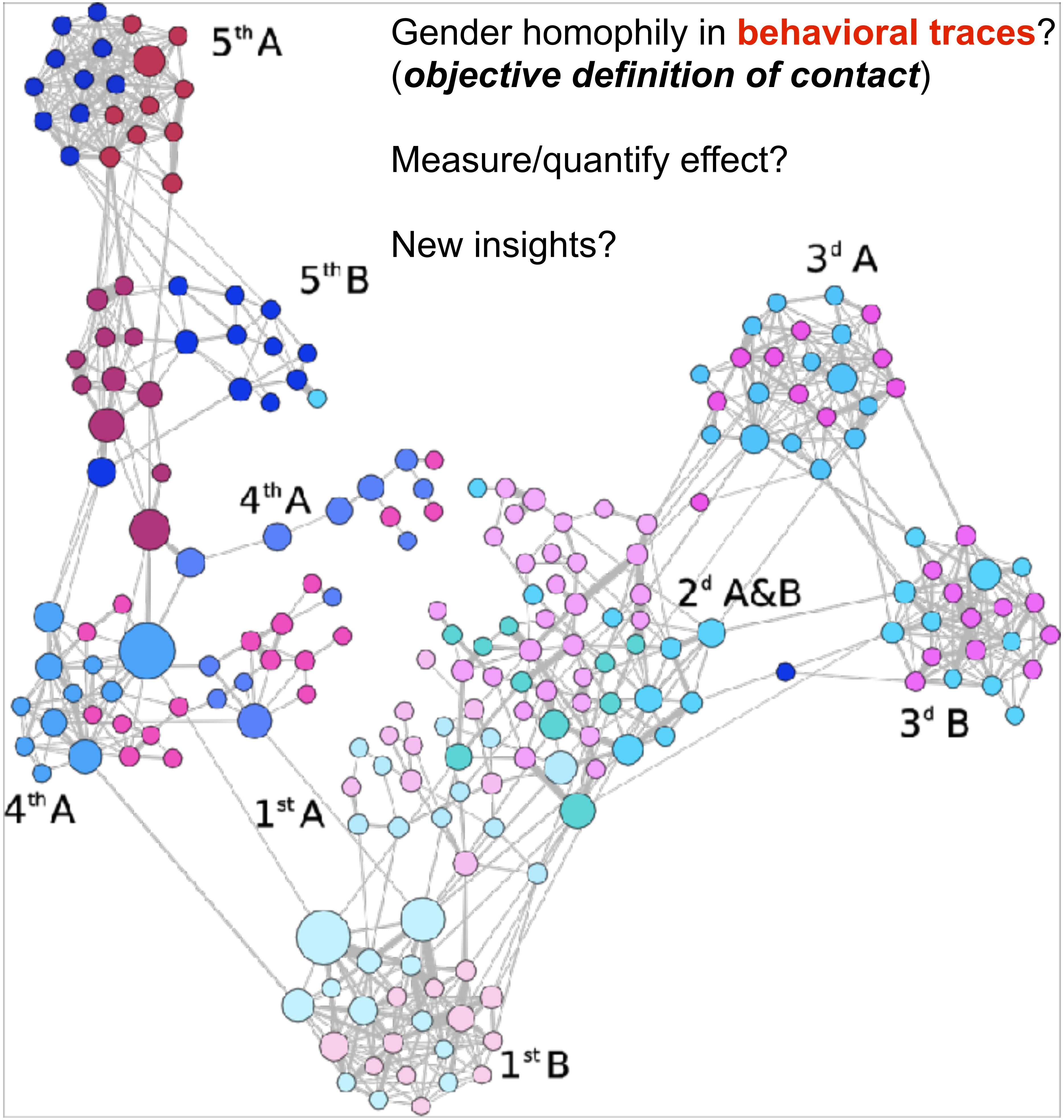
Most studies rely on:

- Questionnaires about friendships (interviews, self-administered questionnaires)
- Diaries of contacts, generally self reported
- (sometimes) direct observation by external observers

Gender homophily in behavioral traces? (***objective definition of contact***)

Measure/quantify effect?

New insights?



Methodology

- Keep only:
  - links within each class
  - contacts *outside of classrooms*
  - “strong” links (at least 5mn in contact over 2 days)

- Measure number of links of *each type*:
  - boy-boy: E_{bb}
  - girl-girl: E_{gg}
  - boy-girl: E_{bg}

- Test hypothesis of gender indifference of links, i.e., compare data with **null models**

Methodology: null models

3 variables: E_{bb} , E_{gg} , E_{bg}

Constraints on null models:

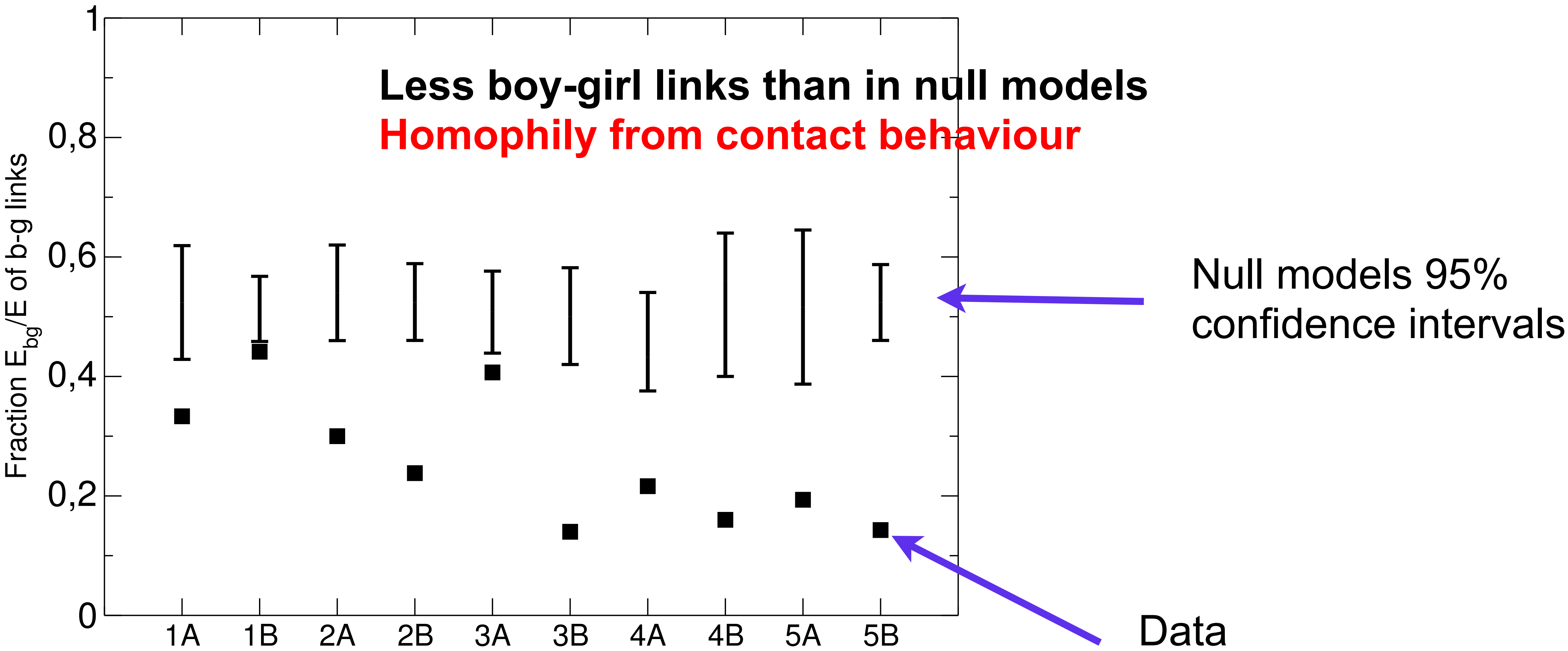
- Fixed numbers of boys and girls
- Fixed total number of links
- Different average degrees of boys and girls

one quantity left to compare (data vs null models): E_{bg}

Null models

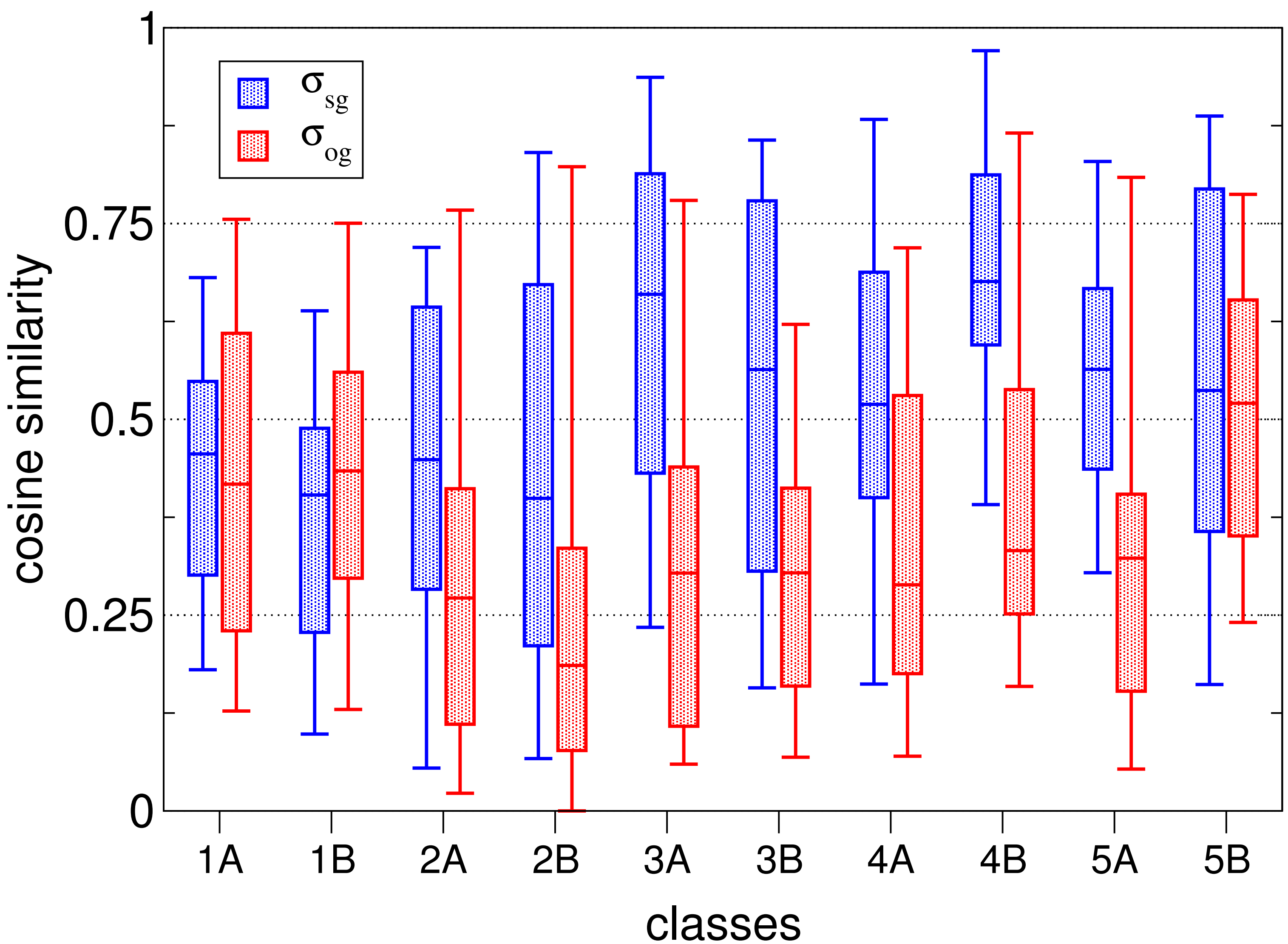
(I) Random network with same degree sequence of original data
(does not conserve degree correlations nor structures such as triangles)

(II) Fixed network structure, reshuffle gender labels among nodes
(similar to QAP)



NB: temporal resolution => further results

Temporal resolution => additional results

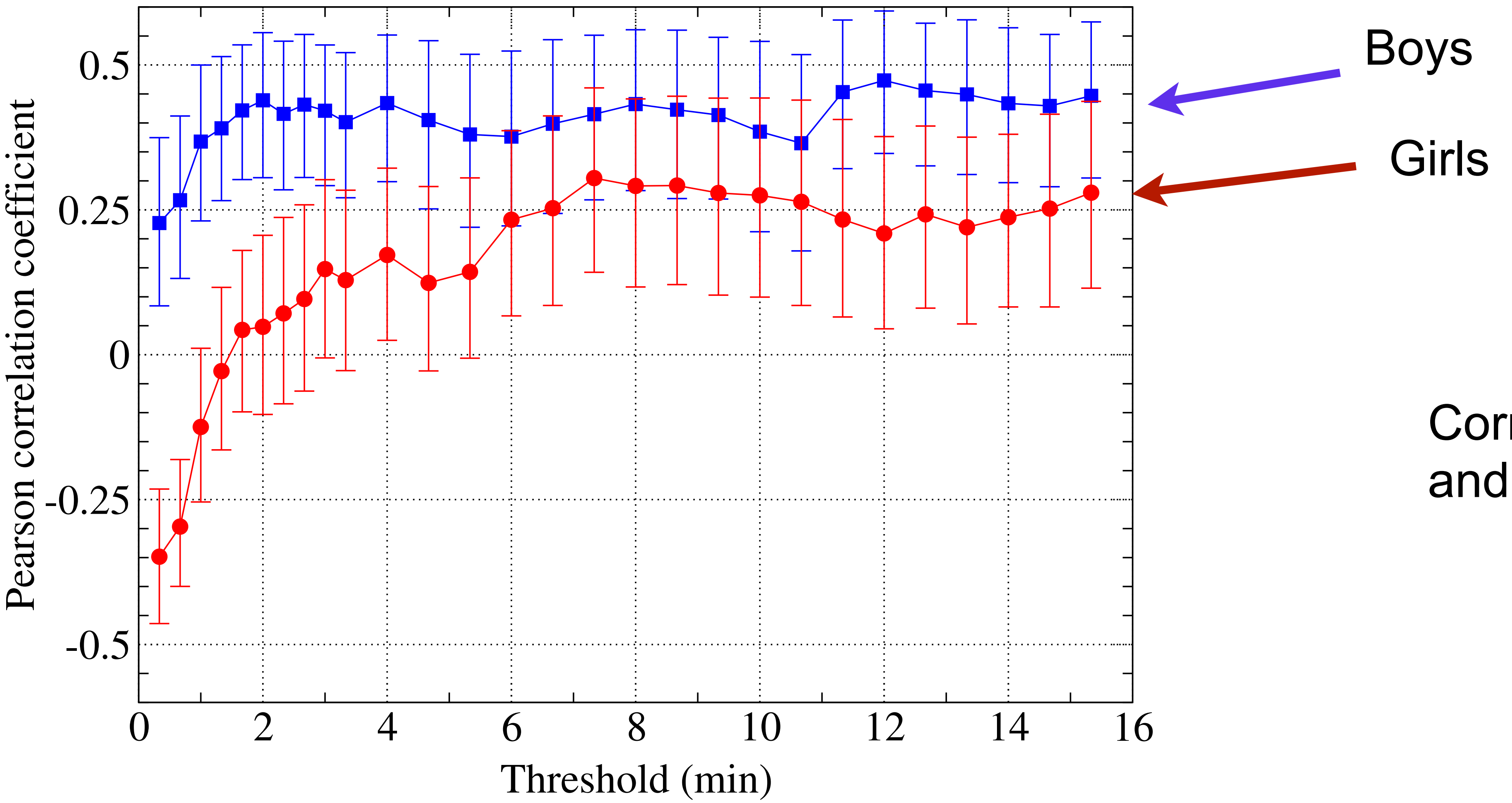


Similarity of neighborhoods from one day to the next:

same-gender neighborhood more stable than opposite-gender neighborhood

J. Stehlé, F. Charbonnier, T. Picard, C. Cattuto, A. Barrat
Gender homophily from spatial behavior in a primary school: a sociometric study.
Social Networks 35(4): 604-613 (2013)

Temporal resolution => additional results



Correlation between age and same-gender homophily

Boys: same-gender homophily *increases* with age for both strong and weak links (expected)

Girls: same-gender homophily *increases* with age for strong links (expected),
decreases with age for weak links (unexpected)
(number of opposite-gender neighbors with weak links increases)

J. Stehlé, F. Charbonnier, T. Picard, C. Cattuto, A. Barrat
Gender homophily from spatial behavior in a primary school: a sociometric study.
Social Networks 35(4): 604-613 (2013)

What about high school students?

Marseilles high school, 2012

Overall, in the contact network:

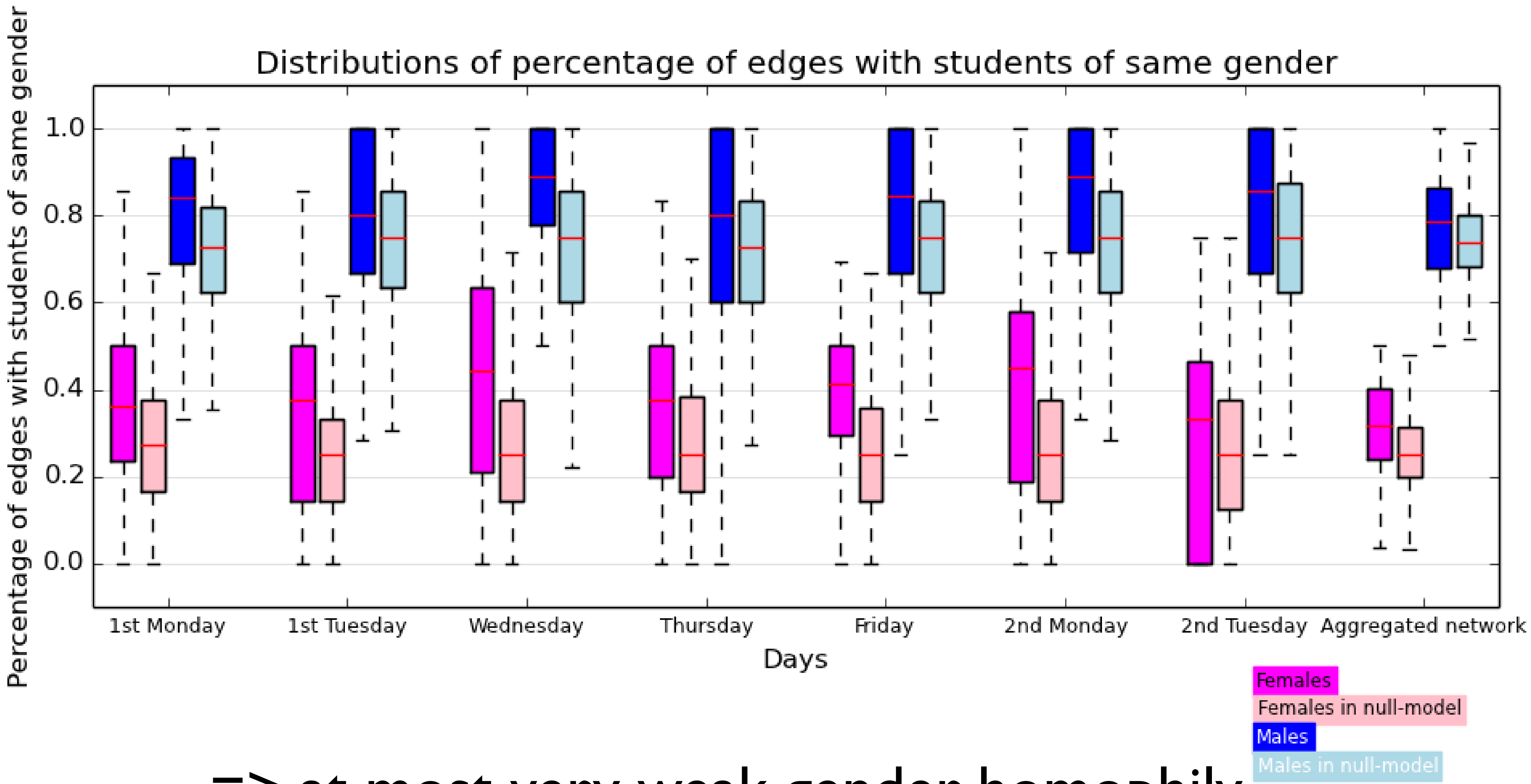
- 58% of edges between male students
- 8% of edges between female students
- 34% of edges between students of different genders

however: strong imbalance (133M, 47F)

J. Fournet, A. Barrat, PLoS ONE (2014)

What about high school students?

Comparison with null model:



=> at most very weak gender homophily

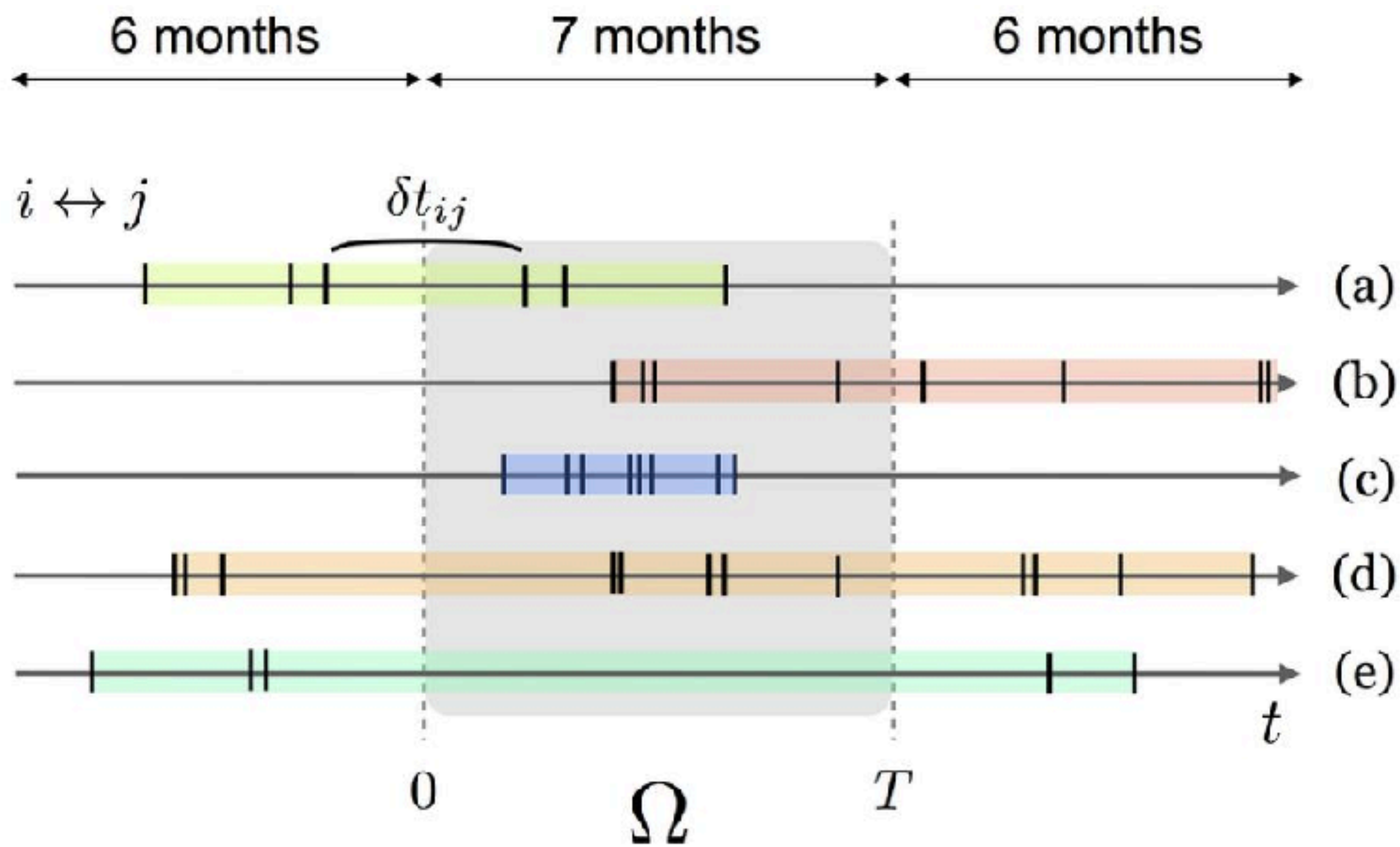
J. Fournet, A. Barrat, PLoS ONE (2014)

“Limited communication capacity unveils strategies for human interaction”

Miritello, Lara, Cebrian, Moro, Sci. Rep. 3:1950 (2013)

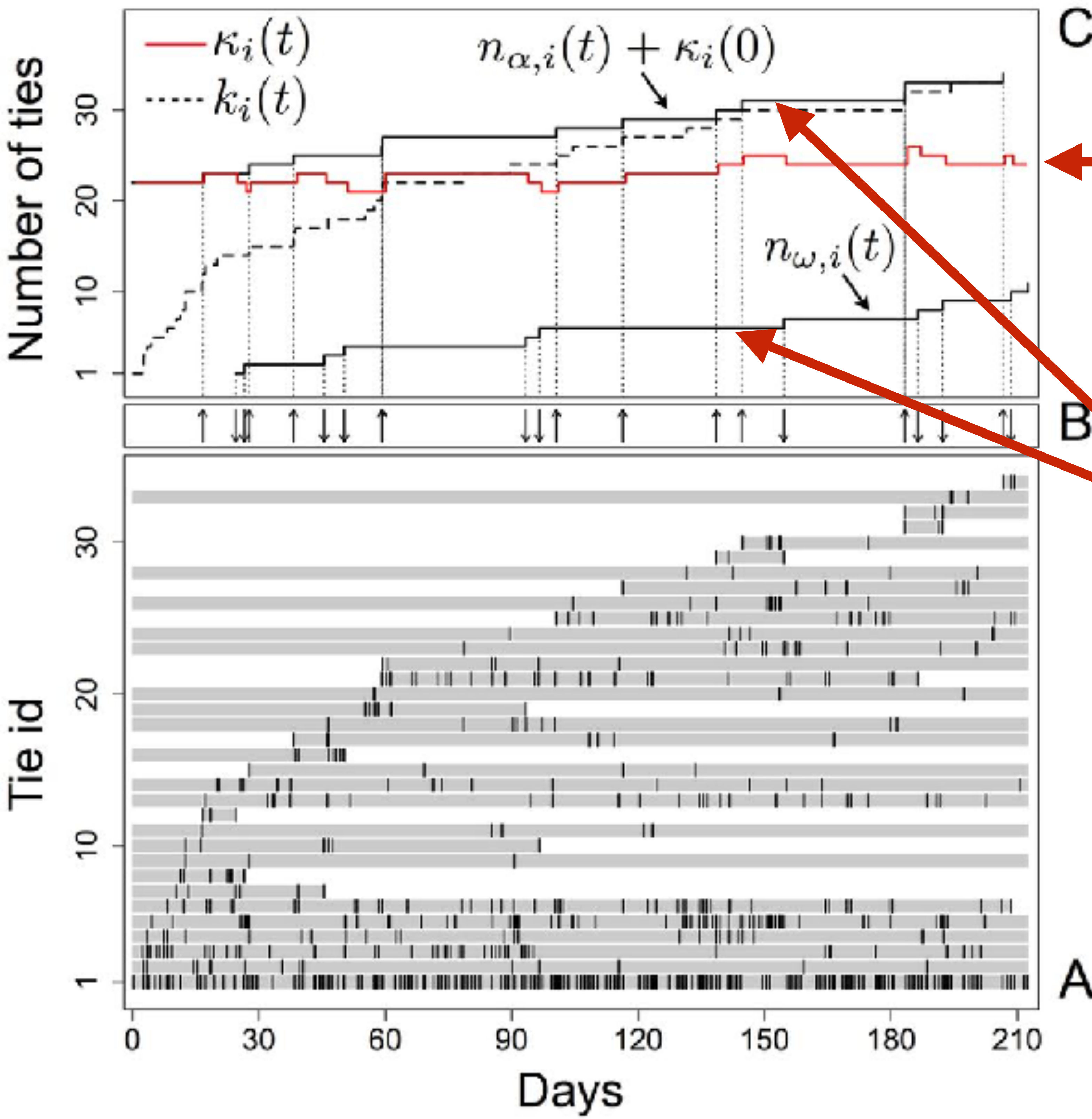
Limited communication capacity

Call Detail Records of 20 million users over 19 months



Miritello et al., Sci. Rep. (2013)

Limited communication capacity



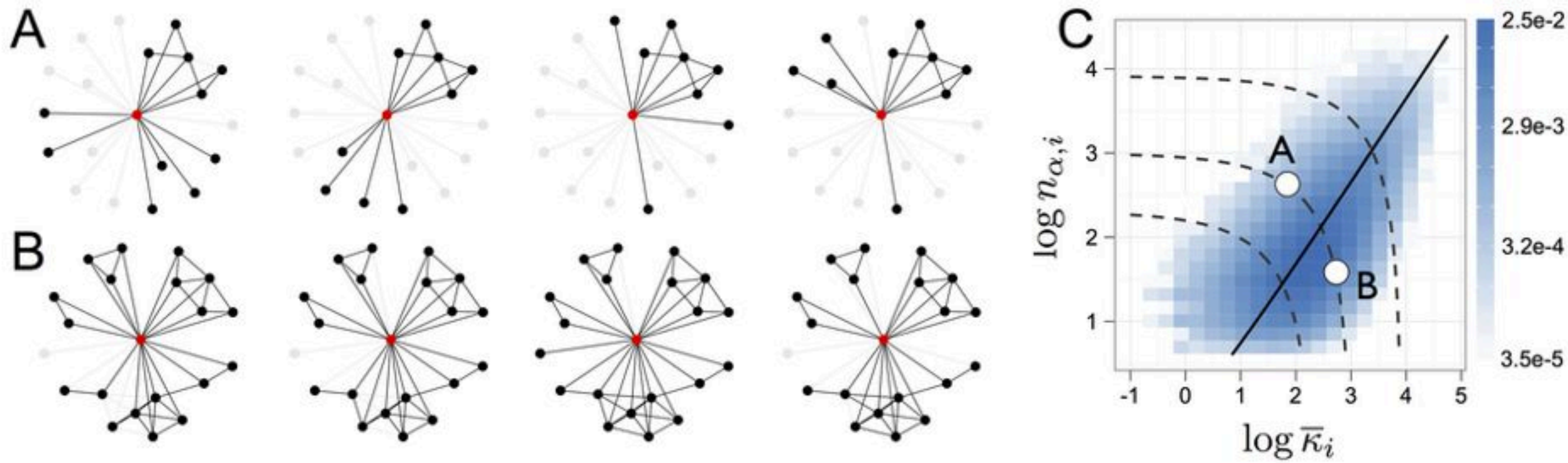
remains roughly constant even if ties are activated and deactivated

number of new alters and number of disappearing alters grow linearly and at the same speed, for a given individual

Miritello et al., Sci. Rep. (2013)

Limited communication capacity and human communication strategies

Social explorers vs. social keepers



+socio-demographic dependence
+impact on information diffusion

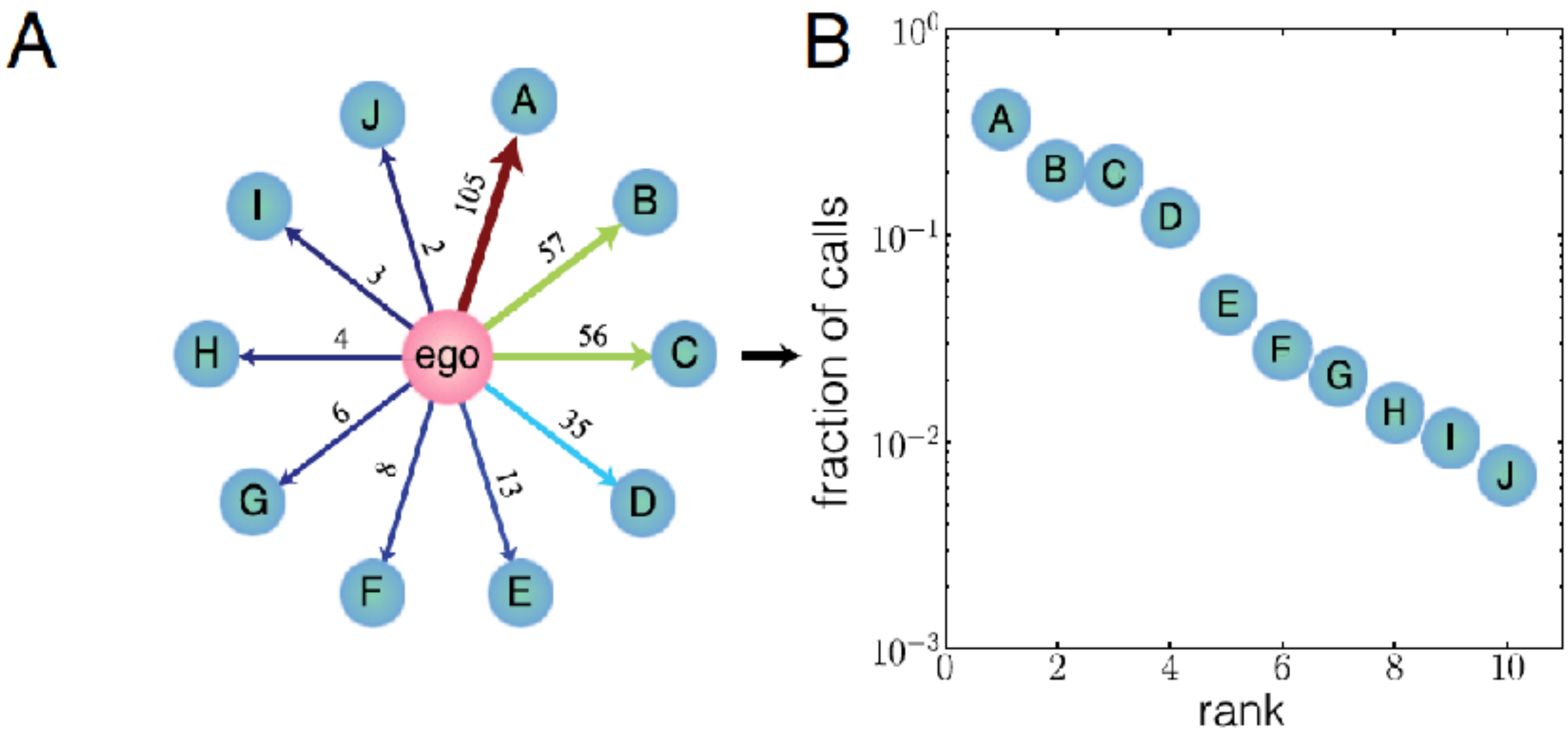
Miritello et al., Sci. Rep. (2013)

“Persistence of social signatures in human communication”

J. Saramäki et al., PNAS 111, 942 (2014)

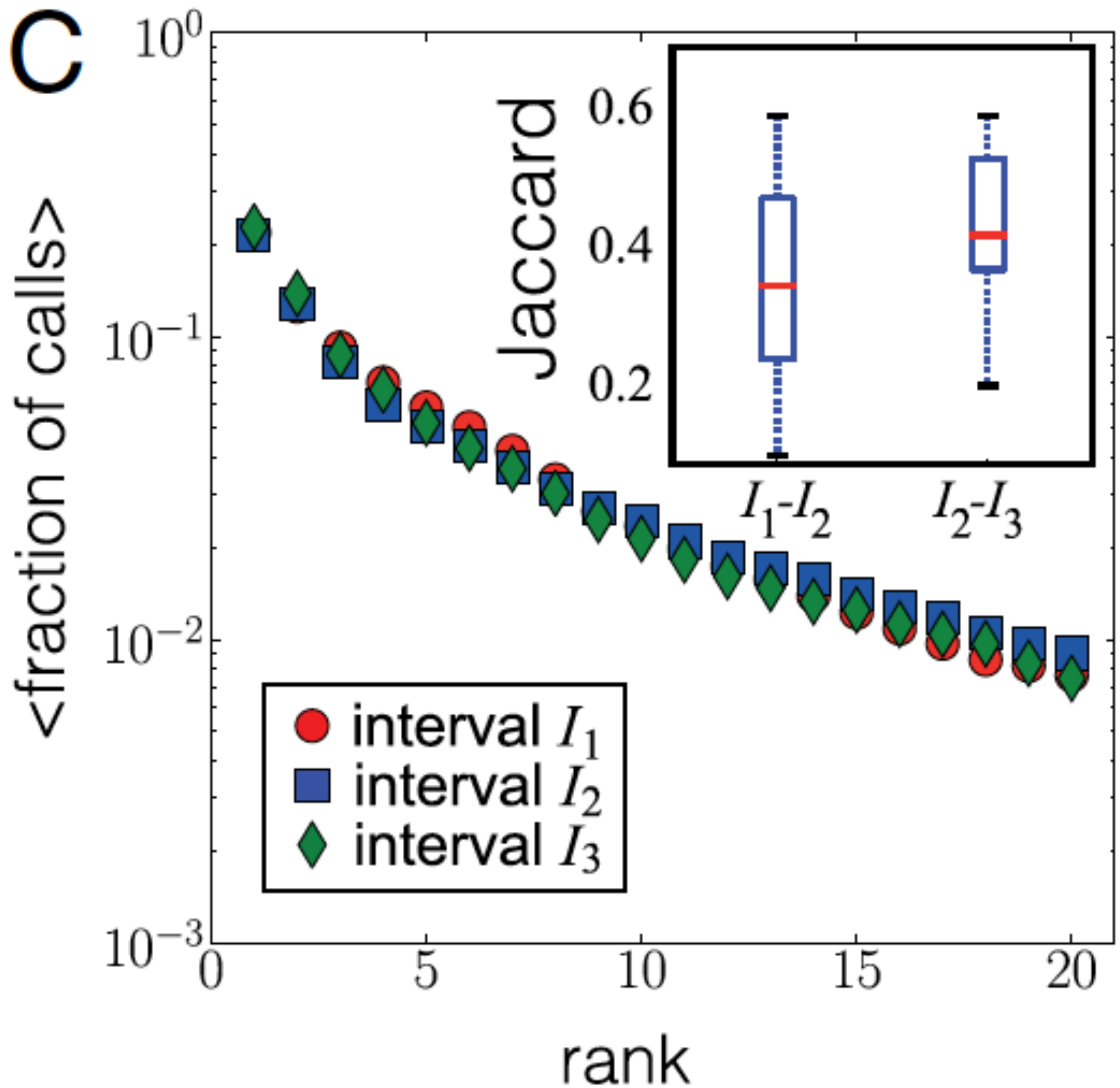
Social signatures in human communication

Data: 18 months dataset, phone calls + surveys,
24 students followed from school to university



Saramäki et al, PNAS 2014

Social signatures in human communication



Robust "signature" despite turnover in neighbourhoods

Saramäki et al, PNAS 2014

Social signatures in human communication

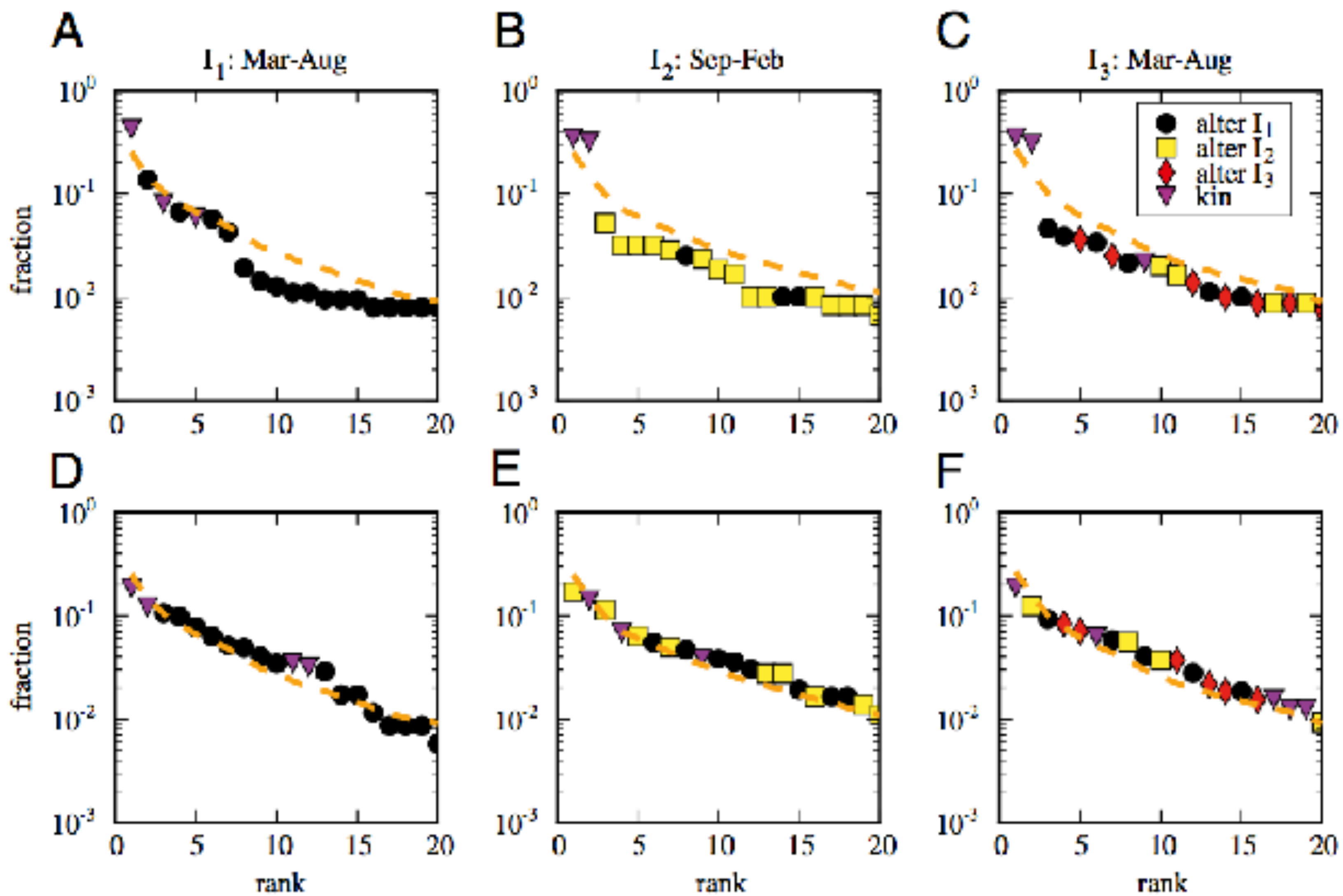


Fig. 2. Individual-level variation in social signatures and their evolution. *Upper* (*A–C*) and *Lower* (*D–F*) depict the time evolution of the social signatures of two different male participants who both went to university in another city. The symbols correspond to alters observed for the first time in intervals I_1 (circles), I_2 (squares), and I_3 (diamonds) or to kin (triangles) as reported by the egos. The large turnover in the networks of the participants is clearly visible. The dashed line indicates the social signature averaged over all 24 egos. In the social signatures depicted in *A–C*, two kin alters receive a higher-than-average fraction of outbound calls, whereas the signatures *D–F* do not deviate much from the average. In both cases, this individual-level variation persists through all time intervals.

Saramäki et al, PNAS 2014

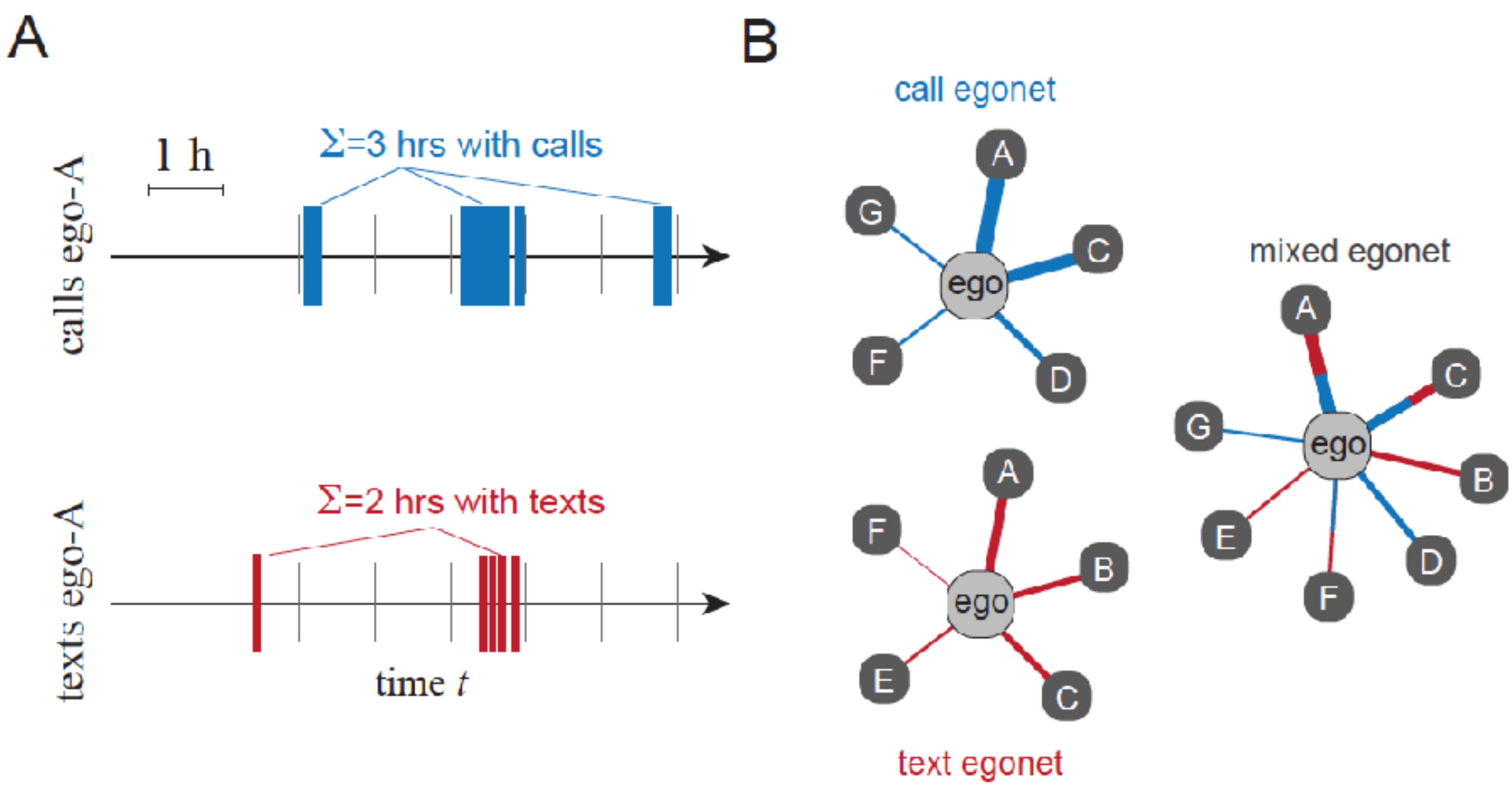
What about different communication channels or social network proxies?

Data at hand => “the” social network of the population (?)

- communications (cell phones, messages, e-mails…)
- co-presence (various spatial resolutions)
- surveys
- online interactions
- …

however, typically: correlated but different

What about different communication channels or social network proxies?



Call and messaging networks differ

Heydari et al., arXiv:1806.02641

What about different communication channels or social network proxies?

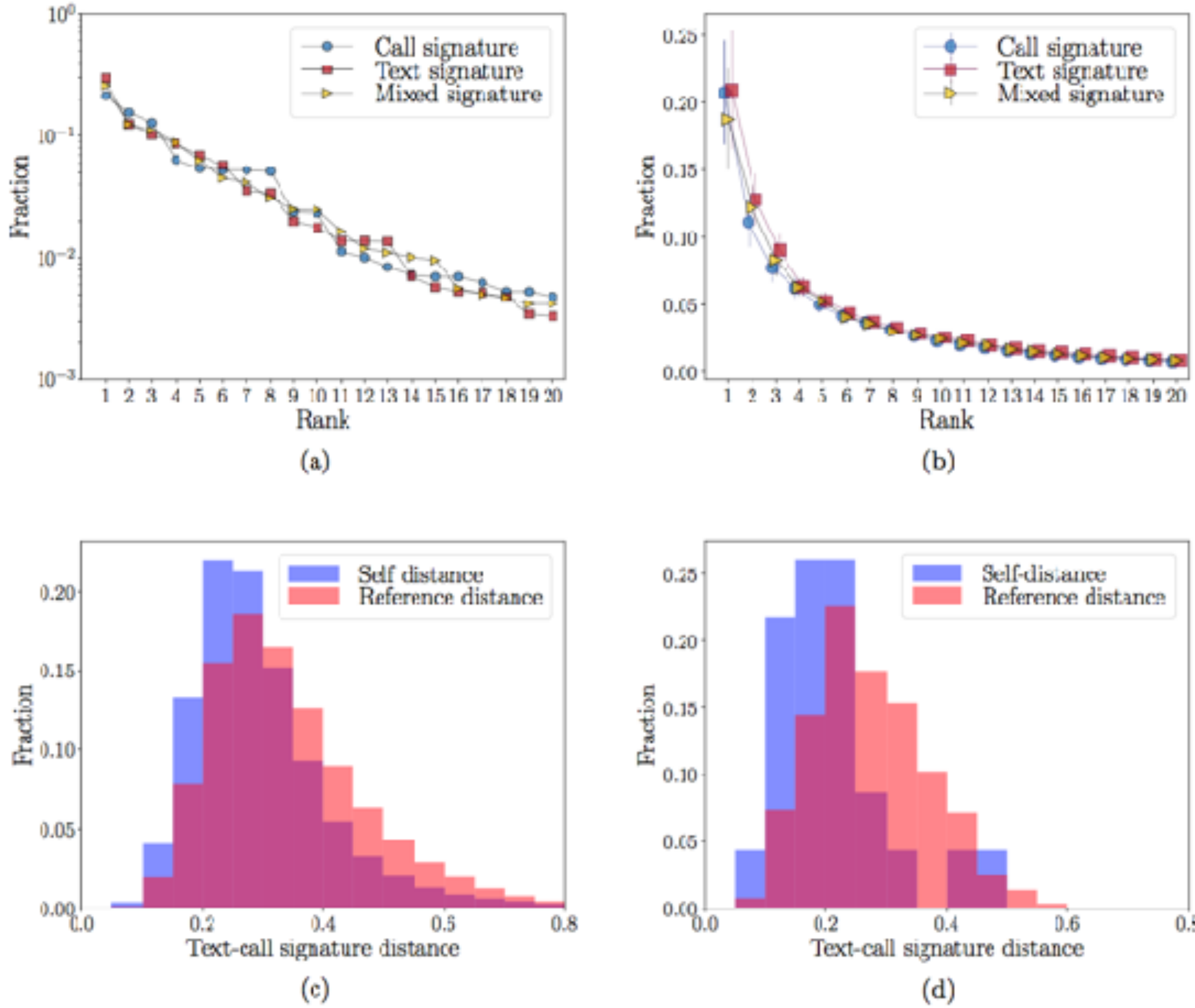


FIG. 3: The similarities of social signatures of different types. Panel (a) shows the call, text and mixed signatures of one person in the Dataset 1. The three signatures look similar. Panel (b) illustrates the average signatures over the population in Dataset 2. The population-level signatures are also fairly similar. Panels (c) and (d) compare the distance distributions of the call and text signatures of same egos with the distributions of call and text signatures of different people as a reference. The call and text signatures of each ego are more similar than pairs of signatures of different people.

Similar signatures for different communication channels of an ego

Heydari et al., arXiv:1806.02641

What about different communication channels or social network proxies?

Another example: Do different layers of a social network convey the same information w.r.t. homophily patterns?

Data on

- Calls
- Surveys
- Co-presence

for students in Singapore over one academic year

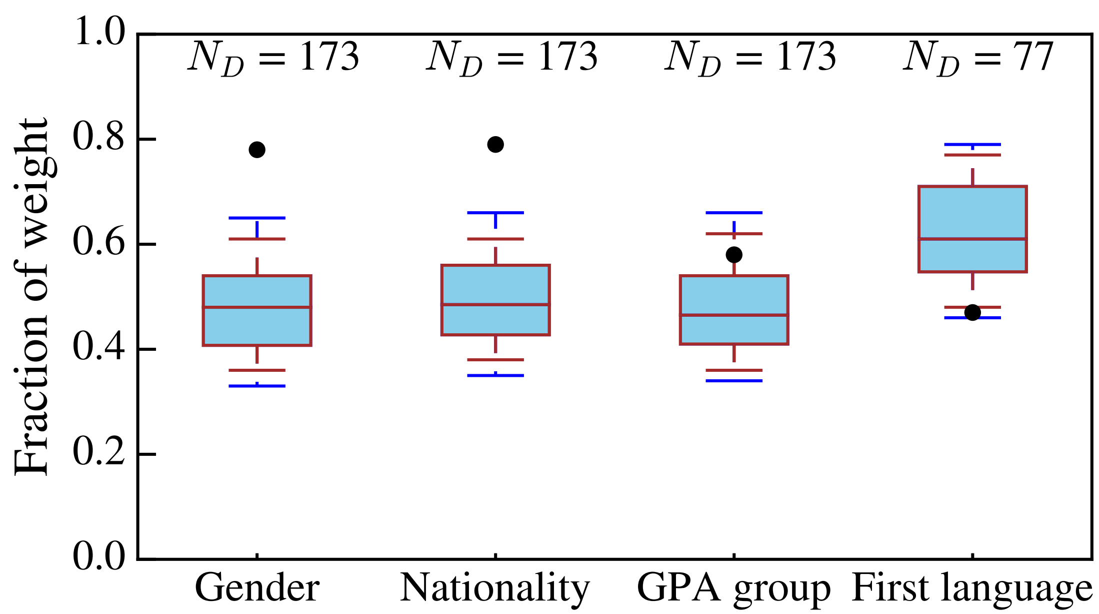
Manivannan et al., arXiv:1801.02874

What about different communication channels or social network proxies?

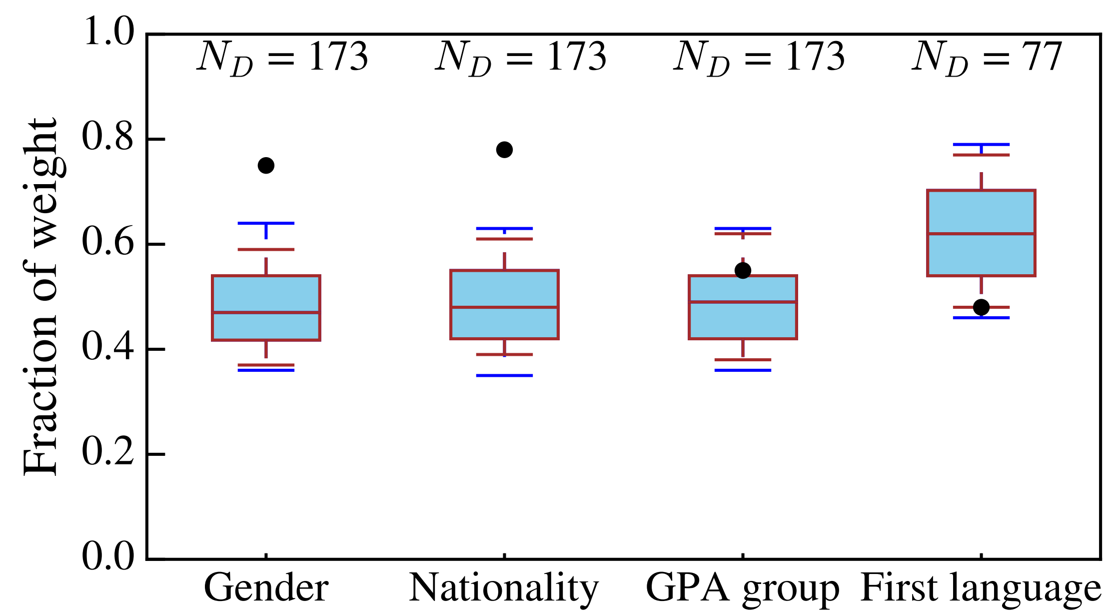
Homophily assessed with respect to a null model

Yearly aggregated network:

(a) Communication network – Aggregated duration



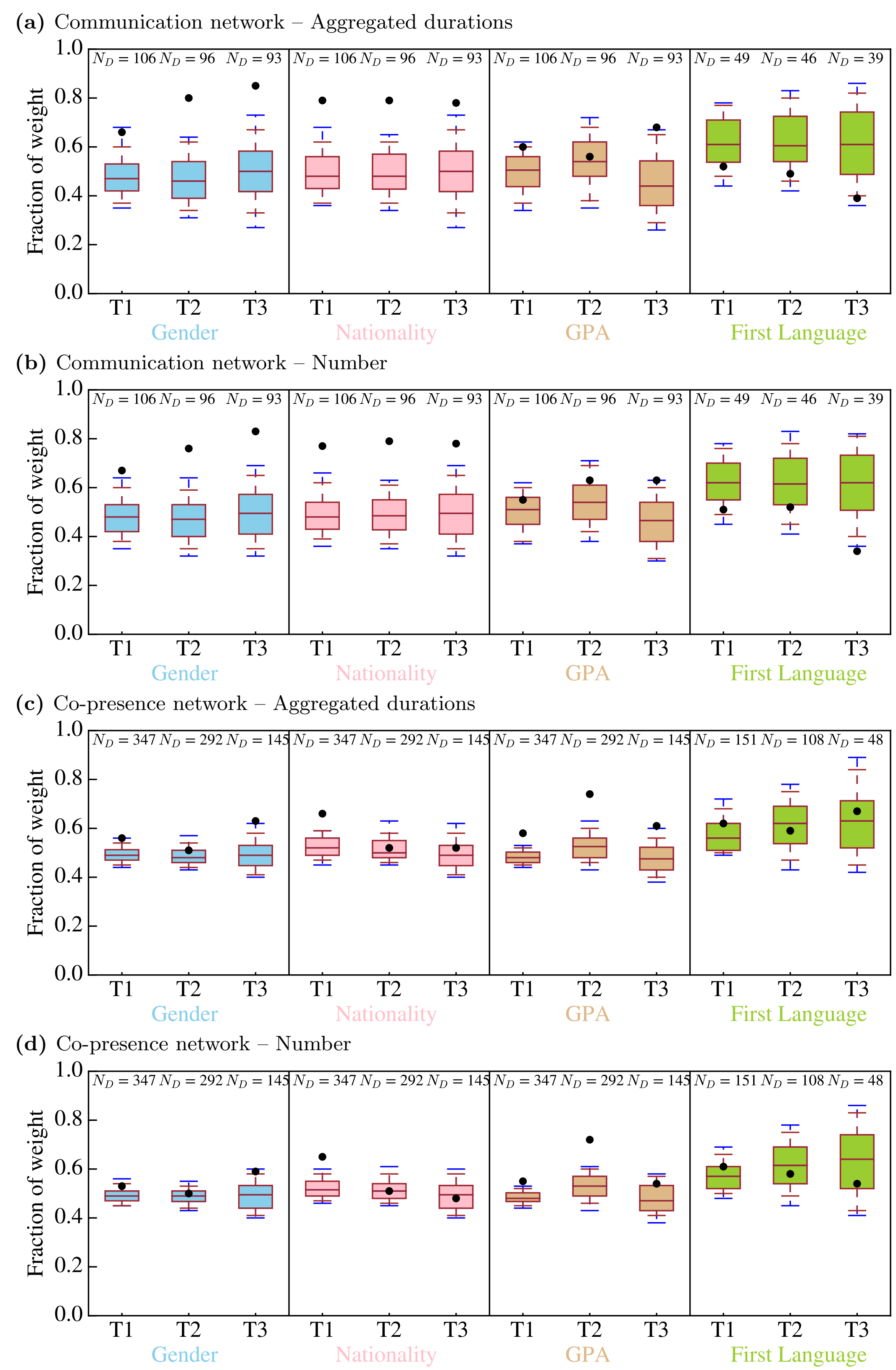
(b) Communication network – Number of calls



Manivannan et al., arXiv:1801.02874

What about different communication channels or social network proxies?

Term aggregated networks: communication vs. co-presence



Manivannan et al., arXiv:1801.02874

What about different communication channels or social network proxies?

Table 4: **Summary of the dyadic homophily patterns found in the different networks, with respect to the various attributes considered.**

Attribute	Weight	Communication T1	Communication T2	Communication T3	Co-presence T1	Co-presence T2	Co-presence T3	Weight	Questionnaires T0	Questionnaires T1	Questionnaires T2	Questionnaires T3
Gender	Volume	S	VS	VS	S	No	VS	Q1	VS	S	S	VS
	Number	VS	VS	VS	W	No	S	Q2	S	VS	VS	VS
Nationality	Volume	VS	VS	VS	VS	No	No	Q1	VS	VS	VS	VS
	Number	VS	VS	VS	VS	No	No	Q2	VS	W	S	W
GPA	Volume	W	No	VS	VS	VS	VS	Q1		VS	W	W
	Number	No	W	S	VS	VS	W	Q2		VS	No	W
First Language	Volume	W_{het}	W_{het}	S_{het}	No	No	No	Q1	No	No	No	No
	Number	W_{het}	W_{het}	VS_{het}	No	No	No	Q2	No	No	W_{het}	No
Loneliness	Volume	W_{het}	W_{het}	No	S_{het}	No	No	Q1	VS	No	No	No
	Number	W_{het}	No	No	W_{het}	No	No	Q2	W_{het}	No	No	No
SACQ	Volume	W_{het}	W_{het}	W	W	S_{het}	W_{het}	Q1	VS	W_{het}	No	No
	Number	No	No	W	S	W_{het}	W_{het}	Q2	W	No	No	No
Classroom Community	Volume	W_{het}	W_{het}	No	W	No	No	Q1	No	No	No	W
	Number	W_{het}	No	No	W	W	No	Q2	W_{het}	W	No	No

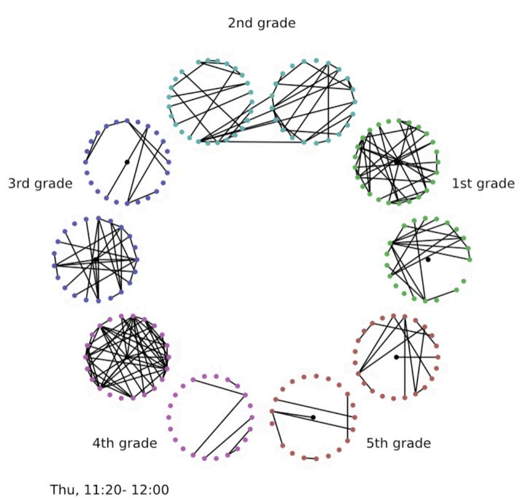
Comparison with a suitable null model
=> communication vs questionnaires: more similar than expected in the null model
=> communication vs co-presence: not more similar than random

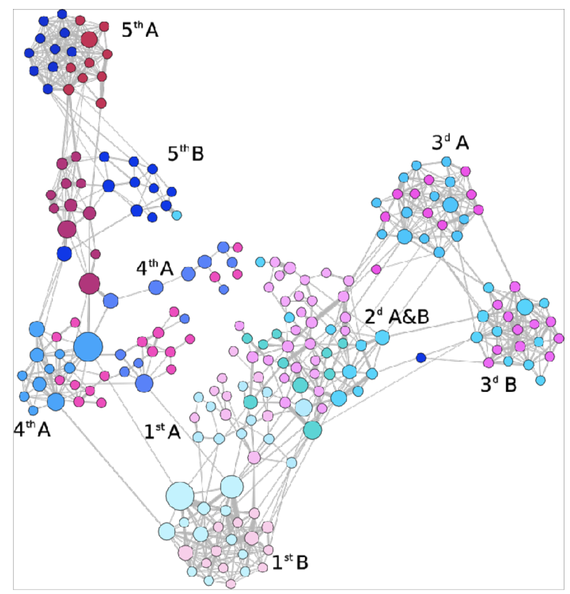
Manivannan et al., arXiv:1801.02874

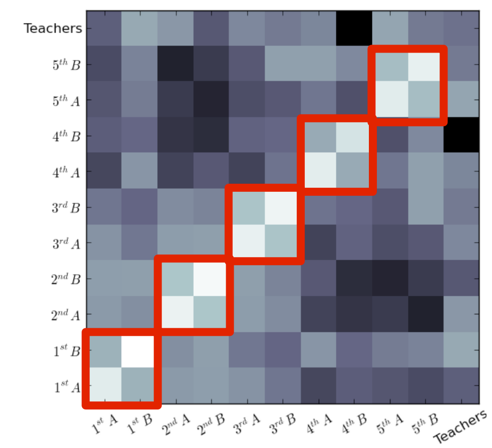
> Another way to use data: Data-driven numerical simulations

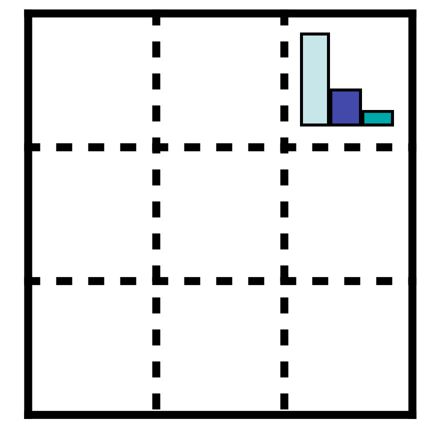
What level of detail, which data representation to use in data-driven simulations?

Various possible representations to feed models









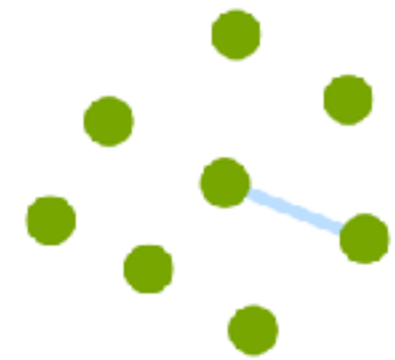
+

J. Stehlé et al., BMC Medicine (2011); A. Machens et al, BMC Inf Dis (2013)

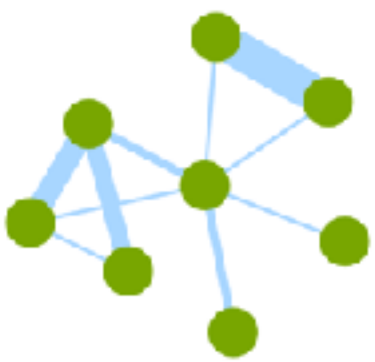
>A concrete example: contact patterns in a hospital

Network representations

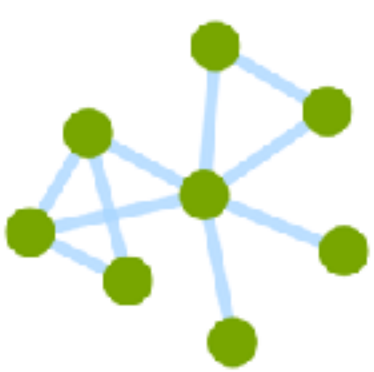
Construction of 3 networks:



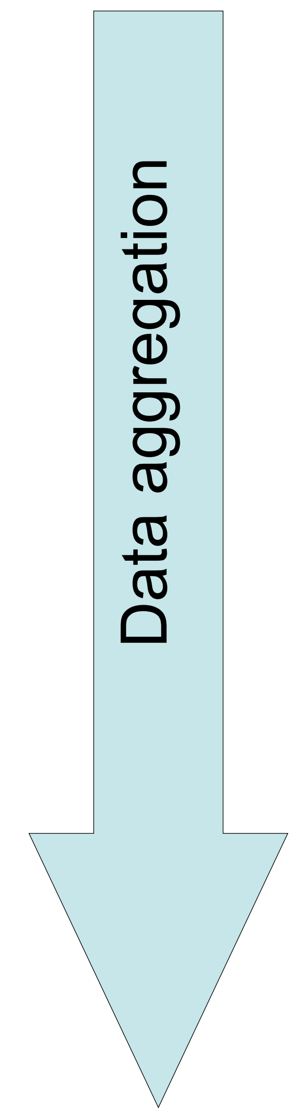
1. Dynamic network (DYN):
   Real sequence of successive contacts



2. Heterogeneous network (HET):
   1-day aggregated network
   A—B if A and B have been in contact
   W_{AB} = cumulative duration of the contacts A-B



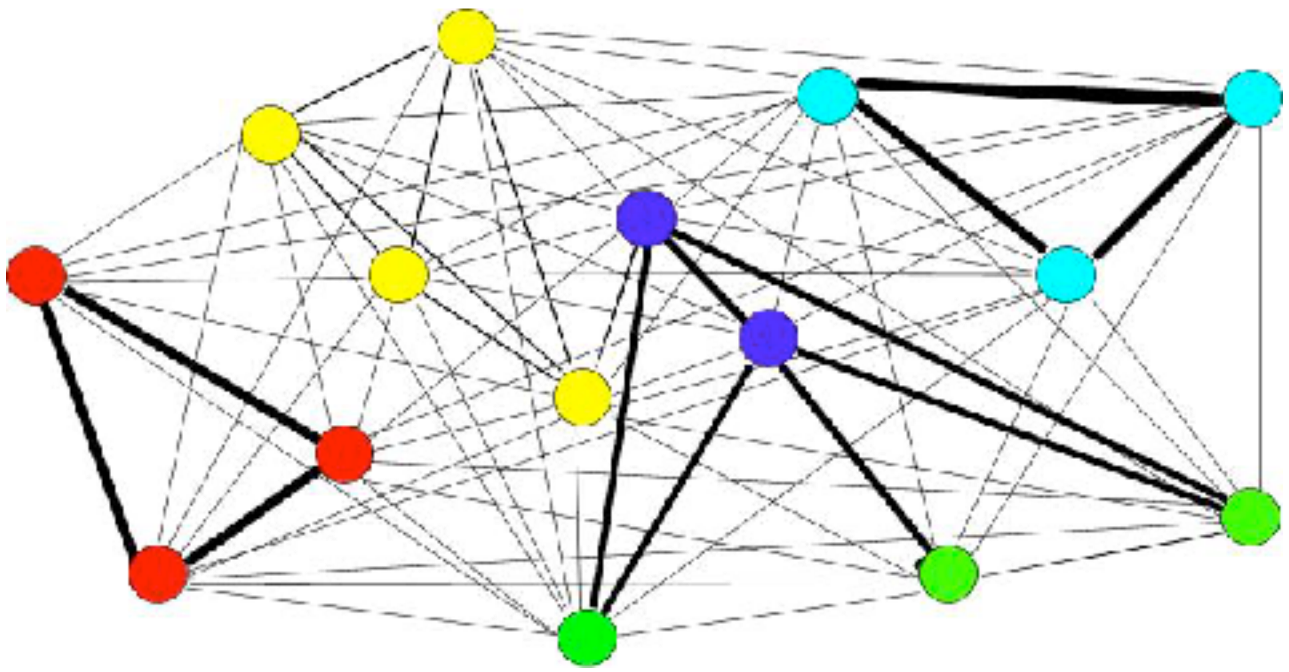
3. Homogeneous network (HOM):
   1-day aggregated network
   A—B if A and B have been in contact
   W_{AB} = average cumulative duration



Networks:

- Take into account network structure at the individual level
- Difficult to generalize

4. Contact matrix

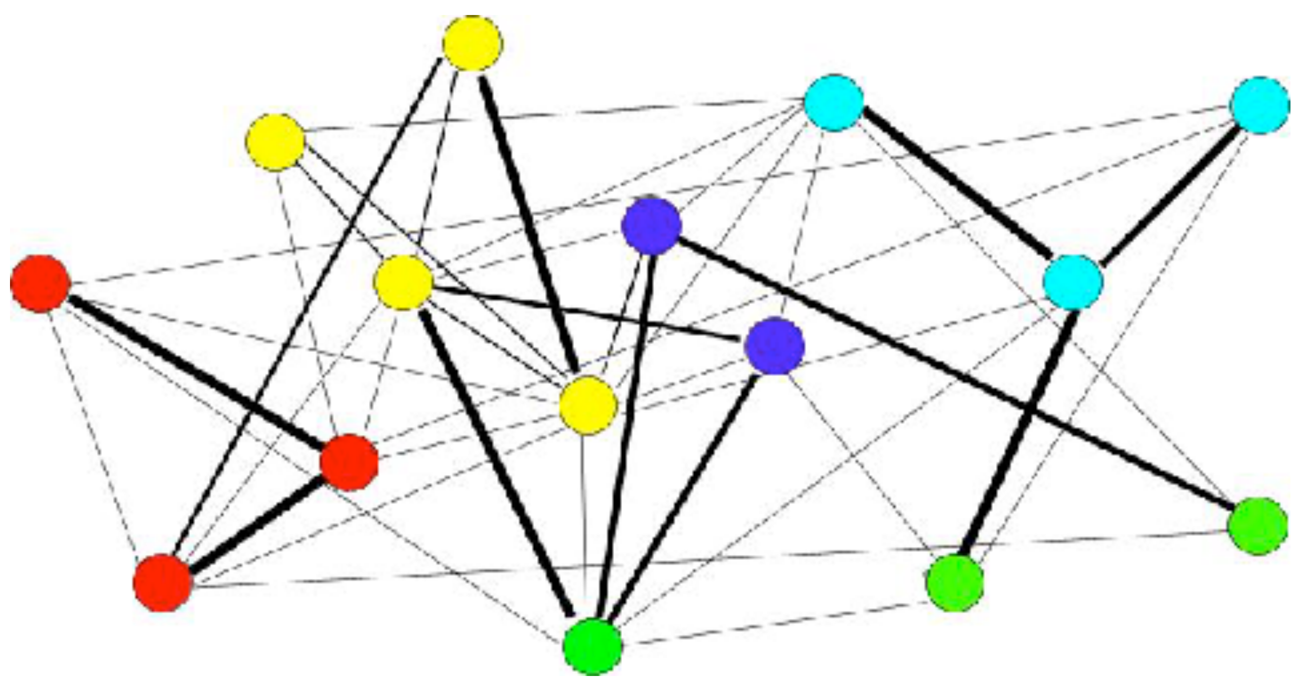


Average contact time in seconds per day

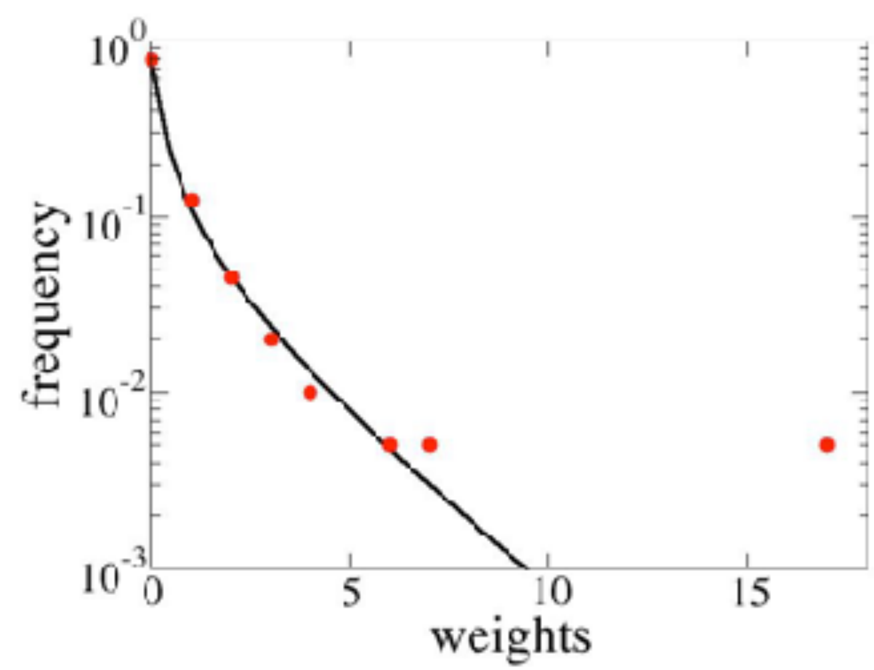
w_{AB}	Assistants	Doctors	Nurses	Patients	Caregivers
Assistants	298	1.16	24.7	0.95	1.92
Doctors	1.16	20.8	3.99	0.95	1.20
Nurses	24.7	3.99	47.3	2.32	2.57
Patients	0.95	0.95	2.32	1.27	46.9
Caregivers	1.92	1.20	2.57	46.9	1.80

- Underlying fully connected network structure
- Takes into account role structure
- Average temporal information, no heterogeneities within each role
- Easy to generalize

5. Intermediate representation: Contact matrix of distributions



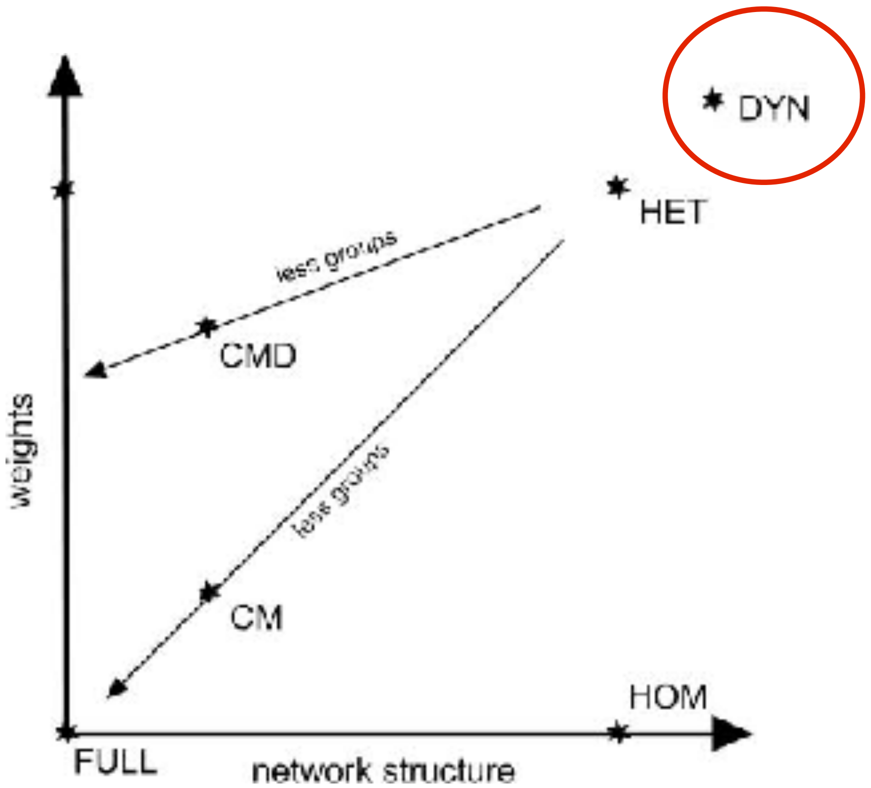
a. Fit each role-pair distribution of weights
(using negative binomials)
b. Create a network in which weights are drawn
from the fitted distribution (NB: including zero weights)



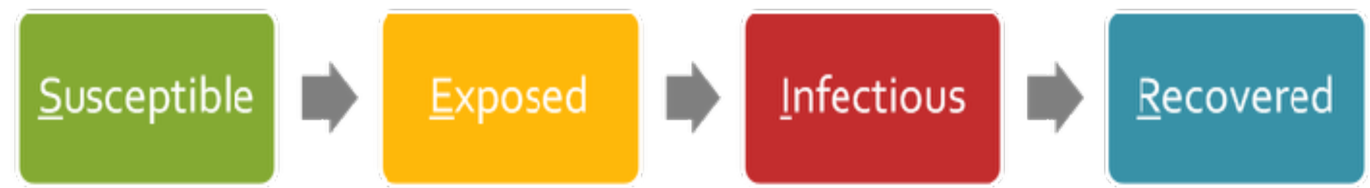
- Underlying realistic network structure
- Takes into account role structure
- Takes into account heterogeneities within each role
- Easy to generalize

Machens et al, BMC Infect. Dis.(2013)

Evaluating the representations?



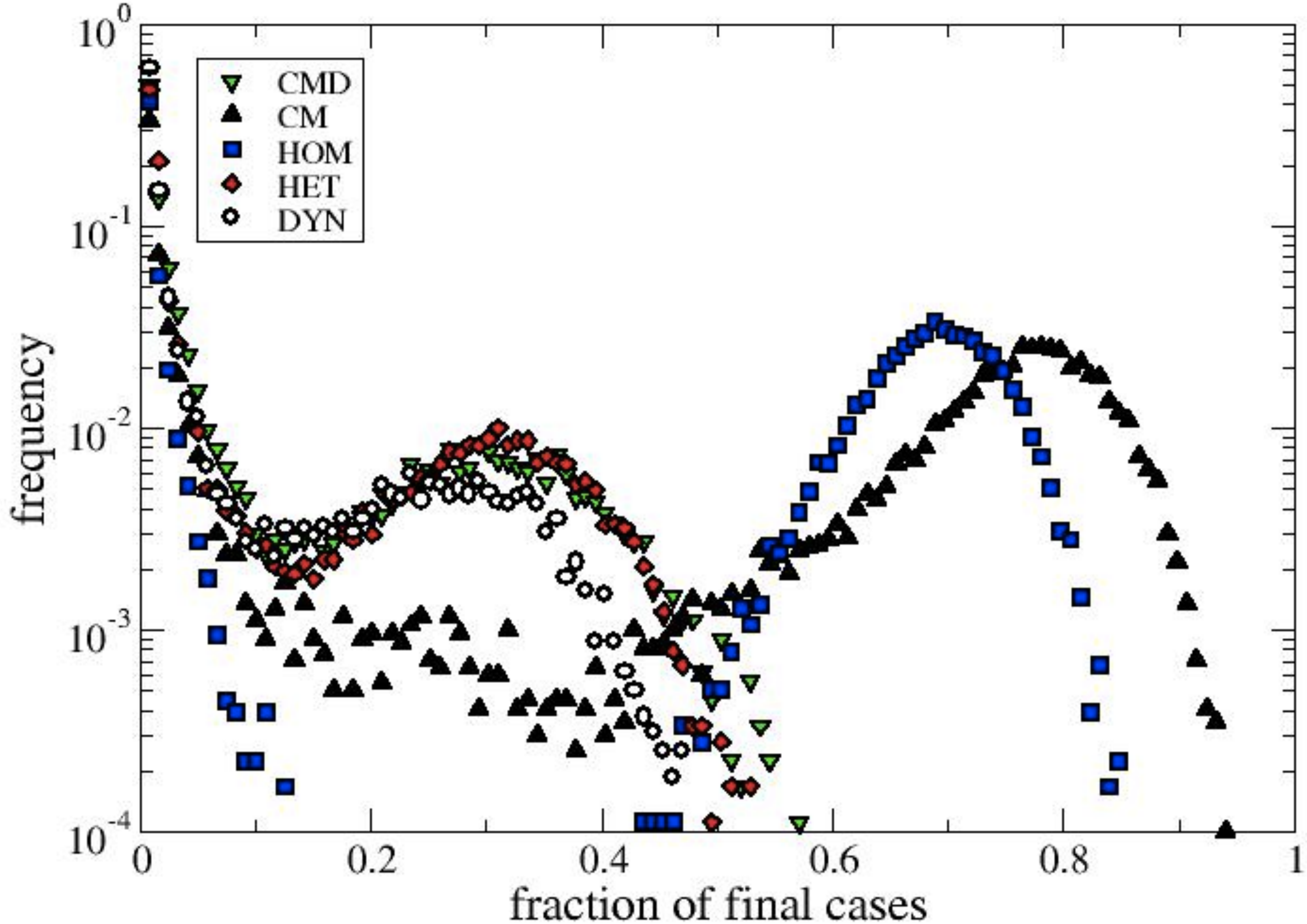
Epidemiological model



- Evaluation of :
  - Extinction probability
  - Attack rate
  - Role of initial seed
  - Attack rate for each group
- Comparison with most realistic DYN representation

Machens et al, BMC Infect. Dis.(2013)

SEIR simulation results

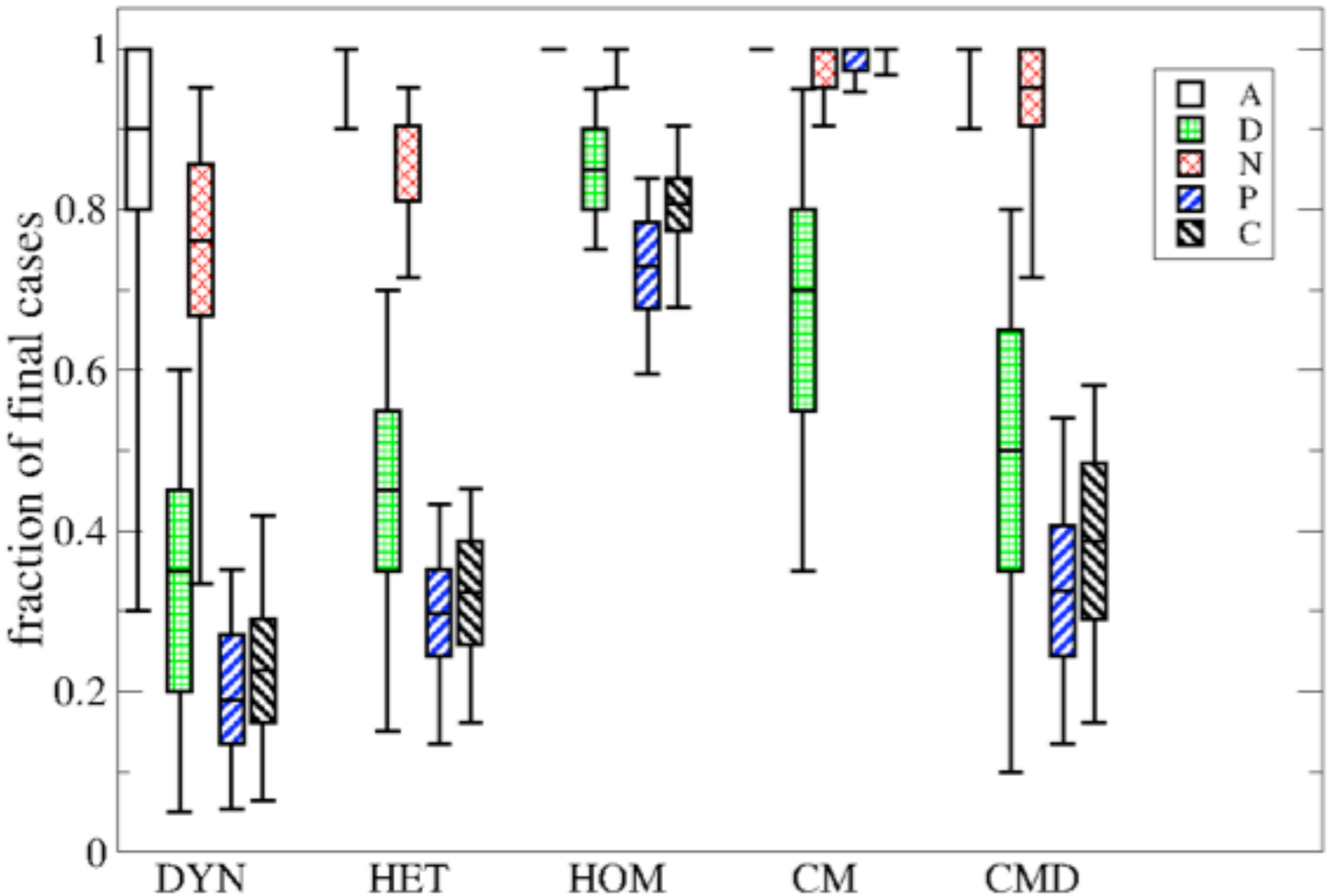


Importance of heterogeneities of contact durations

Machens et al, BMC Infect. Dis.(2013)

SEIR simulation results

Attack rate by groups (for AR > 10%)

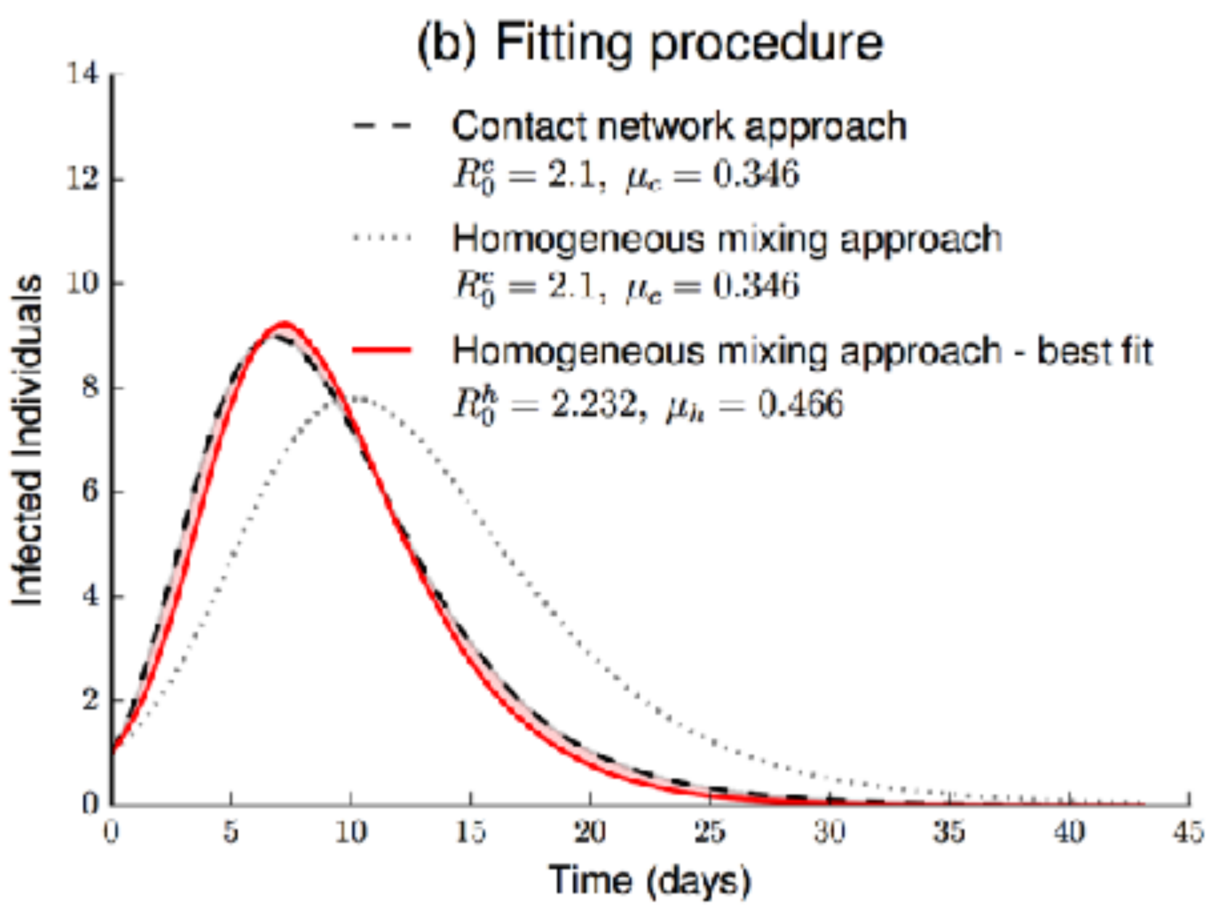


Machens et al, BMC Infect. Dis.(2013)

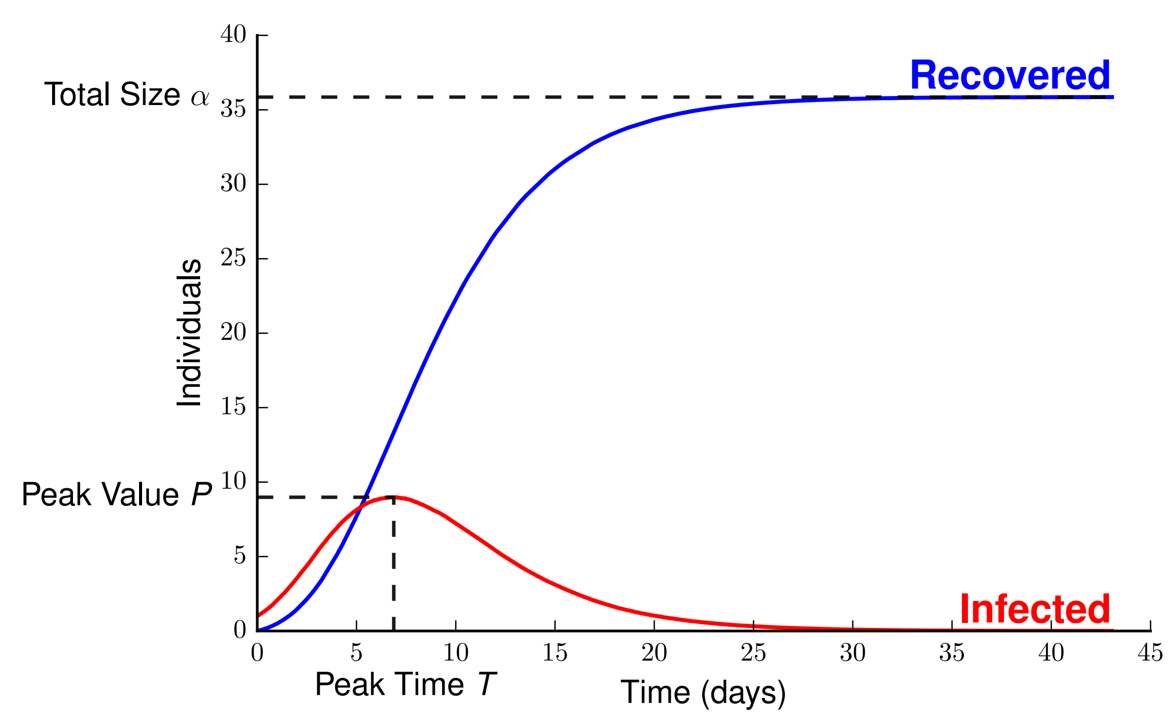
> Recalibrating disease parameters

In simulations, can we replace a realistic contact structure by a homogeneous mixing hypothesis, at the cost of recalibrating parameters?

Bioglio et al., BMC Inf. dis. (2016)



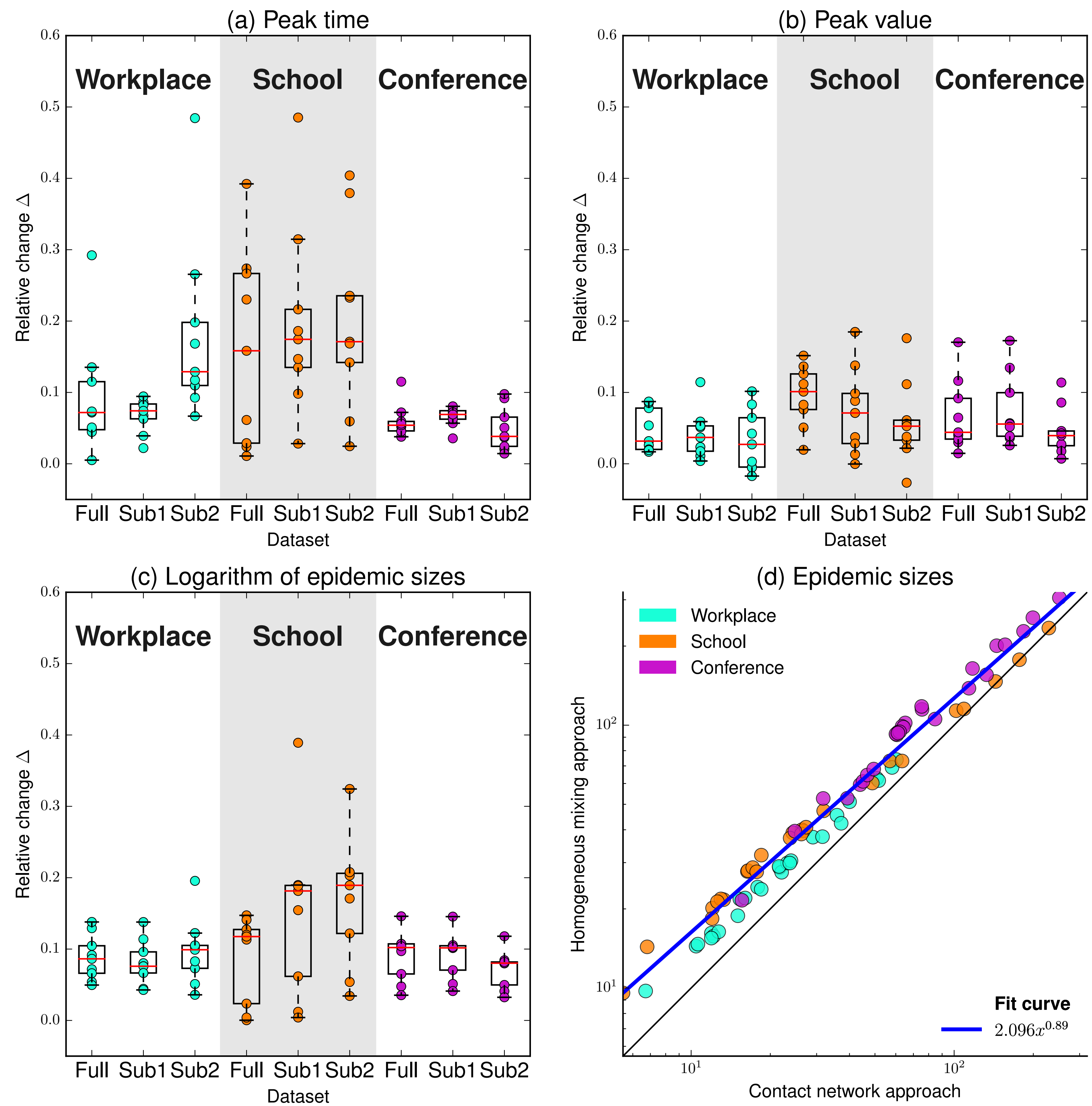
Evaluation



Test on synthetic + real datasets with varying SIR parameters

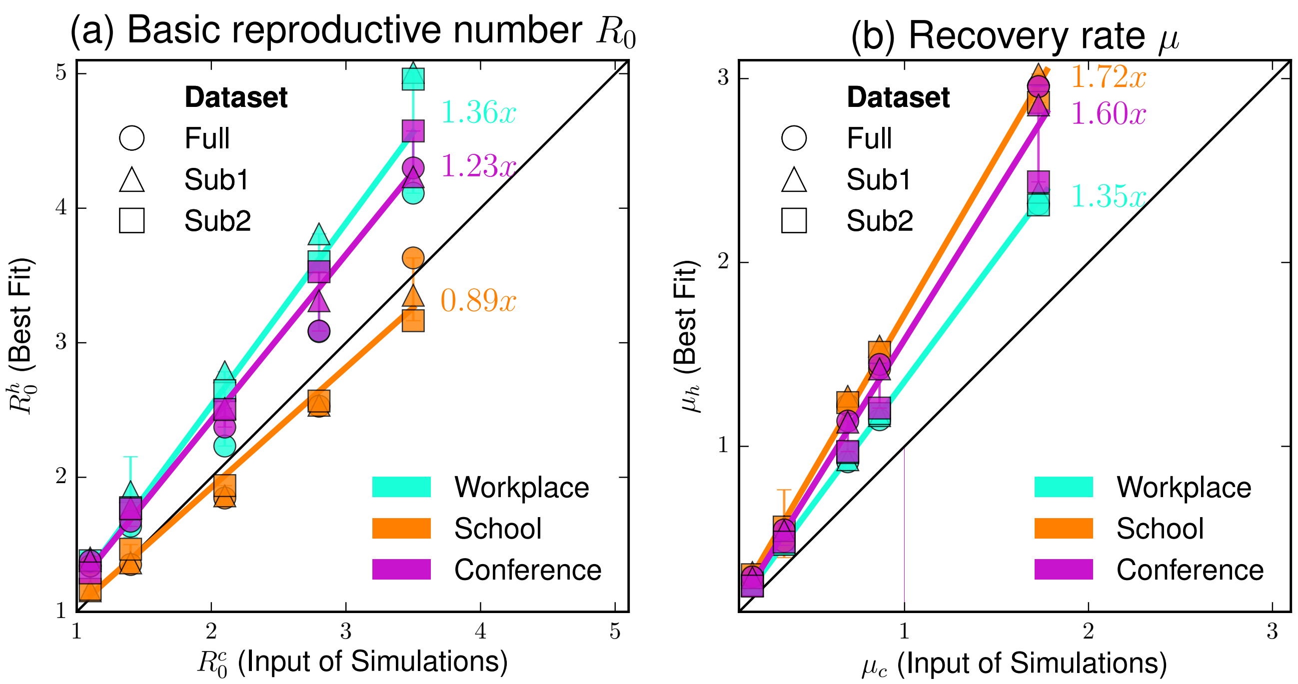
Bioglio et al., BMC Inf. dis. (2016)

Homogeneous mixing: (only) a (good) approximation

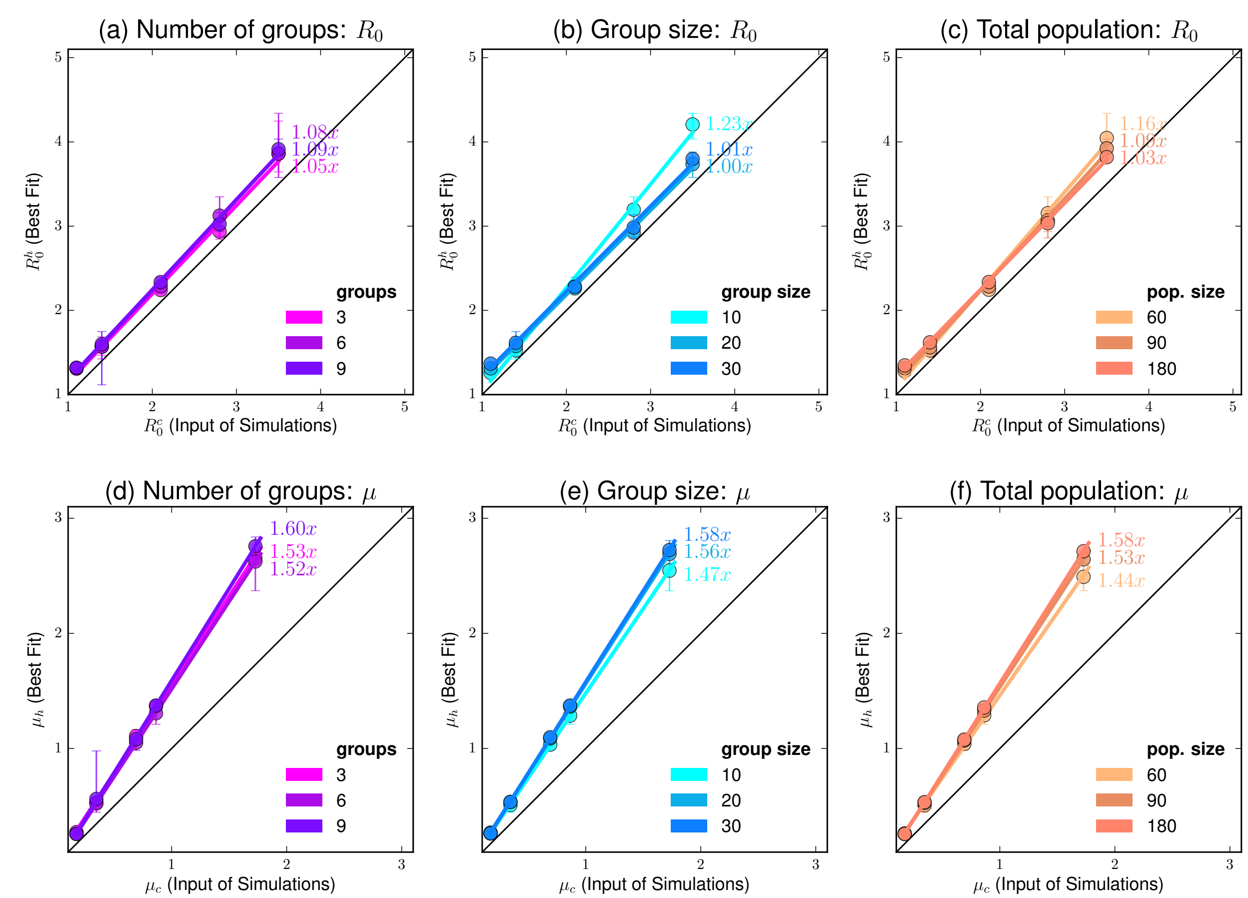


Bioglio et al., BMC Inf. dis. (2016)

Linear recalibration of parameters



Rescaling:
-need to recalibrate both parameters
-setting specific



Need more datasets!

Bioglio et al., BMC Inf. dis. (2016)

> Using data to design and evaluate interventions

Immunization strategies

=> take into account temporal structure

Lee et al., PLOS ONE (2012)
=>inspired by "acquaintance protocol" in static networks

- "*Recent*": choose a node at random, immunize its most recent contact
- "*Weight*": choose a node at random, immunize its most frequent contact in a previous time-window

Starnini et al., JTB (2012)

- aggregate network on [0,T]
- compare strategies
    - immunize nodes with highest k or BC in [0,T]
    - immunize random acquaintance (on [0,T])
    - recent, weight strategies
- vary T
- find saturation of efficiency as T increases

Liu et al., PRL (2014) (activity-driven network model=>analytics)

- target nodes with largest activity
- random neighbour (over an observation time T) of random node

> mesoscale interventions

- act at intermediate scales
- avoid need to identify nodes with specific properties

An example: SEIR + school
Which containment strategies?

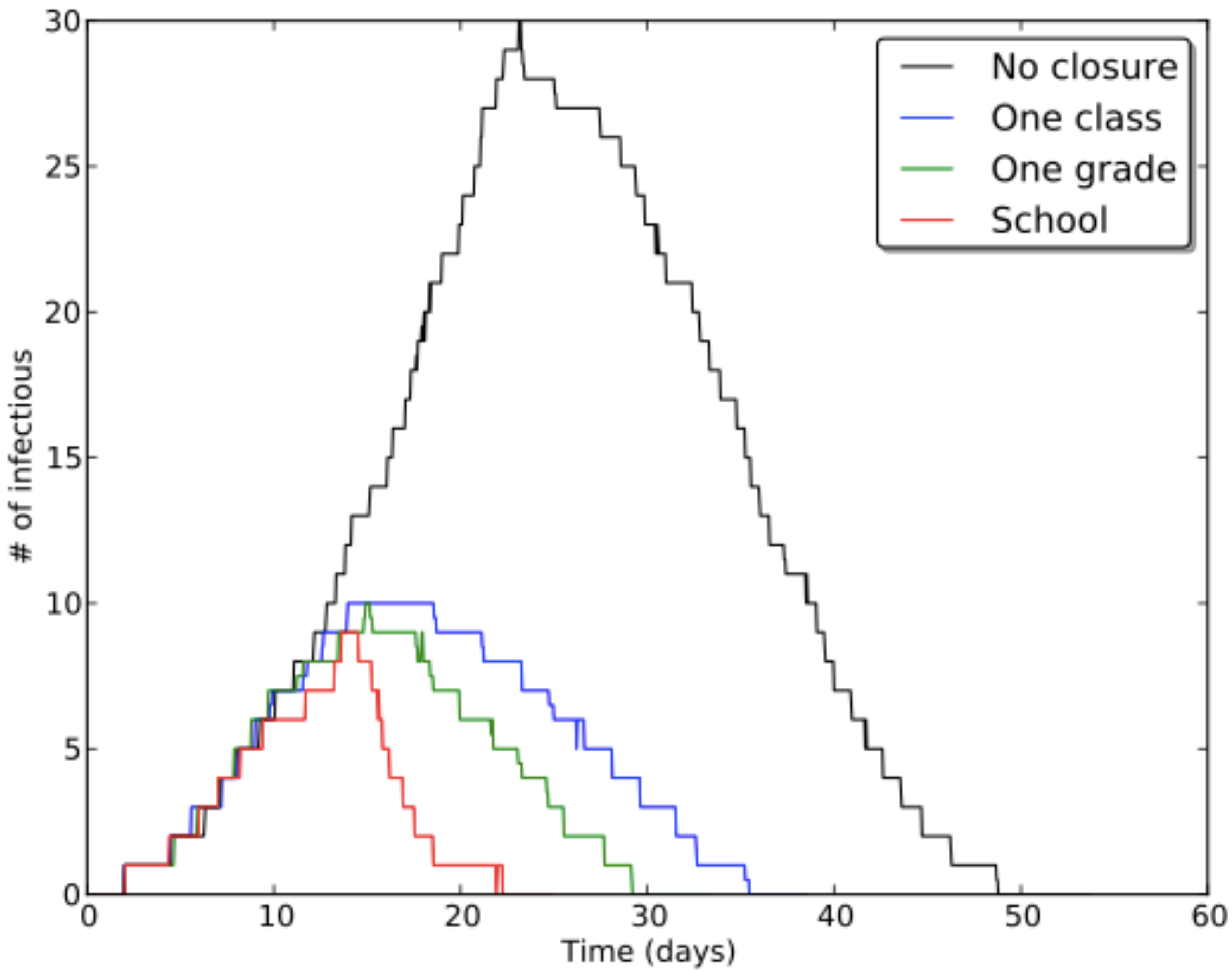
Model:
-SEIR with asymptomatics
-contact data as proxy for possibility of transmission inside school
-when children are out of school: residual homogeneous risk of contamination by contact with population

Containment strategies (suggested by the data):
-detection and subsequent isolation of symptomatic individuals
-whenever symptomatic individuals are detected (more than a given threshold), **closure of**

(i) class
(ii) class + most connected other class (same grade)
(iii) whole school

V. Gemmetto, A. Barrat, C. Cattuto, BMC Inf. Dis. 14, 3841 (2014)

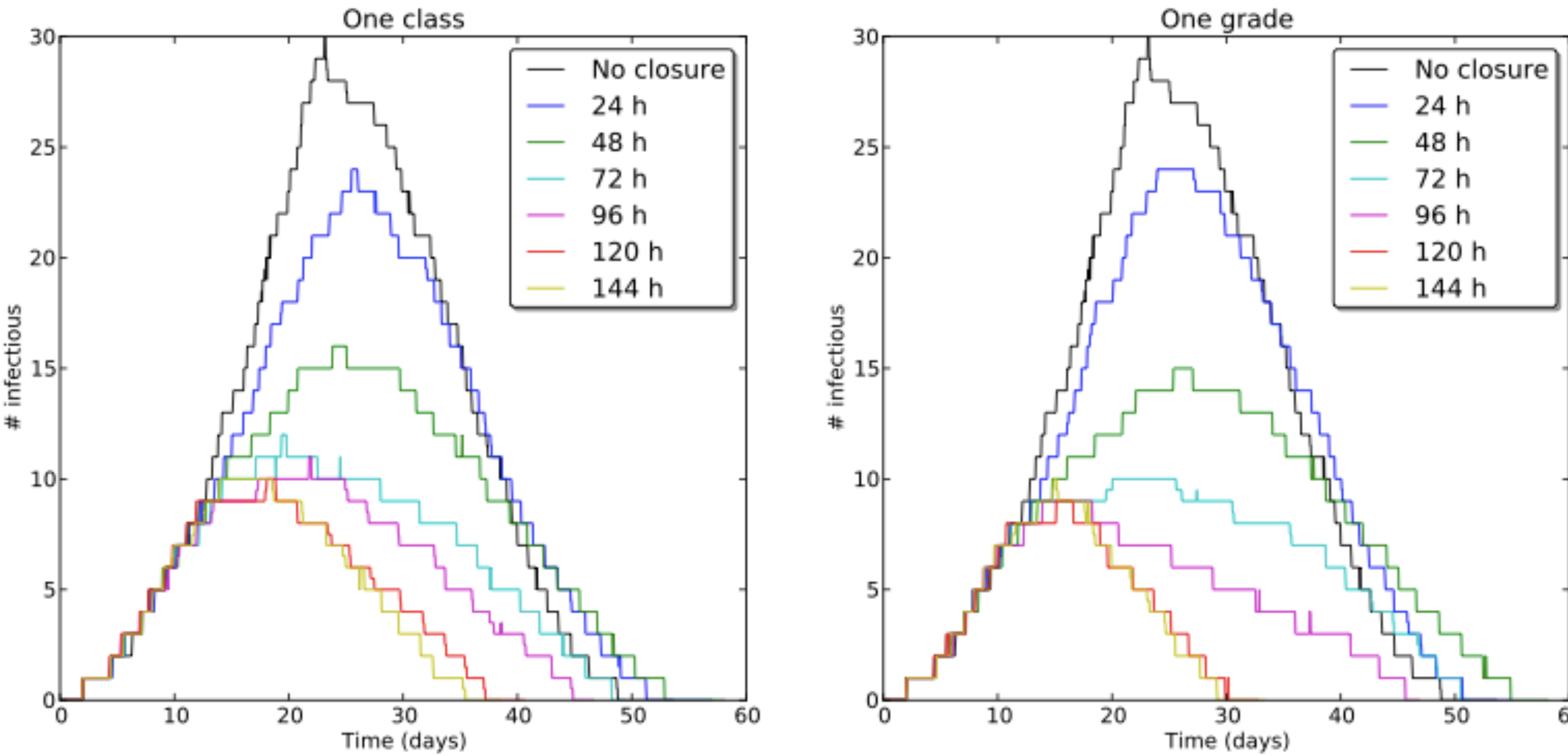
An example: SEIR + school
Which containment strategies?



Average over cases
with AR > 10%

V. Gemmetto, A. Barrat, C. Cattuto, BMC Inf. Dis. 14, 3841 (2014)

An example: SEIR + school
Which containment strategies?



Average over cases
with AR > 10%

V. Gemmetto, A. Barrat, C. Cattuto, BMC Inf. Dis. 14, 3841 (2014)

Which containment strategies?
Comparing class/grade/school closure

Strategy (Threshold, duration)	Targeted class	Targeted grade	School
No closure			34.6
3, 24 h			26.0
3, 48 h			23.2
3, 72 h			14.8
3, 96 h			13.0
3, 120 h			7.5
3, 144 h			5.6
2, 24 h			22.9
2, 48 h			17.8
2, 72 h			14.4
2, 96 h			11.0
2, 120 h			3.2
2, 144 h			1.6

Probability that AR < 10%

Strategy (Threshold, duration)	Targeted class	Targeted grade	School
No closure			179 [149,203]
3, 24 h			170 [151,202]
3, 48 h			162 [43,199]
3, 72 h			146 [28,198]
3, 96 h			120 [27,195]
3, 120 h			67 [26,192]
3, 144 h			55 [25,180]
2, 24 h			173 [139,198]
2, 48 h			170 [62,199]
2, 72 h			149 [48,201]
2, 96 h			141 [31,196]
2, 120 h			133 [30,195]
2, 144 h			57 [25,192]

Average AR when AR > 10%

V. Gemmetto, A. Barrat, C. Cattuto, BMC Inf. Dis. 14, 3841 (2014)

An example: SEIR + school
Which containment strategies?

Closure strategy (Threshold, duration)	Targeted class	Targeted grade	Whole school
3, 24h	6.2	6.6	10.0
3, 48h	7.6	8.0	14.3
3, 72h	8.2	9.7	16.1
3, 96h	11.3	13.7	22.4
3, 120h	12.2	13.5	26.5
3, 144h	13.3	13.9	27.9
2, 24h	5.8	5.8	10.0
2, 48h	6.5	7.6	13.5
2, 72h	6.4	8.1	16.1
2, 96h	8.5	9.4	21.5
2, 120h	8.5	10.6	24.3
2, 144h	8.3	9.8	25.3

Cost in number of lost class-days

V. Gemmetto, A. Barrat, C. Cattuto, BMC Inf. Dis. 14, 3841 (2014)

Two-sided role of the data

- "Simple" stylized facts, easy to generalise
  - – more contacts within each class than between classes
  - – which classes are most connected

=> from low-resolution data

=> suggests strategies

- Complex, detailed data set
  - – detailed sequence of contact events

=> high-resolution data

=> validation of strategies through detailed and realistic numerical simulations

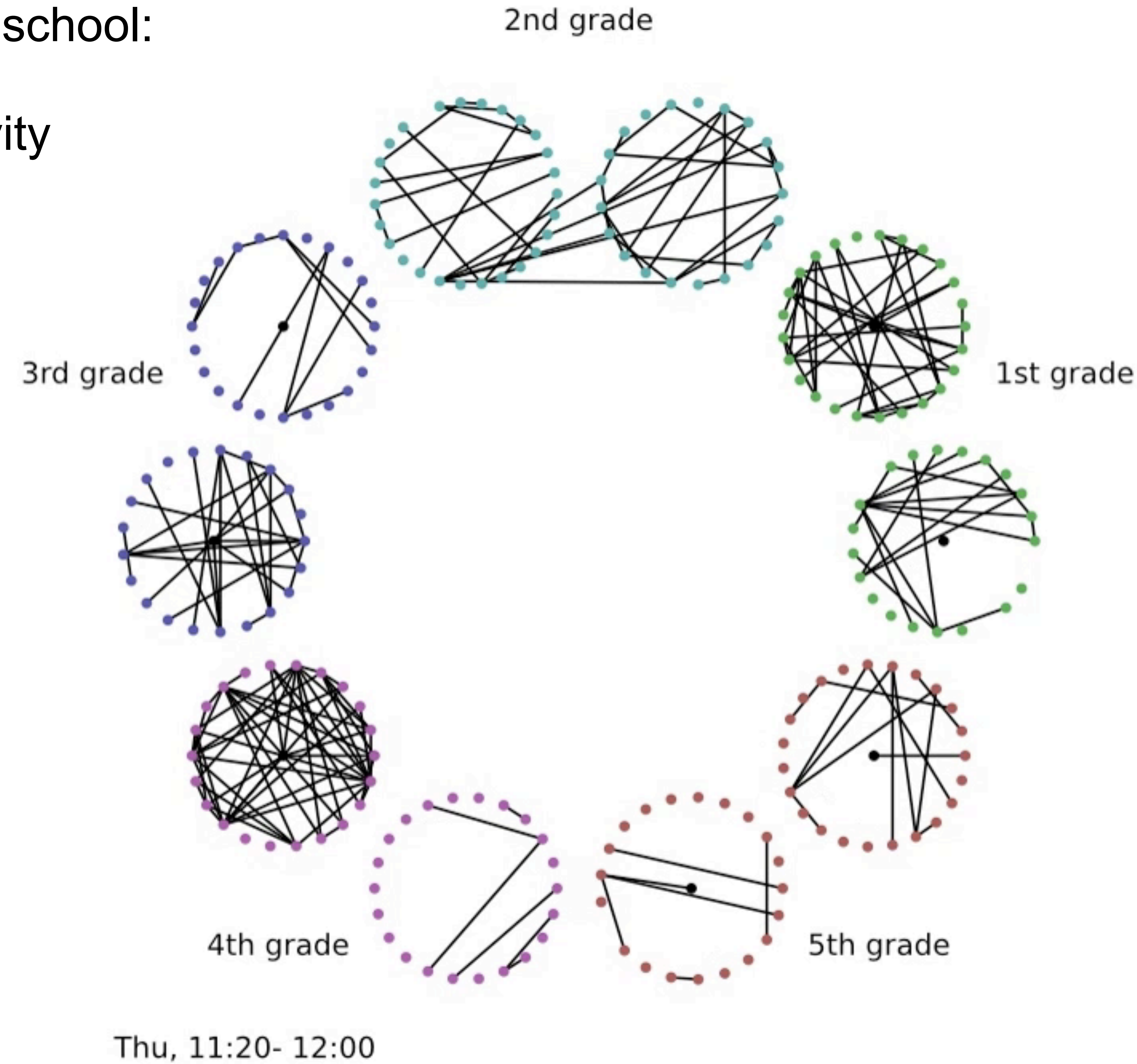
NB: Fumanelli et al. (2016) & Ciavarella et al. (2016) show the validity of these results using agent-based models at larger scale.

> mesoscale interventions

- 2

contacts in a primary school:

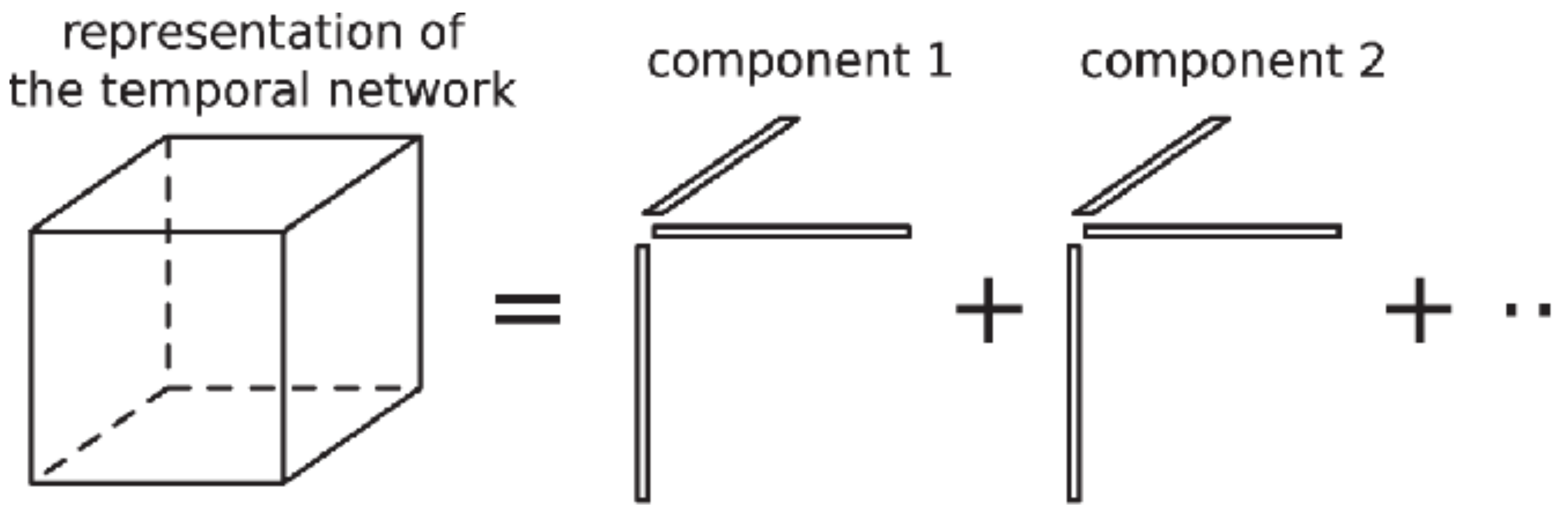
- classes
- correlated link activity



J. Stehlé et al. PLoS ONE 6(8):e23176 (2011)

Decomposition of temporal network

Kruskal decomposition



$$T \approx \tilde{\mathcal{T}} = \sum_{r=1}^{R} \mathcal{S}_r = \sum_{r=1}^{R} \mathbf{a_r} \circ \mathbf{b_r} \circ \mathbf{c_r}$$

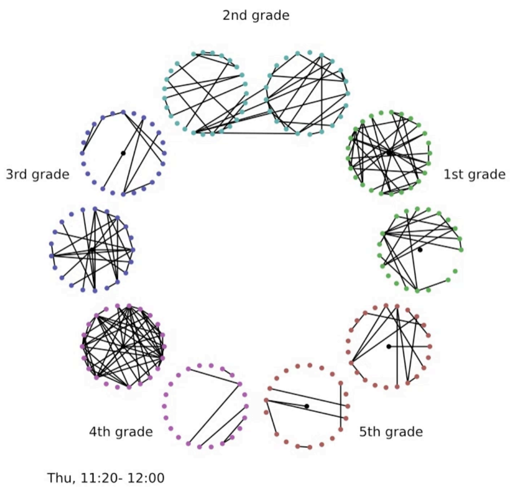
$$t_{ijk} \approx \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}$$

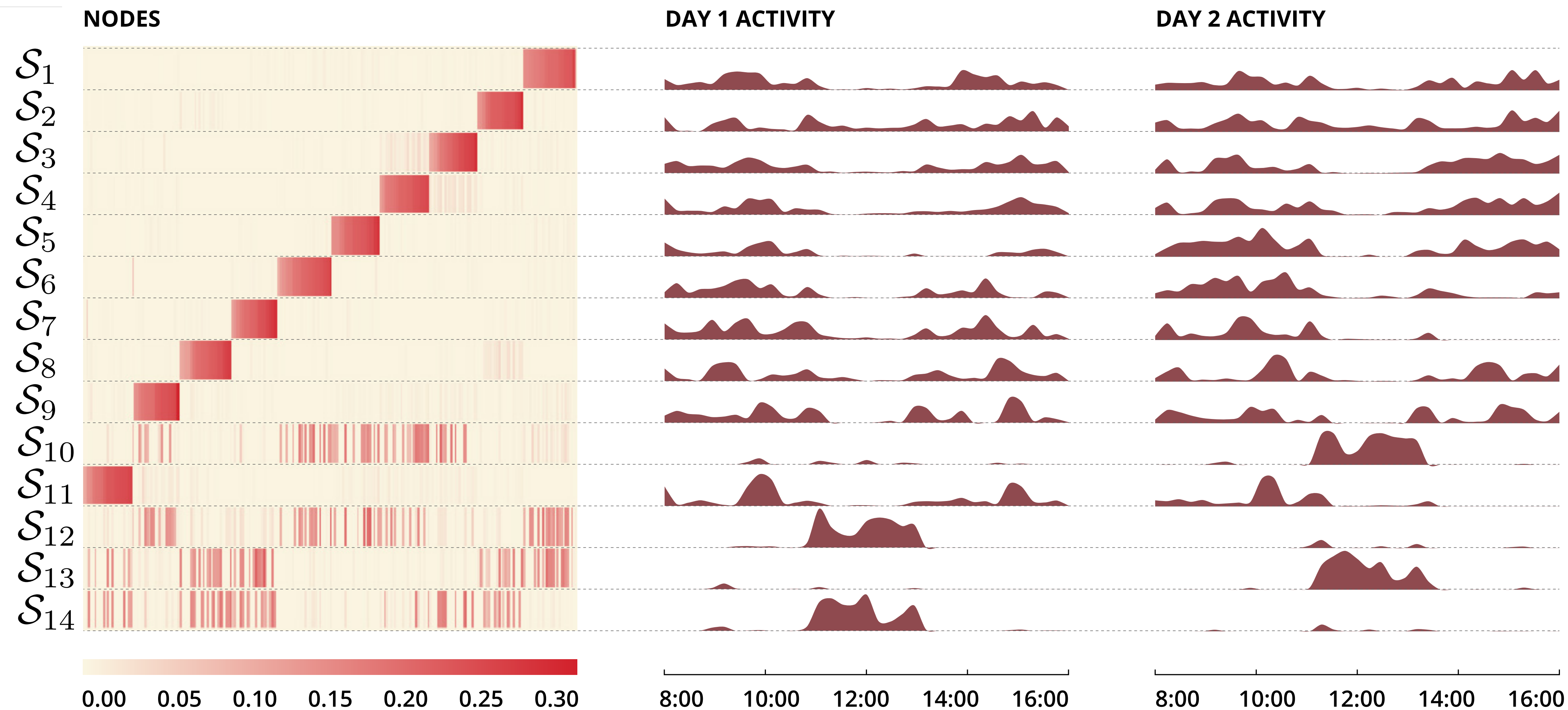
undirected network: a=b

a_{ir}= membership of node i to component r
c_{kr}=timeline of component r

L. Gauvin et al., PLoS ONE (2014)

Primary school case study





10 classes

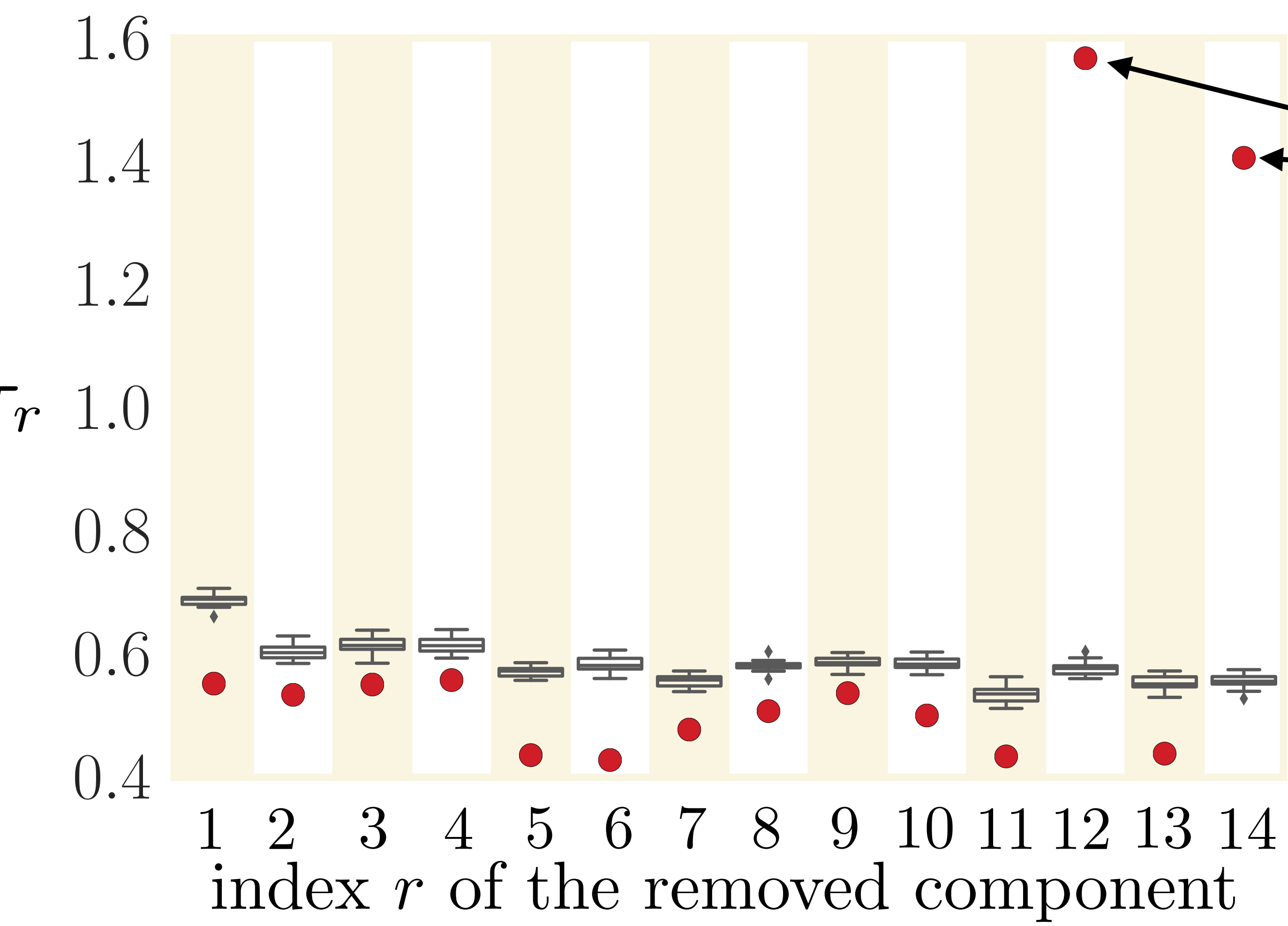
4 mixed-membership components = breaks

L. Gauvin et al., PLoS ONE (2014)

Intervention: removal of a component

Altered temporal network $$\tilde{\mathcal{T}}^s = \sum_{r \neq s} \mathcal{S}_r$$

Stochastic SI model, epidemic delay ratio $$\tau_r = \left\langle \frac{t_j^r - t_j}{t_j} \right\rangle$$



Components 12 and 14=1st day lunch break

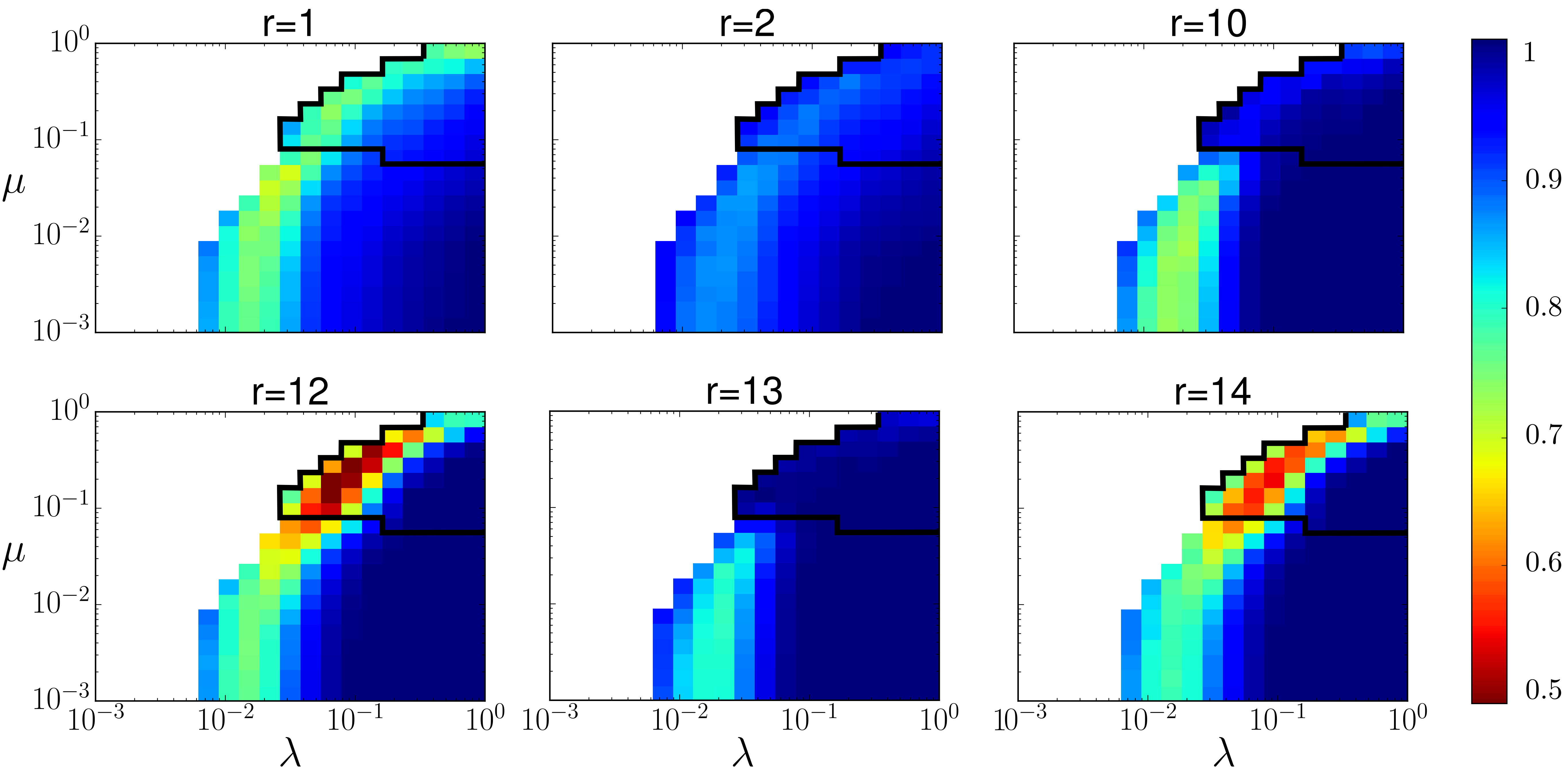
Comparison with null model=removal of randomly chosen links

NB: components 12,14 carry less weight than 1 or 2

L. Gauvin et al., arXiv:1501.02758
Revealing latent factors of temporal networks for mesoscale intervention in epidemic spread,

Intervention: removal of a component

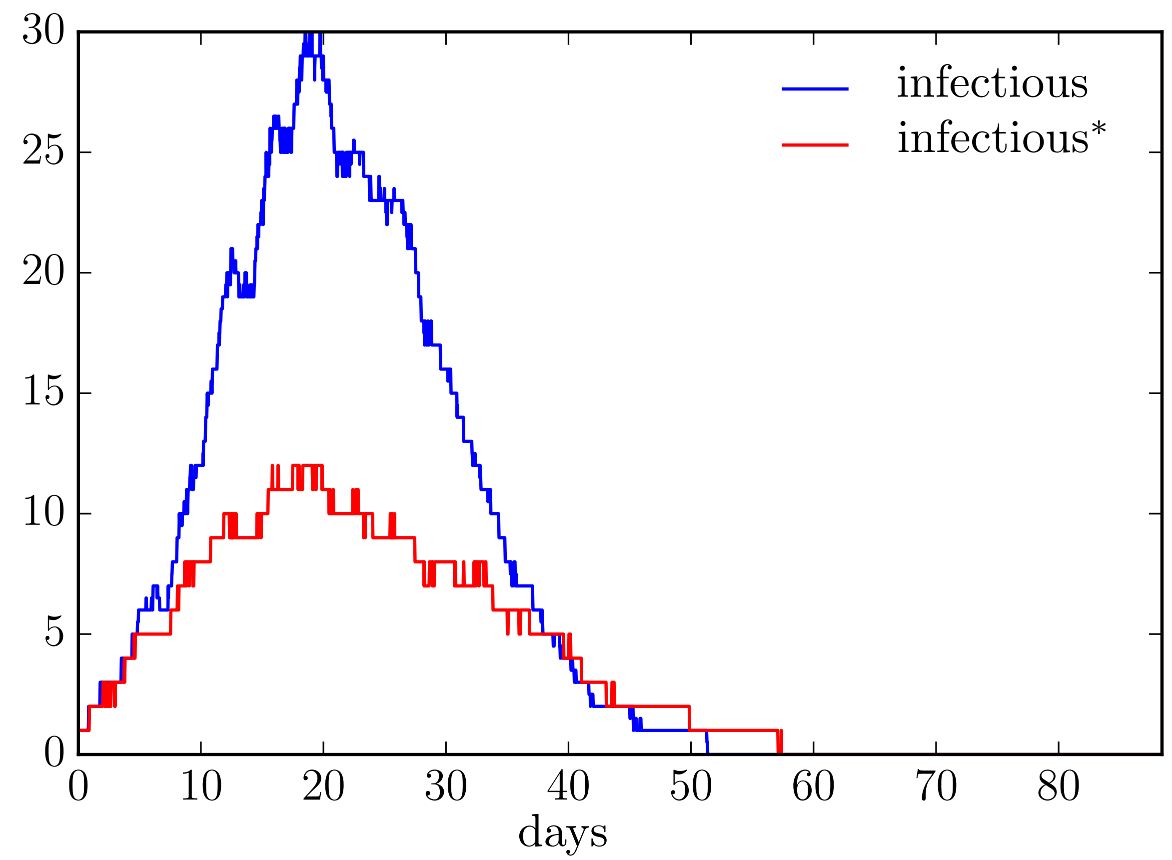
SIR model, epidemic size ratio



L. Gauvin et al., arXiv:1501.02758
Revealing latent factors of temporal networks for mesoscale intervention in epidemic spread,

Reactive targeted intervention

- SEIR
- contact data as proxy for possibility of transmission inside school
- when children are out of school: residual homogeneous risk of contamination by contact with population
- isolation of infectious cases at the end of each day
- if 2 infectious are detected:
    - **removal** of mixing components
    - **replacement** by equivalent amount of class activities



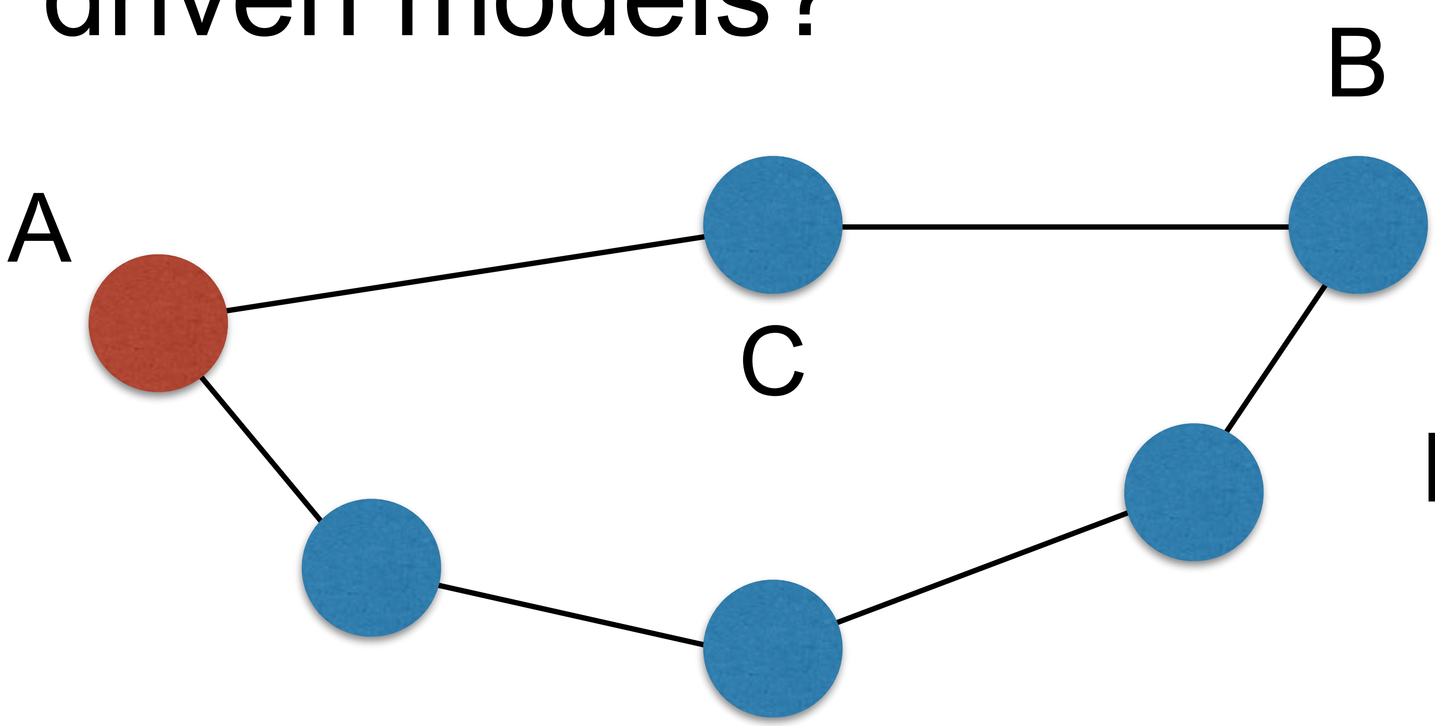
L. Gauvin et al.,
Revealing latent factors of temporal networks for mesoscale intervention in epidemic spread,
arXiv:1501.02758

> Incomplete data

Data is often incomplete because of population sampling
(not everyone takes part to the data collection)

Consequences of missing data….

- Biases in measured statistical properties?
- Biases in the estimation of the outcomes of data-driven models?



$d_{AB,\ real} = 2$

$d_{AB,\ sampled} = 4$

Risk of contamination A->B
real >> sampled
(C acts as an immunized node)

Issues

I- Evaluation of biases due to missing data

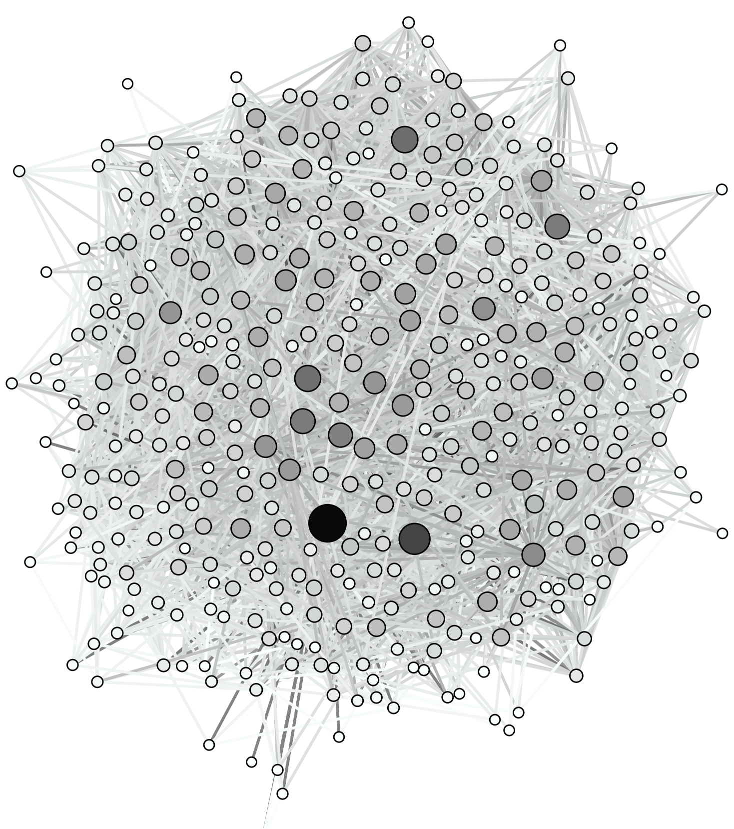
II- Compensating for biases

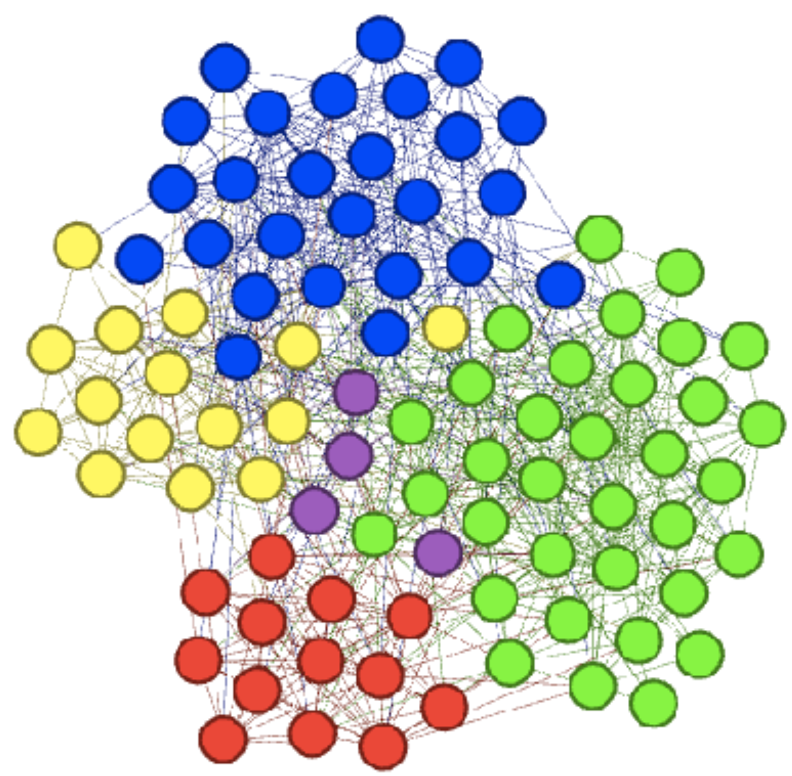
Methodology: resampling

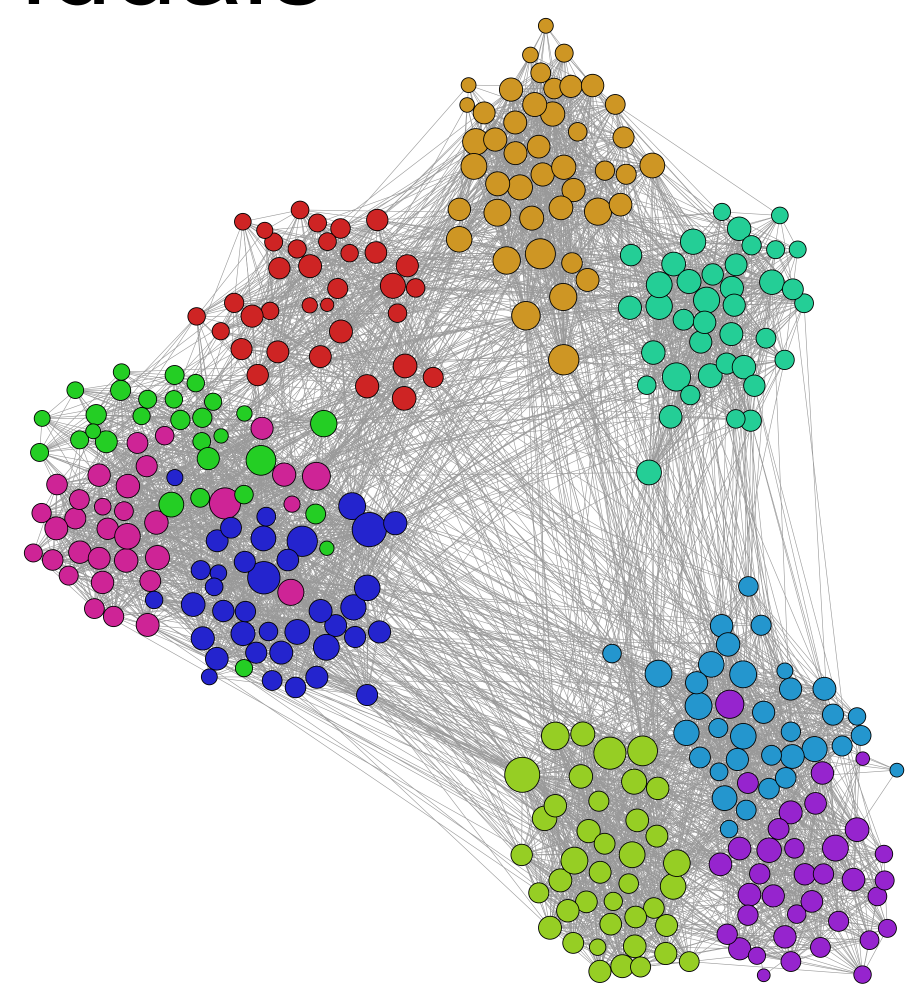
- Remove nodes (individuals) at random
- Measure statistical properties of contacts between remaining individuals
- Run simulations of spreading processes (SIS or SIR models)
- Compare outcomes of simulations on initial and resampled temporal networks

Data sets: temporal networks of face-to-face proximity

- Conference: 2 days, 403 individuals
- Offices: 2 weeks, 92 individuals
- High School: 1 week, 326 individuals

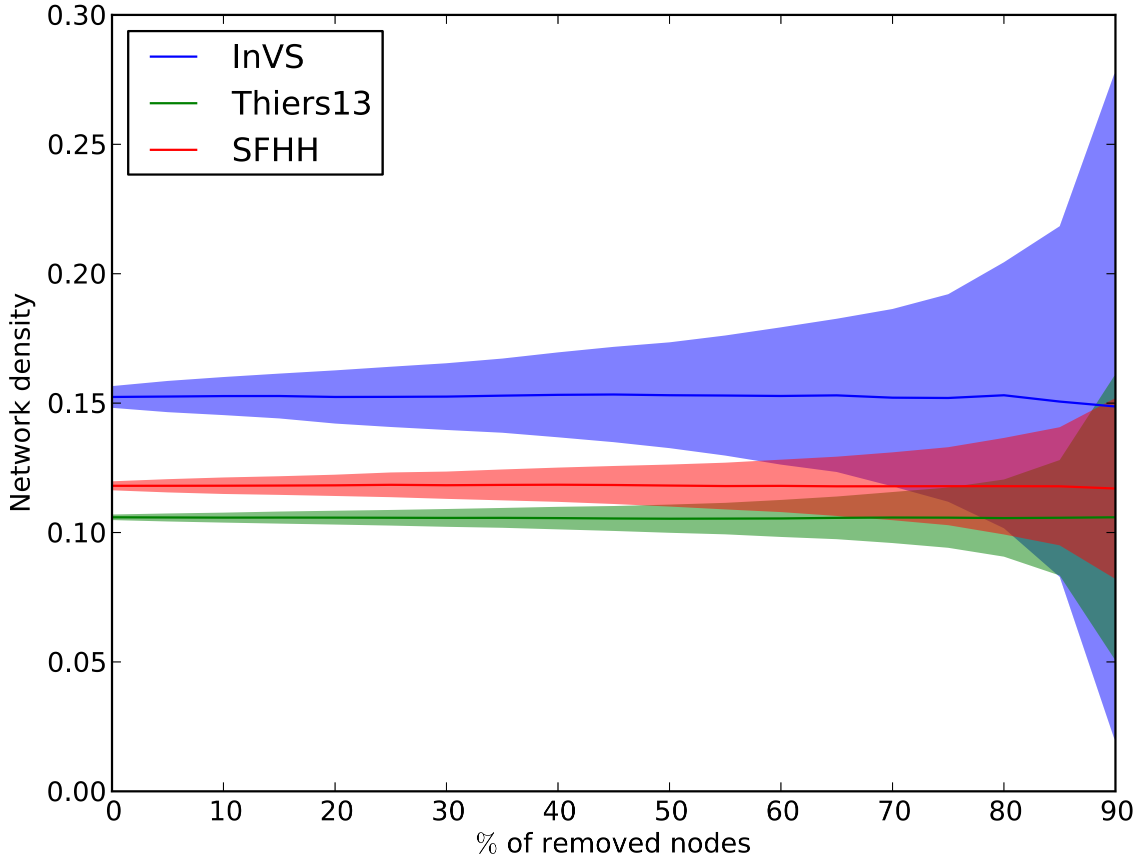


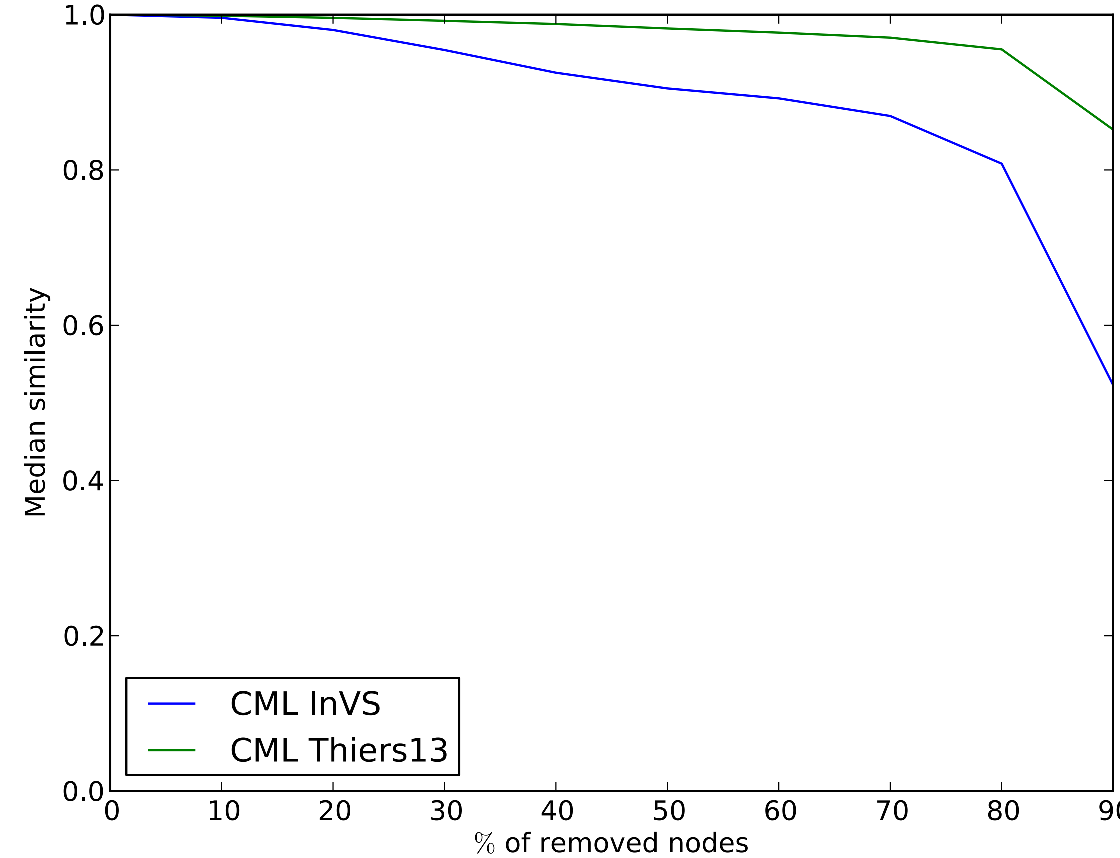




SocioPatterns.org

Impact of sampling on contact network properties and mixing patterns





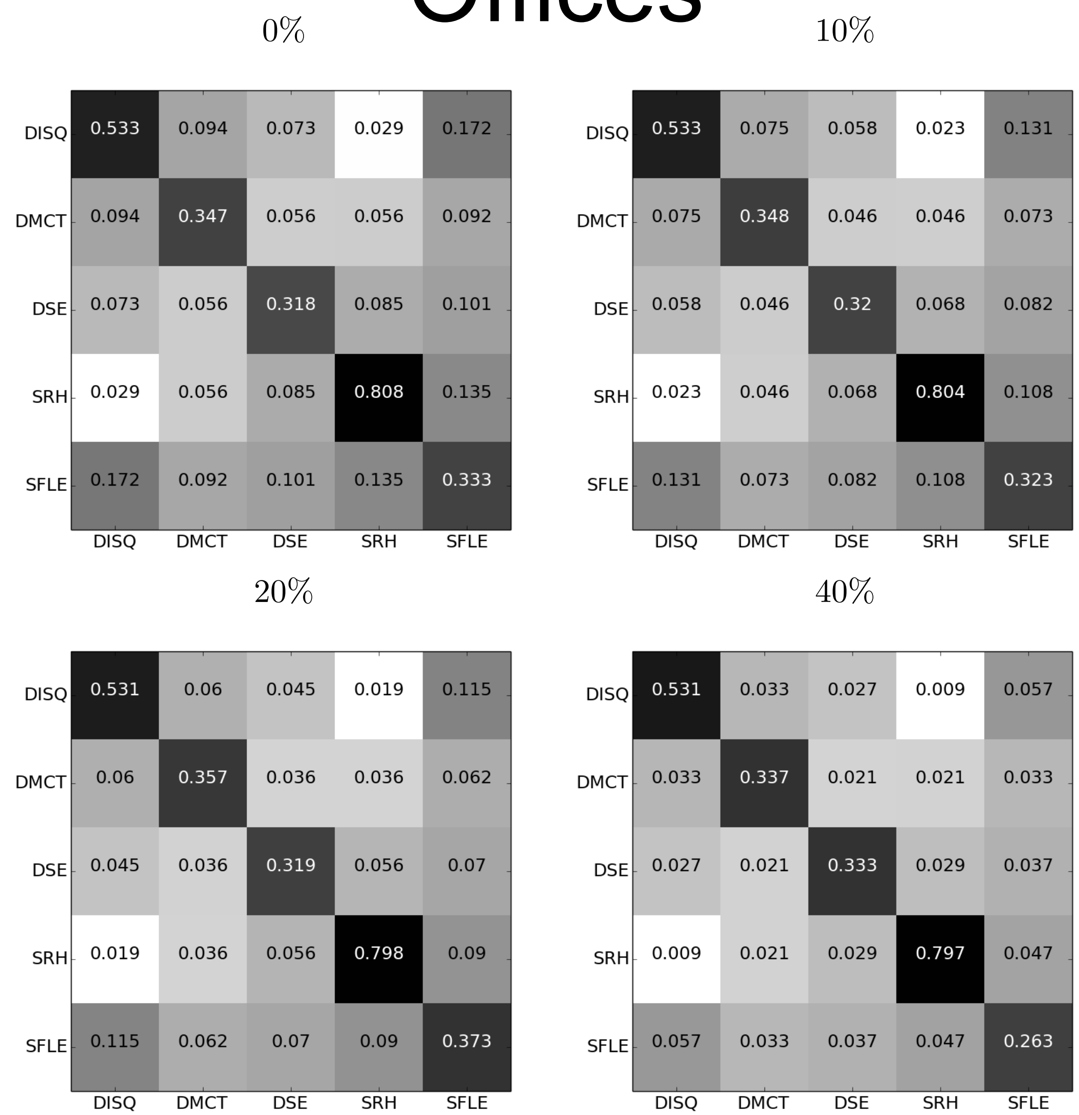
Random sampling keeps:

- **Temporal statistical properties** (distributions of contact and inter-contact durations, distribution of aggregate durations, of # contacts per link)
- **Density** of network
- **Contact matrices of link densities** (for structured populations)

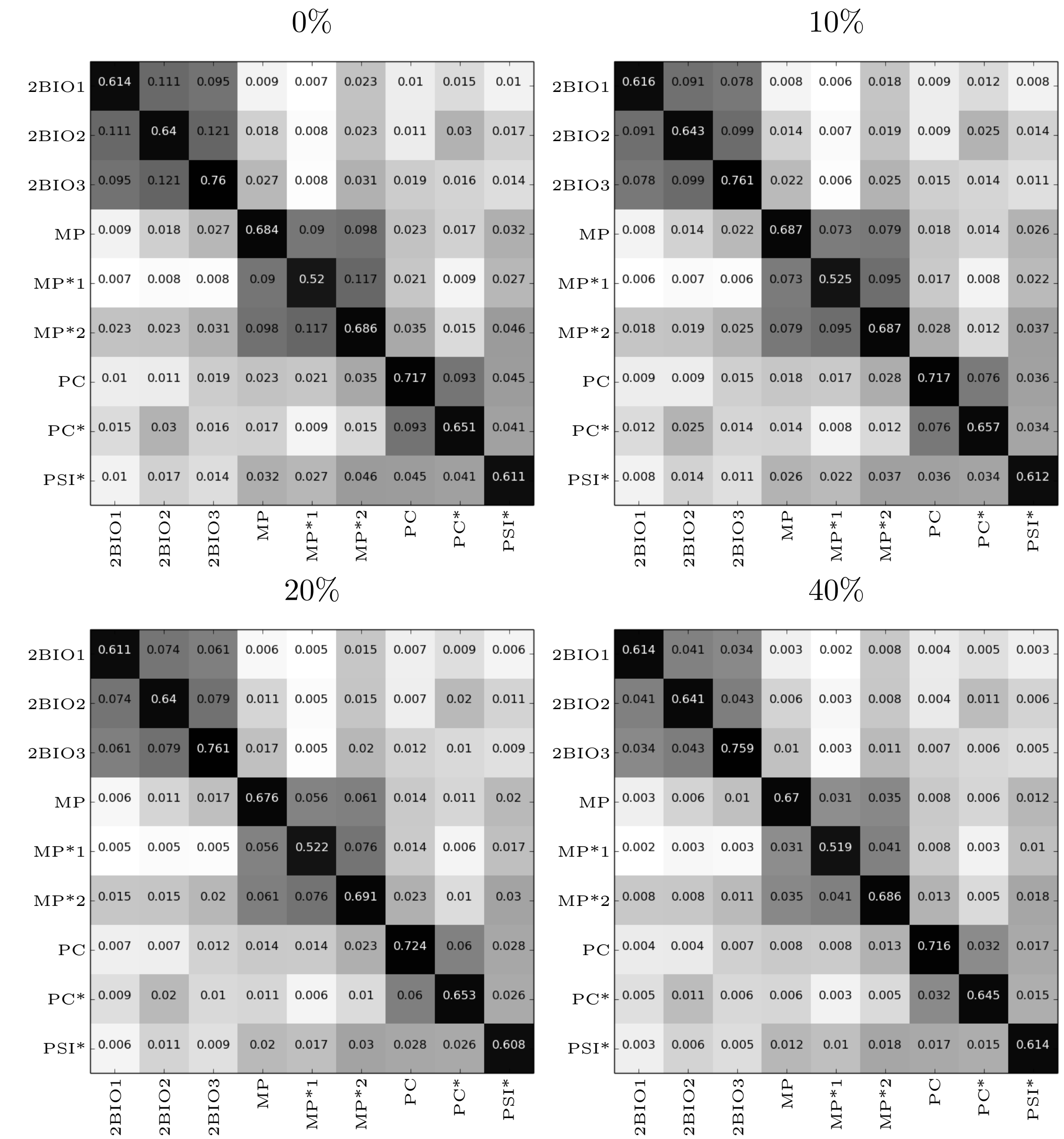
known to be crucial for spreading processes

Impact of sampling on contact network properties and mixing patterns

Offices



High School



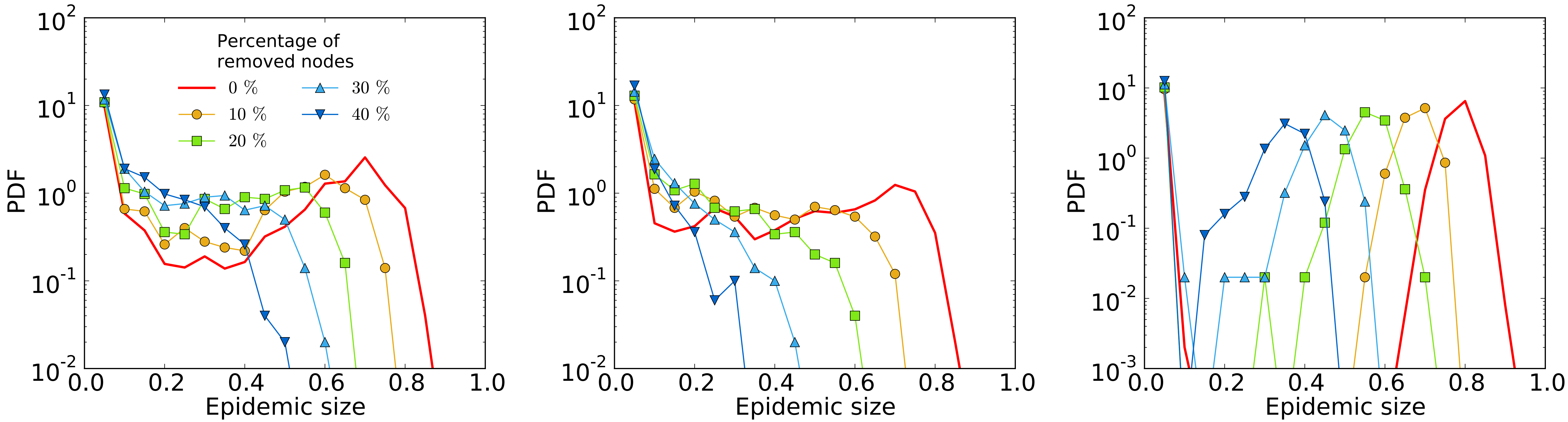
Random sampling keeps:

- **Temporal statistical properties** (distributions of contact and inter-contact durations, distribution of aggregate durations, of # contacts per link)
- **Density of network**
- **Contact matrices of link densities** (for structured populations)

known to be crucial for spreading processes

Impact of sampling on simulations of spreading processes

SIR model, distribution of epidemic sizes



Offices High School Conference

Strong **underestimation** of

- risk of large epidemics
- average epidemic size

(absent individuals act as if they were immunised => herd immunisation effect)

How to estimate the outcome of spreading processes from incomplete data?

Do the incomplete data contain enough information?

Methodology

- build ***surrogate*** contact networks, using ***only*** properties of sampled (incomplete) data:
  - Measure ***properties of the (sampled) incomplete data***
  - Add missing nodes (**assumption: class/department known**)
  - Build links between missing and present nodes, ***in order to maintain the properties measured in the incomplete data***
  - On each link, ***create a timeline of contacts respecting the statistical properties of the incomplete data***

- perform simulations on surrogate data sets

NB: surrogate contacts, here we do not try to infer (unknown) real contacts

Which properties / how much detail?

- Method “**0**”: measure **density + average weight**, add back nodes and links in order to keep ***density*** fixed (=>increases average degree, ***simplest compensation method***); ***Poissonian timelines of contact events on links;***

- Method “**W**”: measure **density** + **statistics** of link’s **weights**, add back nodes and links to keep density fixed, ***weights on surrogate links taken from measured ones, Poissonian timelines on links;***

- Method “**WS**”: measure **link density contact matrix** + statistics of link’s **weights**, **separating intra- and inter-groups links**; add back nodes and links to keep ***contact matrix fixed, weights on surrogate links taken from measured ones, separating intra-group and inter-group links, Poissonian timelines on links;***

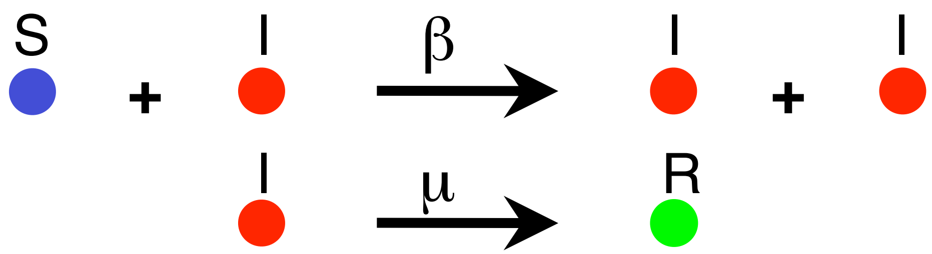
- Method “**WT**”: measure **density** + statistics of **numbers of contacts per link** and of **contact and inter-contact durations**; add back nodes and links as in method “**0**”; for each surrogate link, create a ***surrogate timeline by extracting a number of contacts at random and alternate contact and inter-contact durations;***

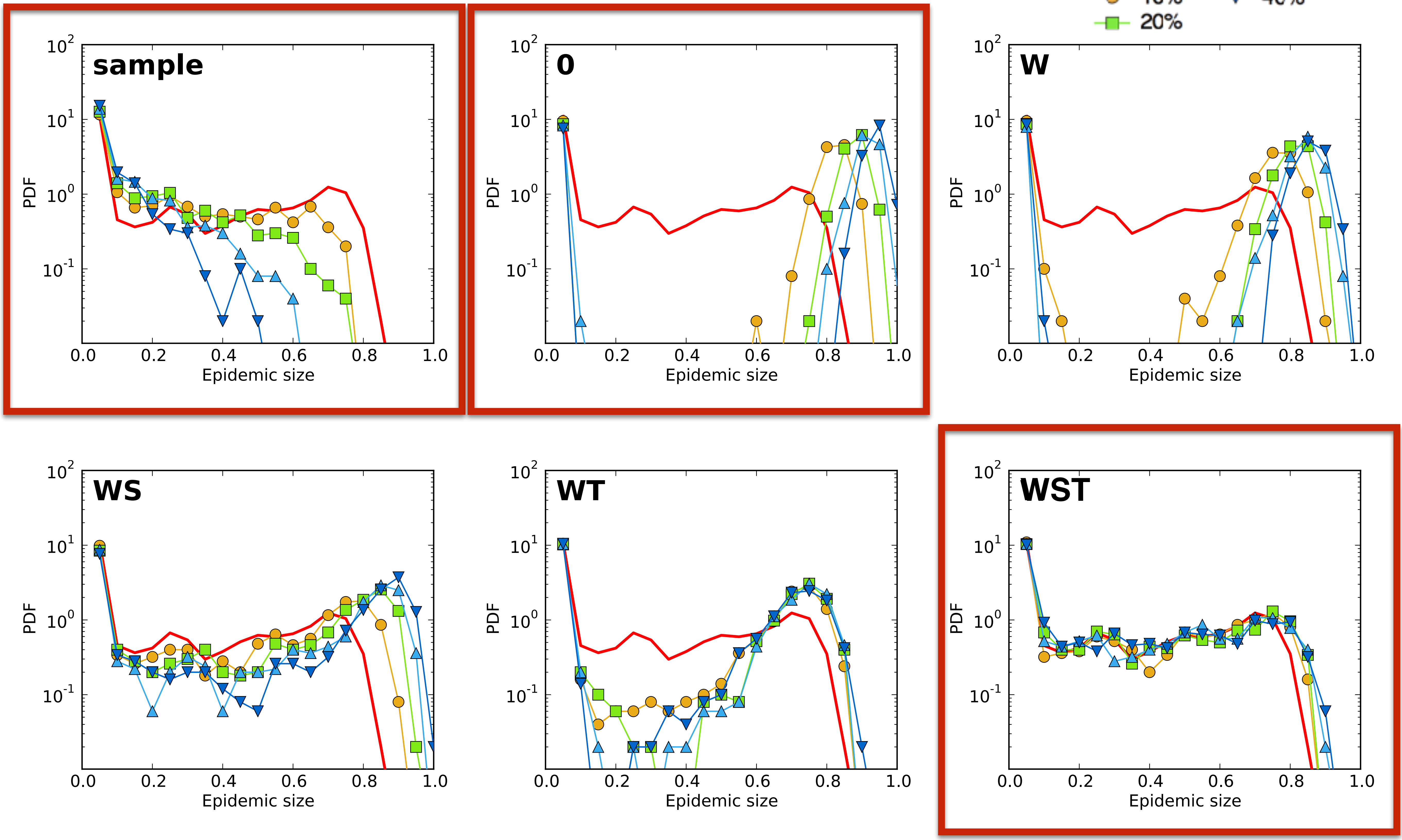
- Method “**WST**”: measure **link density contact matrix** + statistics of **numbers of contacts per link** and of **contact and inter-contact durations**, **separating intra- and inter-groups links**; add back nodes and ***surrogate links as in “WS”***, create ***surrogate timelines as in “WT”.***

Increasing information

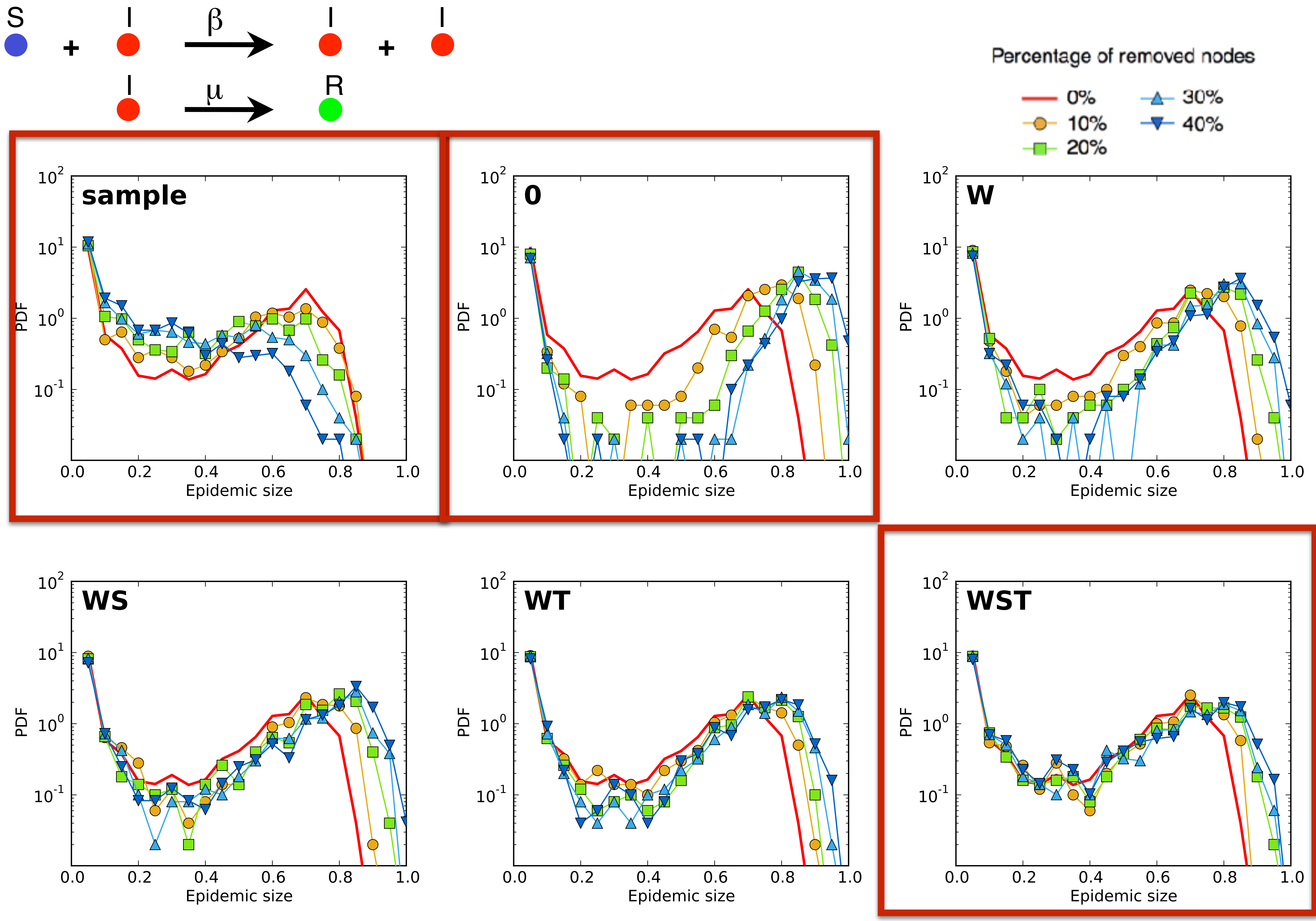
Results

SIR model, distribution of epidemic sizes (high school)

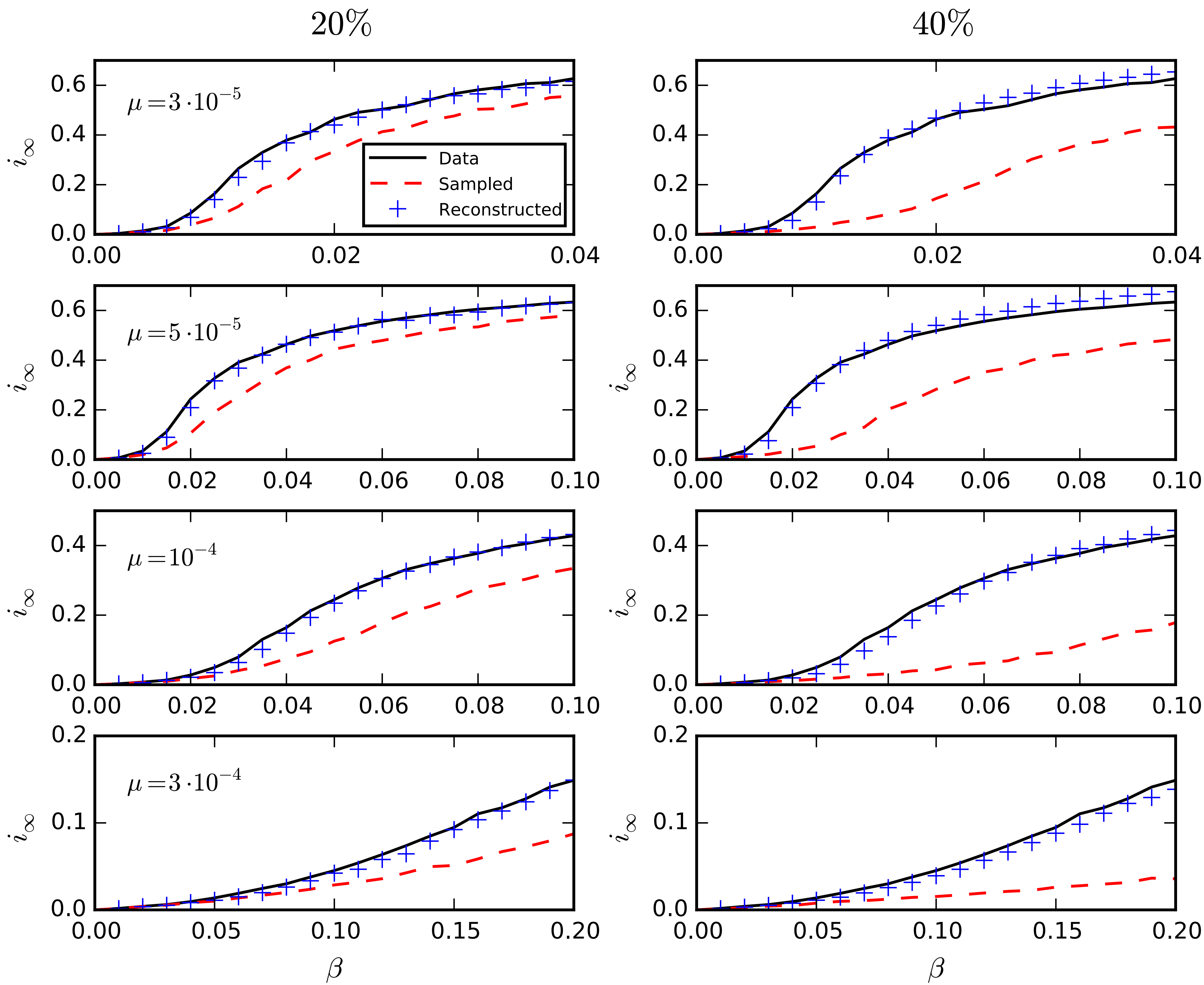




SIR model, distribution of epidemic sizes (offices)



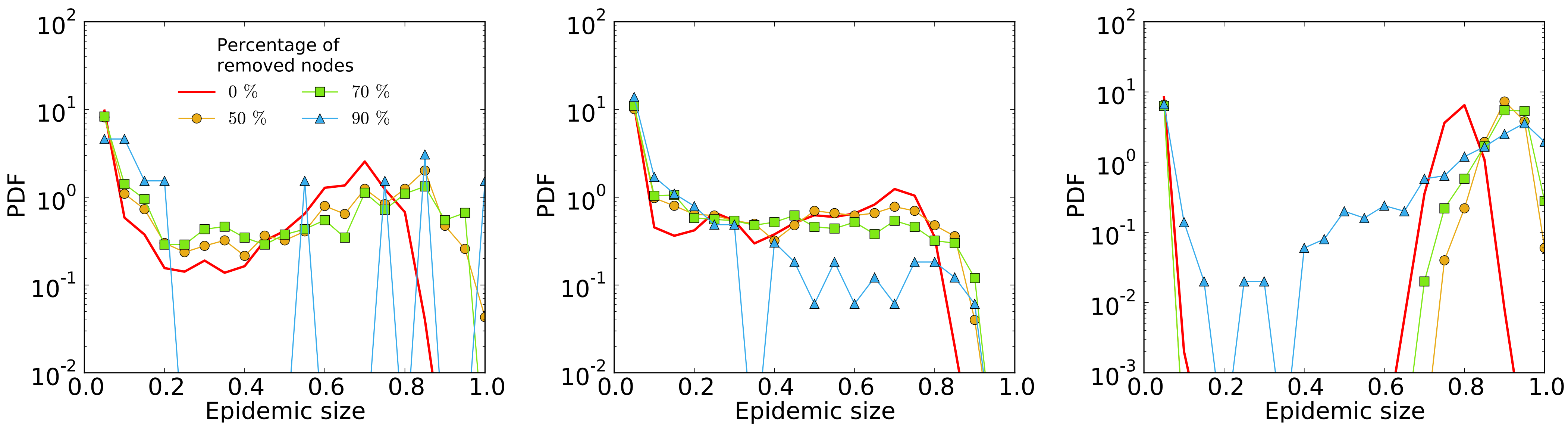
SIS case


20%
40%
$\mu = 3 \cdot 10^{-5}$
$\mu = 5 \cdot 10^{-5}$
$\mu = 10^{-4}$
$\mu = 3 \cdot 10^{-4}$
Data
Sampled
Reconstructed
i_∞
β

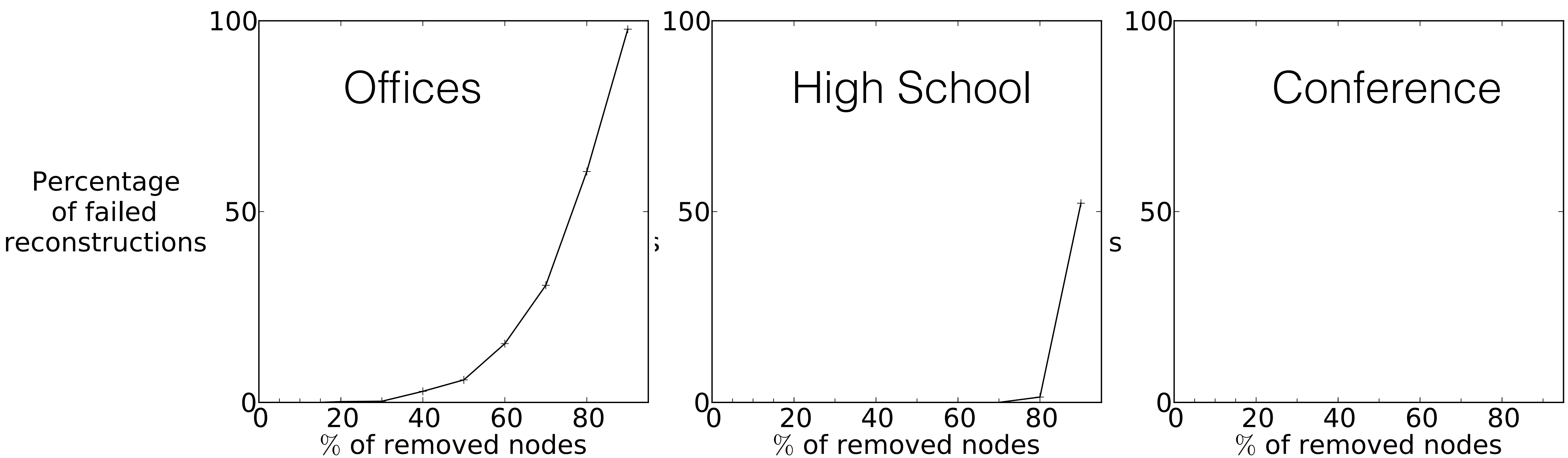
Limitations

Very low sampling rate => worse estimation of contact patterns

I- Larger deviations



II- Reconstruction procedure fails (not enough statistics in sampled data)



Limitations

Systematic overestimation of

- maximal size of large epidemics
- probability and average size of large outbreaks

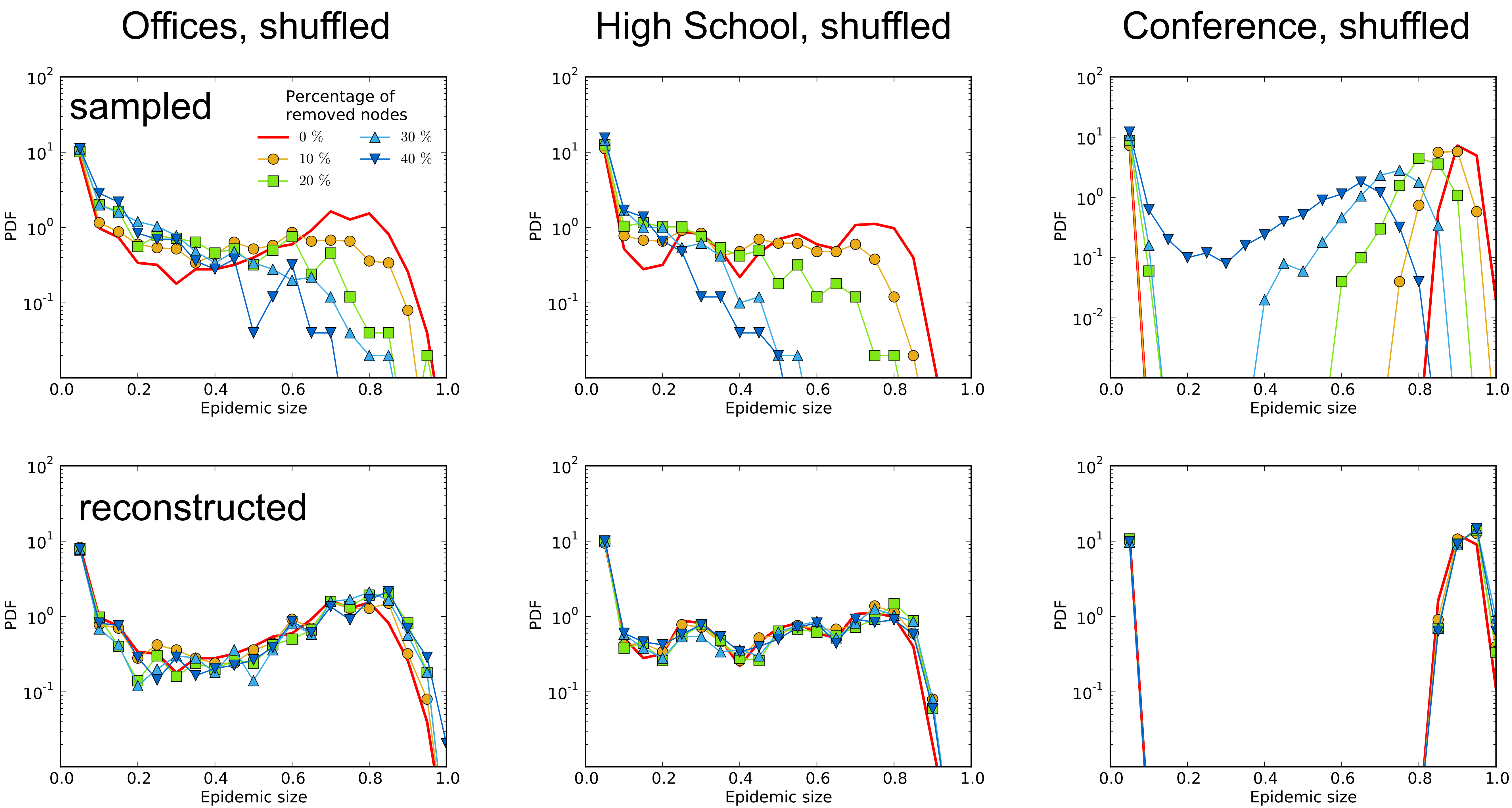
Due to correlations (structure and/or temporal) ?

- present in data
- not reproduced in surrogate data

To test this hypothesis:

- use reshuffled data (structurally or temporally) in which correlations are destroyed,
- perform resampling and construction of surrogate data,
- simulate spreading.

Using structurally reshuffled data



=> overestimation due to the presence of small-scale structures in real data, difficult or even impossible to reproduce in surrogate data

Using incomplete contact data

- Population sampling
  - can strongly underestimate the outcome of spreading processes
  - however, maintains important properties of the temporal contact network (distribution of contact/inter-contact durations, of aggregate durations, of # of contacts per link), and of the aggregated network (overall network density, structure of contact matrix)

- Possible estimation of the real outcome of simulations of spreading processes using surrogate contacts for the missing individuals
  - measure properties of sampled data
  - build surrogate links and contacts that respect the measured statistics
  - leads to a small overestimation of the outcome due to non captured correlations

M. Génois, C. Vestergaard, C. Cattuto, A. Barrat
arXiv:1503.04066, Nat. Comm. 6:8860 (2015)

What if no metadata (group structure) is available?

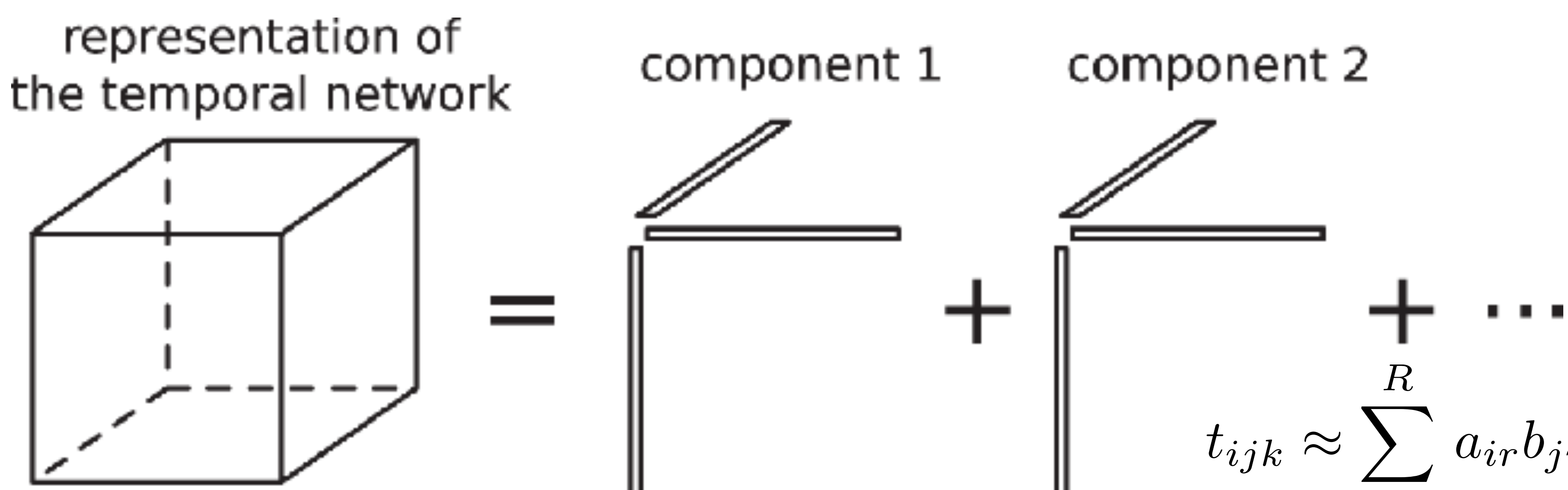
Hypothesis: data missing for certain nodes at certain times
Known: which nodes have missing data

Using NTF

- Transform the incomplete data into a 3-way tensor
- Perform the Kruskal decomposition of the incomplete data
- Deduce structures, their timelines, and to which structure each pair (i,j), where either i or j has missing data, belongs

- From the timeline of the structure to which (i,j) belongs, deduce when (i,j) should be active and add back the corresponding contact event => surrogate temporal network

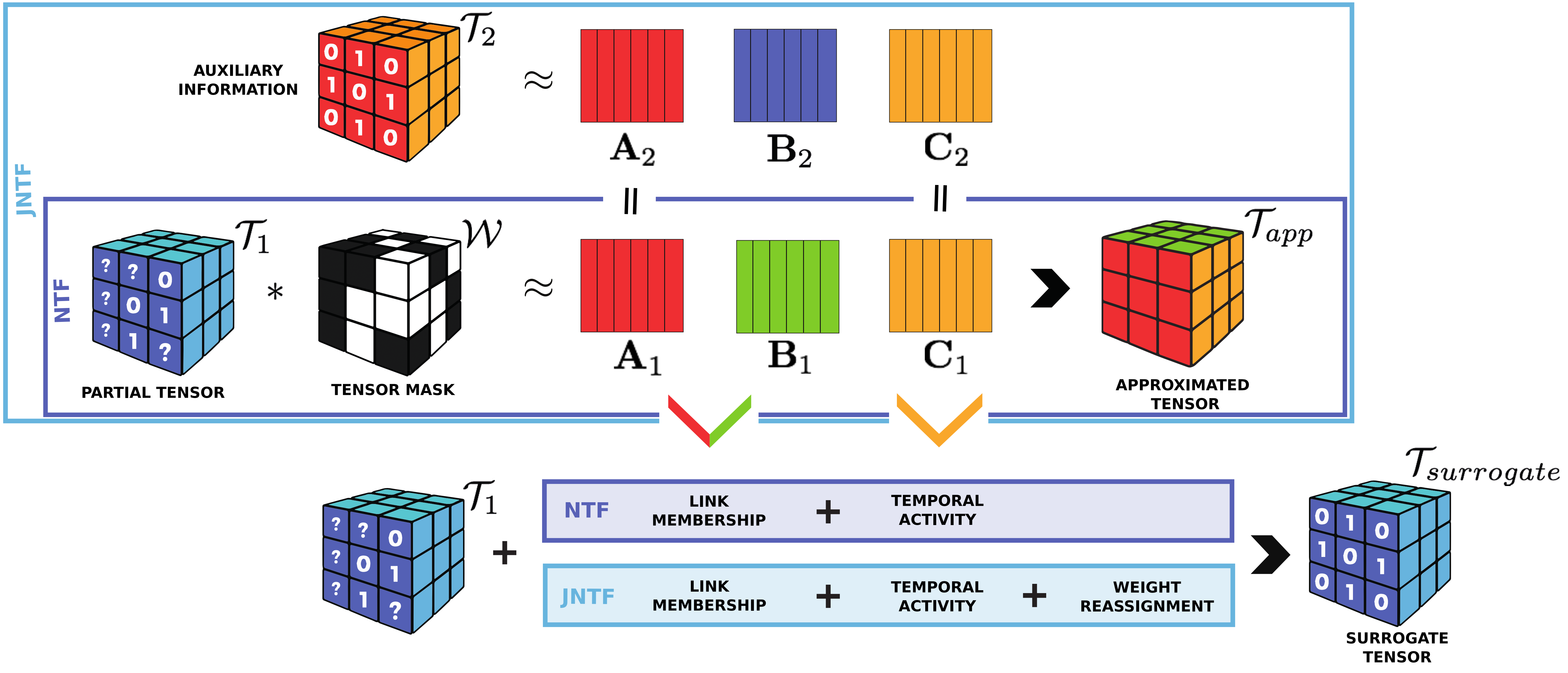
$$T \approx \tilde{\mathcal{T}} = \sum_{r=1}^{R} \mathcal{S}_r = \sum_{r=1}^{R} \mathbf{a_r} \circ \mathbf{b_r} \circ \mathbf{c_r}$$



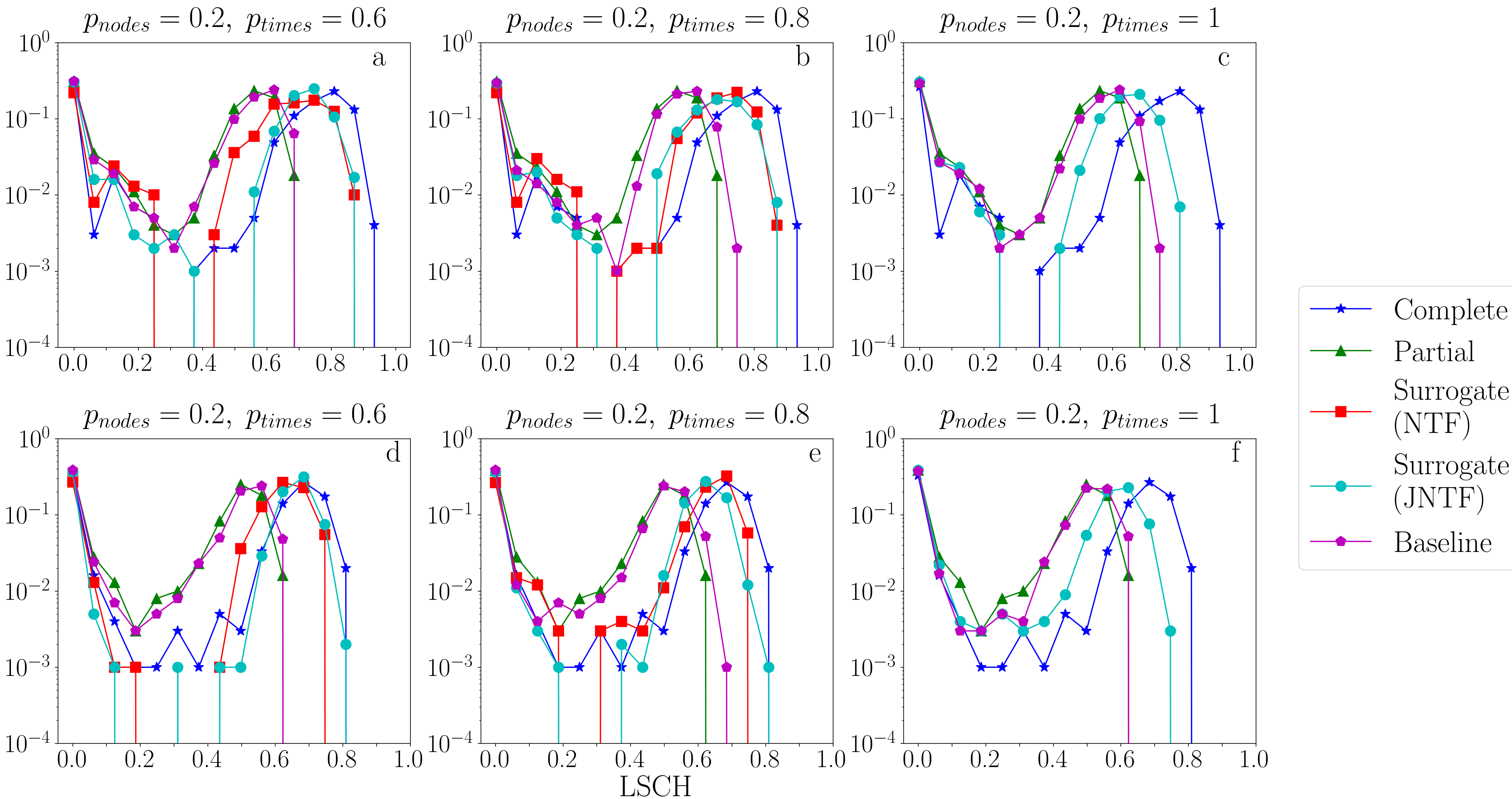
$$t_{ijk} \approx \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}$$

NTF and JNTF

Auxiliary information = approximate node locations



Using NTF & JNTF



NB: no information on group structure

> Using (incomplete) contact diary data

Many data sets come from contact diaries:

- how do the reported contacts compare with contacts measured with sensors?

- how do simulations using different types of data compare?

Comparing data from different sources

Simultaneous data collection among the same persons (high school students)

- wearable sensors
- contact diaries

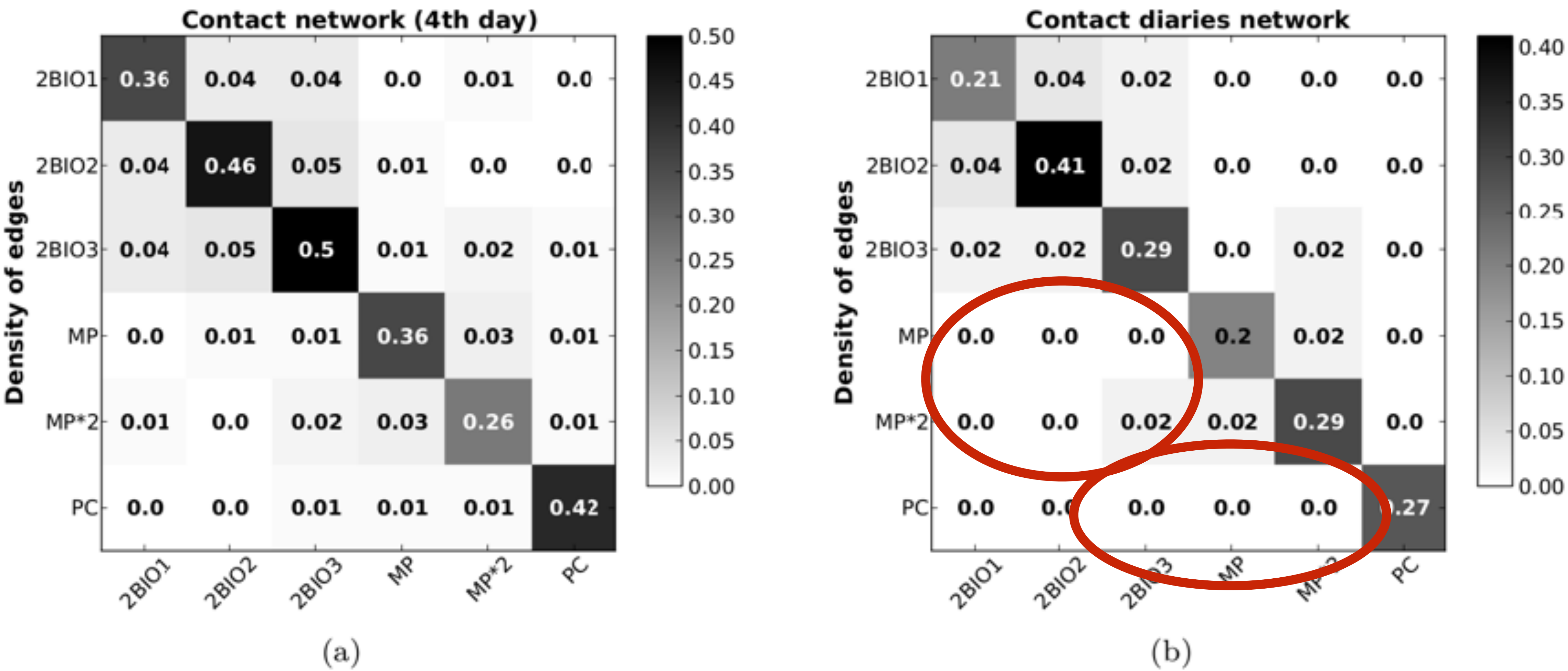
Main results (robust across several case studies):

- smaller participation rate for diaries
- strong underestimation of the number of contacts
- reporting probability increases with contact duration
- overestimation of the contact durations

- long enough contacts are all reported in diaries
- overall structure of network maintained, but less cohesive

T. Smieszek et al., How should social mixing be measured? Comparing web-based survey and sensor-based methods. BMC Infectious Diseases 14:136 (2014)
R. Mastrandrea, J. Fournet, A. Barrat. Contact Patterns in a **High School**: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. PLOS ONE 10:e0136497 (2015)
M. Leecaster et al., Estimates of Social Contact in a **Middle School** Based on Self-Report and Wireless Sensor Data, PLoS ONE 11(4): e0153690 (2016)
T. Smieszek et al., Contact diaries versus wearable proximity sensors in measuring contact patterns at a **conference**: method comparison and participants' attitudes, BMC Infectious Diseases 16:341 (2016)

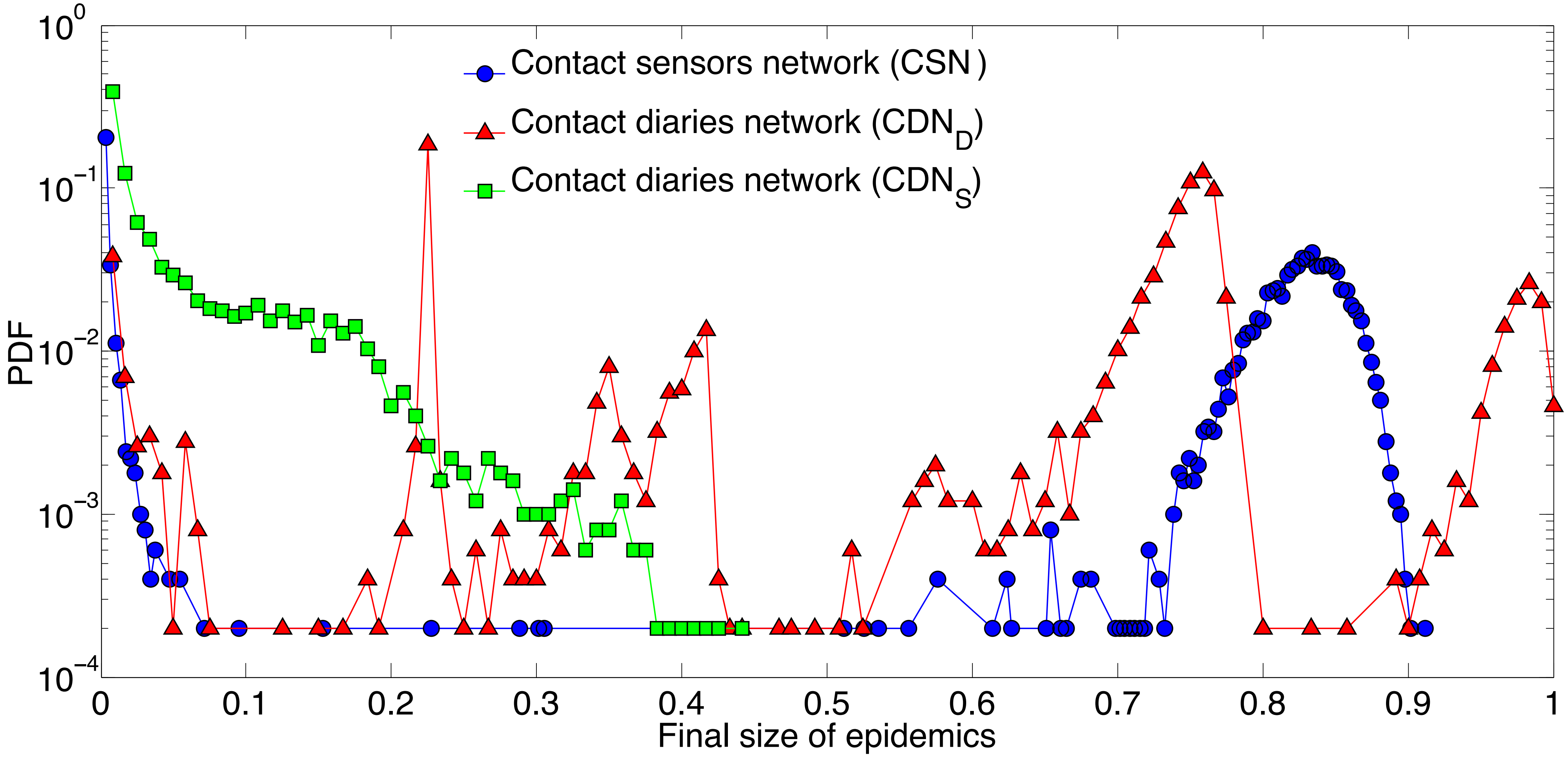
Comparing data from different sources



- Similar mixing patterns (contact matrices)
- Less contacts between classes in the contact diaries

R. Mastrandrea, J. Fournet, A. Barrat. Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. PLOS ONE 10:e0136497 (2015)

Impact on simulations of spreading processes



Several ingredients

- population sampling
- less cohesive structure/stronger community structure
- distribution of weights

R. Mastrandrea, A. Barrat. How to estimate epidemic risk from incomplete contact diaries data?
PLoS Comput Biol, 12: e1005002 (2016)

How to compensate?

Issues

- population sampling

    surrogate nodes and links (between missing and present nodes) in order to preserve the contact matrix of densities

- no links between some class pairs (unrealistic)

    replace by realistic values

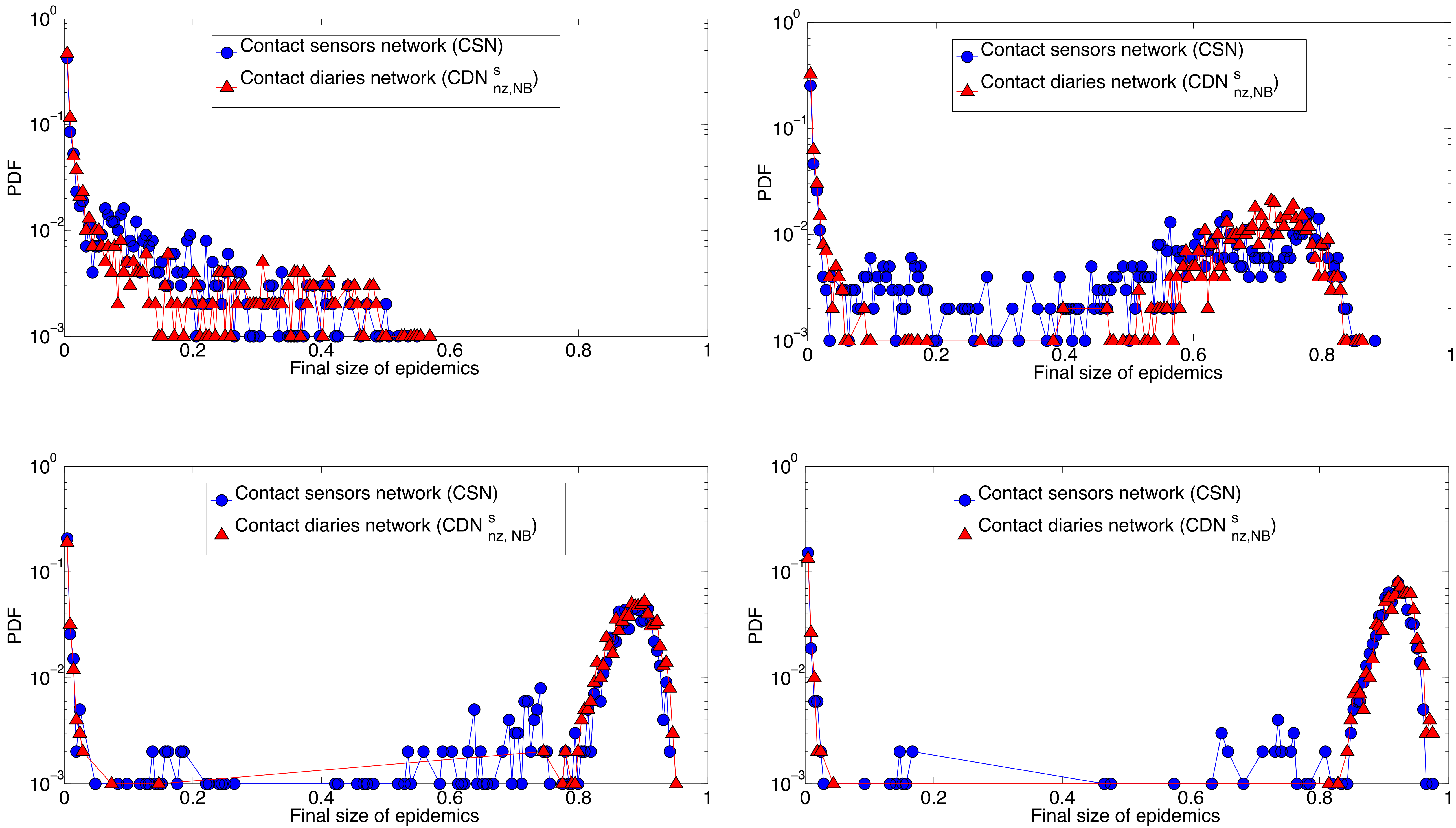
- unknown weights

    use the empirical distribution of weights
    or
    use a pool of publicly available data

Obtention of a surrogate contact network
Simulations of spreading processes on these surrogate data
Comparison with spreading processes on sensor contact data

R. Mastrandrea, A. Barrat. How to estimate epidemic risk from incomplete contact diaries data? PLoS Comput Biol, 12: e1005002 (2016)

Results of simulations of an SIR model



R. Mastrandrea, A. Barrat. How to estimate epidemic risk from incomplete contact diaries data? PLoS Comput Biol, 12: e1005002 (2016)

Using data from contact diaries Summary

- Possibility to *compensate* for low response rate and underreporting through surrogate nodes and links
- Use *extra knowledge* of population structure (to know if link densities are realistic or not)
- Use *publicly available data* on contact durations

NB:

- **case study** of high school data, need to validate with other combined data sets
- for data concerning friendship relations, much worse sampling rate => procedure is not sufficient, sensor data is needed (J. Fournet & A. Barrat, Sci Rep 2017)

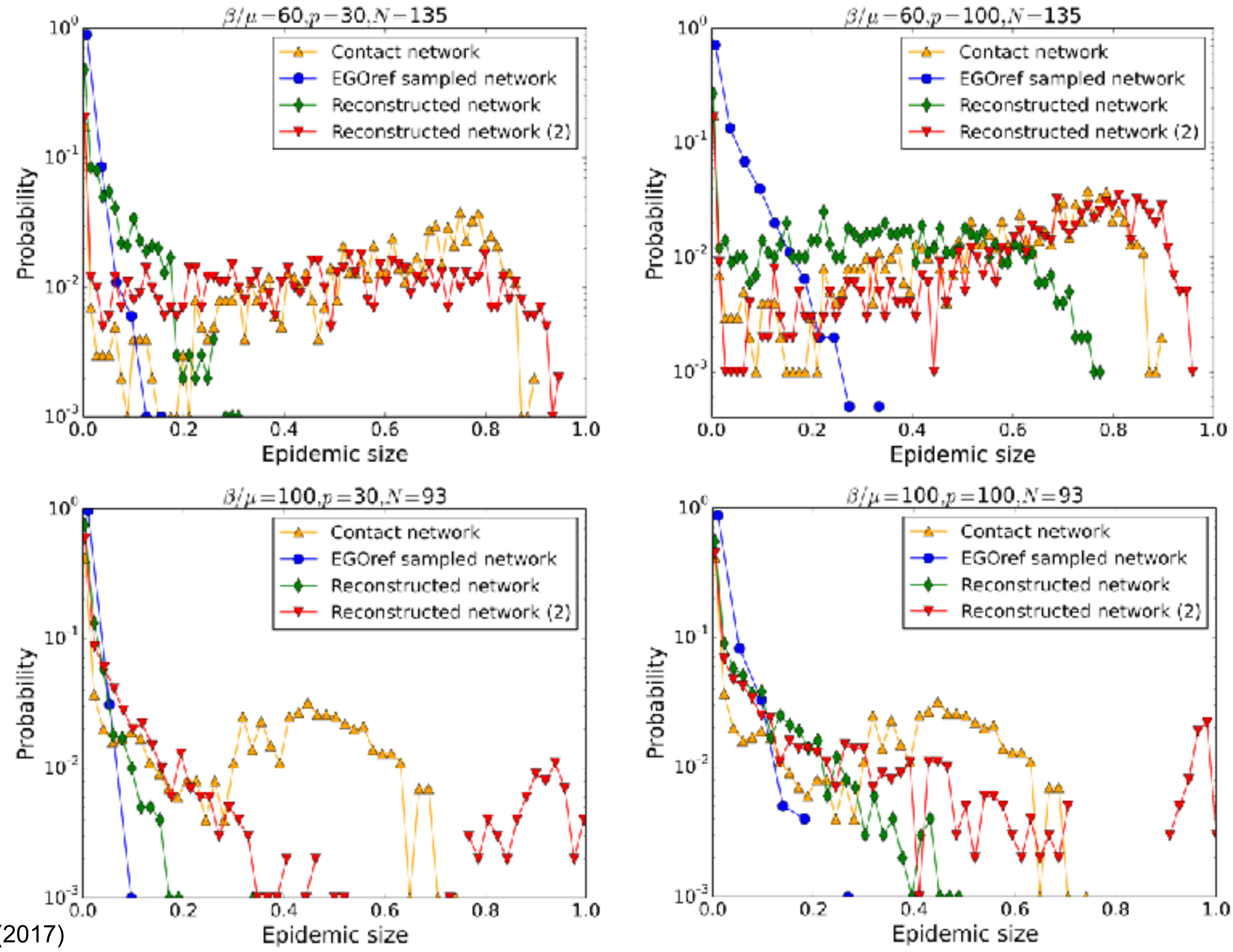
R. Mastrandrea, A. Barrat. How to estimate epidemic risk from incomplete contact diaries data? PLoS Comput Biol, 12: e1005002 (2016)

Non-uniform link sampling

- Simple reconstruction method from biased data
- Sensor data from one group + rescaling of density contact matrix

High-school

Offices



J. Fournet, A. Barrat, Sci. Rep. 7:9975 (2017)

Other/ perspectives / in progress

- “**Atlas**” **of human interactions**
  - Collection of datasets (www.sociopatterns.org)
  - Comparing with self-reported contacts in various contexts
- **How to feed data into models**
  - Agent-based models of contact patterns reproducing empirical phenomenology
  - Using co-presence data (easier to get) instead of contact data
  - Case of spatial sampling, case of non-uniform sampling
  - Novel data representations
  - Finding structures in data
  - Bridging scales (contacts + mobility)
  - Effect of sampling on containment strategy evaluation
  - Other processes
- **Public health**
  - Contact patterns and health-related behavior (hand-washing)
  - Merging contact data and virological data
  - Designing and testing data-driven mitigation strategies

SocioPatterns

Perspectives

Social networks of animals,
animal behaviour
cognitive sciences

with V. Gelardi, N. Claidière, J. Fagot, CNRS & Aix-Marseille University, France

SocioPatterns

Temporal networks: Still very open field!

Data
Structures in data
Incompleteness of data
Models
Processes on temporal networks
…

Datasets: www.sociopatterns.org/datasets

Physics Reports

Volume 519, Issue 3, October 2012, Pages 97–125

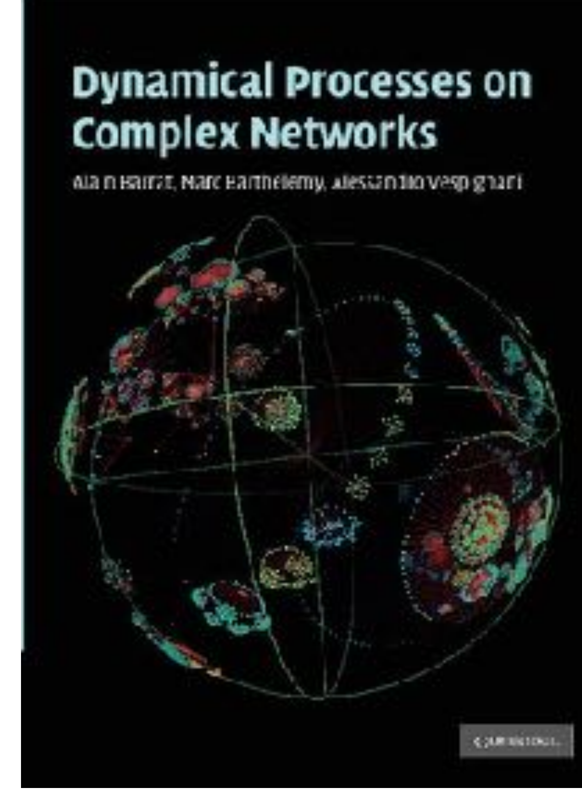
Temporal networks

Petter Holme[a, b, c], Jari Saramäki[d]

Show more

http://dx.doi.org/10.1016/j.physrep.2012.03.001

Get rights and content



The European Physical Journal B

September 2015, 88:234

Modern temporal network theory: a colloquium

Authors | Authors and affiliations

Petter Holme

Colloquium

First Online: 21 September 2015

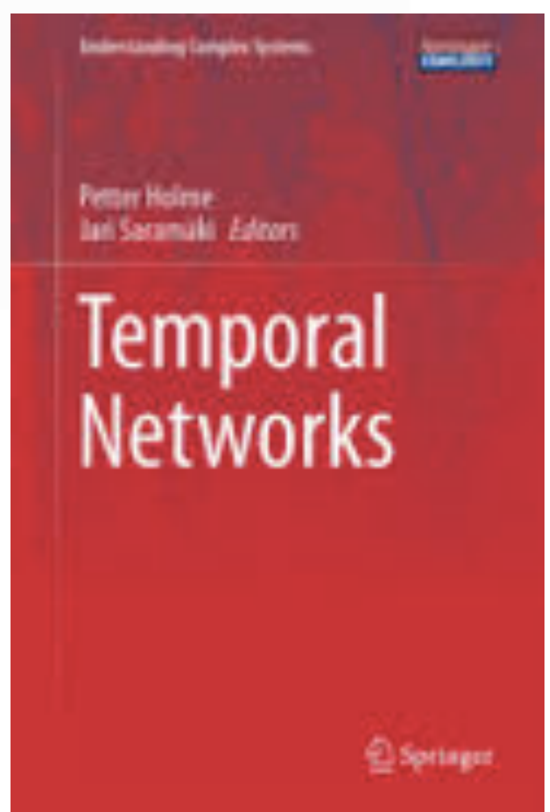
DOI: 10.1140/epjb/e2015-60657-4

Cite this article as:

Holme, P. Eur. Phys. J. B (2015) 88: 234. doi:10.1140/epjb/e2015-60657-4

Part of the following topical collections:

- Topical issue: Temporal Network Theory and Applications



SpringerLink

Search

Home • Contact Us

Look Inside | Get Access

Regular Article The Dynamics Of Networks: General Theory

The European Physical Journal Special Topics

September 2013, Volume 222, Issue 6, pp 1295-1309

First online: 18 September 2013

Empirical temporal networks of face-to-face human interactions

A. Barrat, C. Cattuto, V. Colizza, F. Gesualdo, L. Isella, E. Pandolfi, J.-F. Pinton, L. Ravà, C. Rizzo and 4 more

SocioPatterns.org