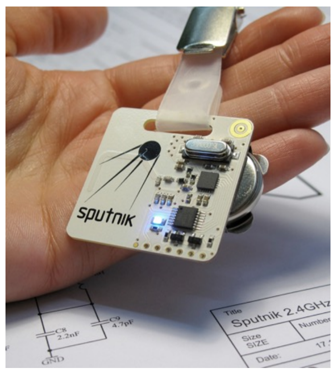
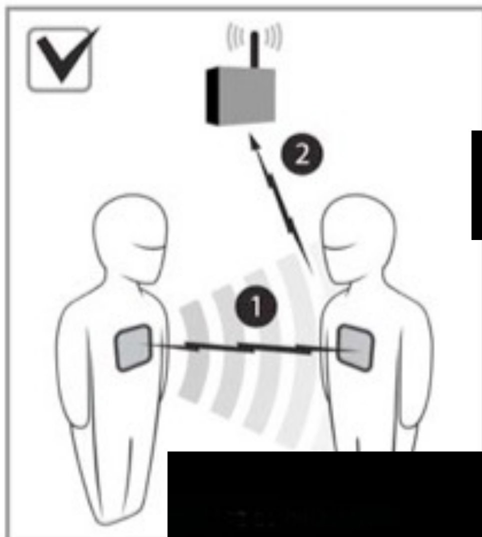
LIFE DATA EPIDEMIOLOGY

lect. 5: temporal networks

Chiara Poletto
polettoc@gmail.com

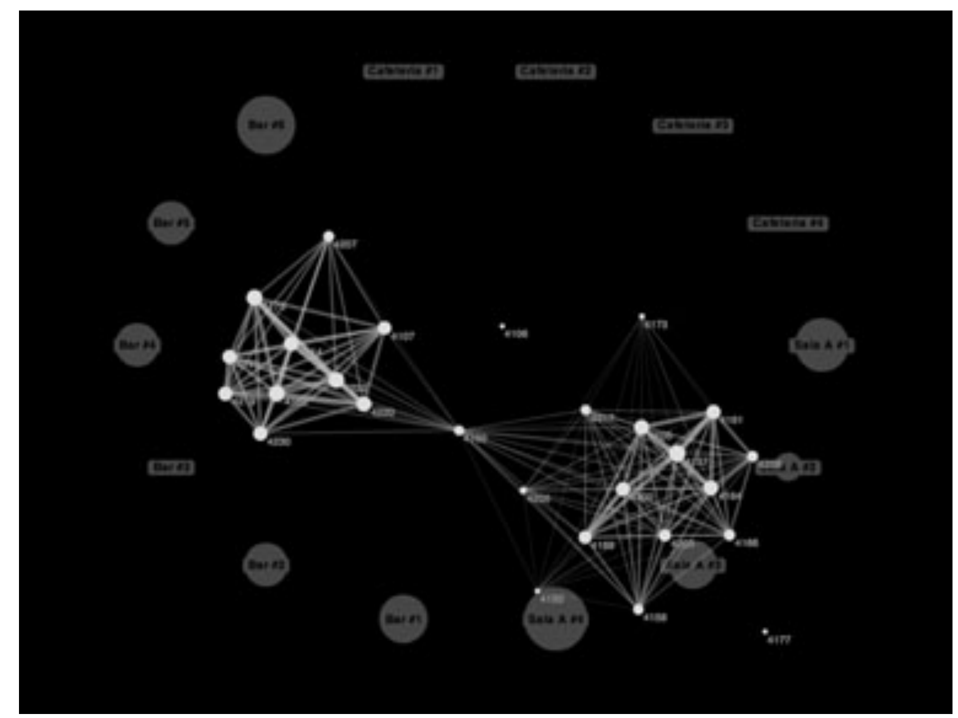
high resolution network data





face-to-face contacts

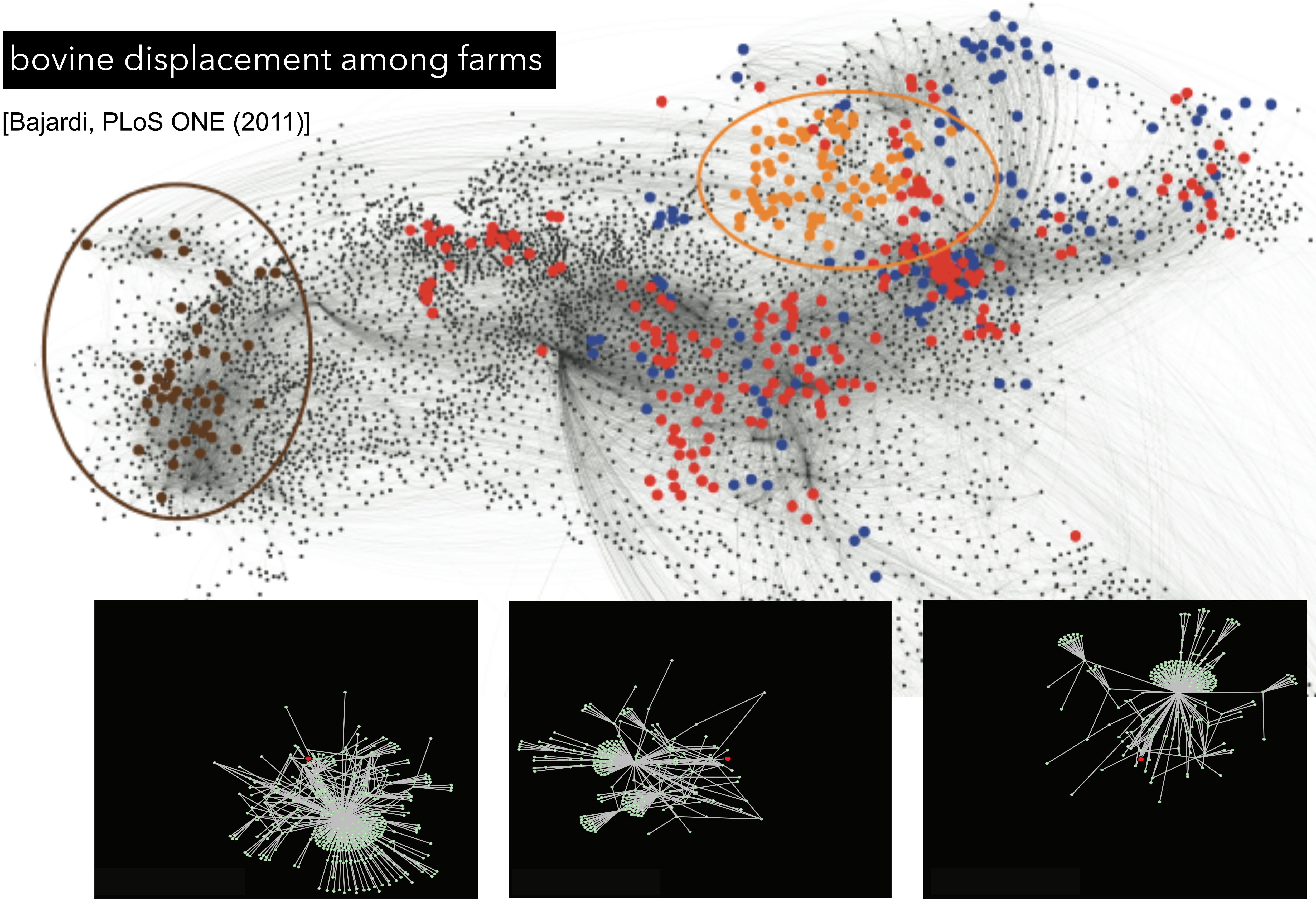
RFID technology



schools - workplaces - hospitals - museums - conferences - households - rural Africa

[**Sociopatterns.org**]

high resolution network data



[Bajardi, PLoS ONE (2011)]

high resolution network data

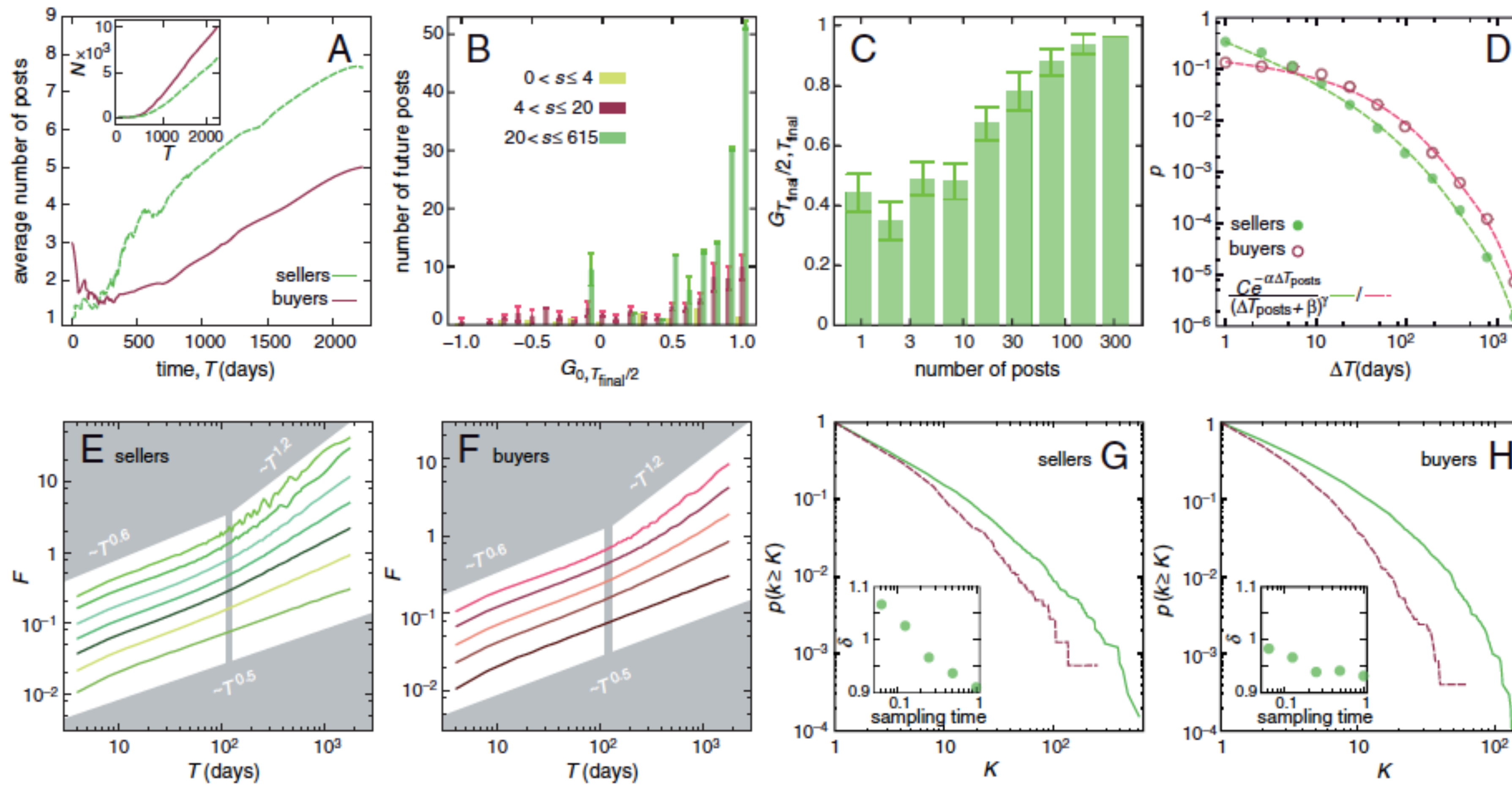


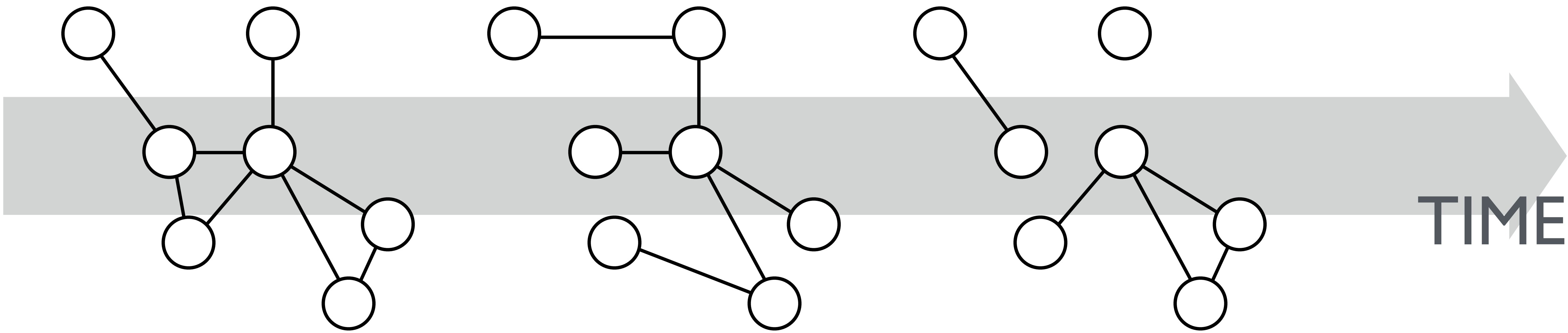
Fig. 1. Statistics of the dynamics of the community. (*A*) Time evolution of the average number of posts by sex buyers and about sex buyers. The *Inset* shows the growth in the number of sex sellers and sex buyers in the data. (*B*) The number of new posts according to the previous average grade at $T_{\text{final}}/_{2=1,116 \text{ days}}$ for three different activity levels, or total number of posts, s. The R^2-values of these data are 0.19 ($0 < s \le 4$), 0.29 ($4 < s \le 20$), and 0.33 ($20 < s$). (*C*) The average future grade of sellers as a function of their number of contacts at half of the total sampling time (the data is logarithmically binned along the abscissa). (*D*) Shows the distribution of the time elapsed between two posts T_{posts} for buyers and sellers. Many posts were written during the same day, respectively, $p(T_{\text{posts}} = 0) = 0.495$ and $p(T_{\text{posts}} = 0) = 0.246$. The distributions are well fitted by $p(T_{\text{posts}}) = C\exp(-\alpha T_{\text{posts}}) = (T_{\text{posts}} + \beta)^{\gamma}$, with: $C = 2.9 \pm 0.5$ days$^{\gamma}$, $\alpha = 0.0023 \pm 0.0001$ days^{-1}, $\beta = 3.1 \pm 0.4$ days, and $\gamma = 1.49 \pm 0.04$ (for sellers); and $C = 12 \pm 8$ days$^{\gamma}$, $\alpha = 0.0021 \pm 0.0002$ days^{-1}, $\beta = 18 \pm 4$ days, and $\gamma = 1.5 \pm 0.1$ (for buyers). (*E*) and (*F*) shows statistics the DFA fluctuation function as a function of the time-scale ΔT for sellers and buyers, resp. The different curves correspond to different activity levels—from bottom to top they represent less than 3, 3–7, 8–20, 21–54, 55–148, 149–403, and more than 403 posts (about sellers or from buyers) resp. *Black Lines* are inserted for reference. $T^{1/2}$ corresponds to uncorrelated interaction. (*G*) and (*H*) show degree distributions for sex sellers (*G*) and buyers (*H*) cumulative degree distributions for the full sampling time (*Solid Line*) and a yearlong window (starting one year after the full dataset; *Dashed Line*) for sex sellers and -buyers, resp. The *Insets* show the exponent of preferential attachment (Eq. 1).

internet mediated prostitution

sexual contacts between 6,624 escorts and 10,106 sex buyers extracted from an online community

[LEC. Rocha, et al, PNAS 2009]

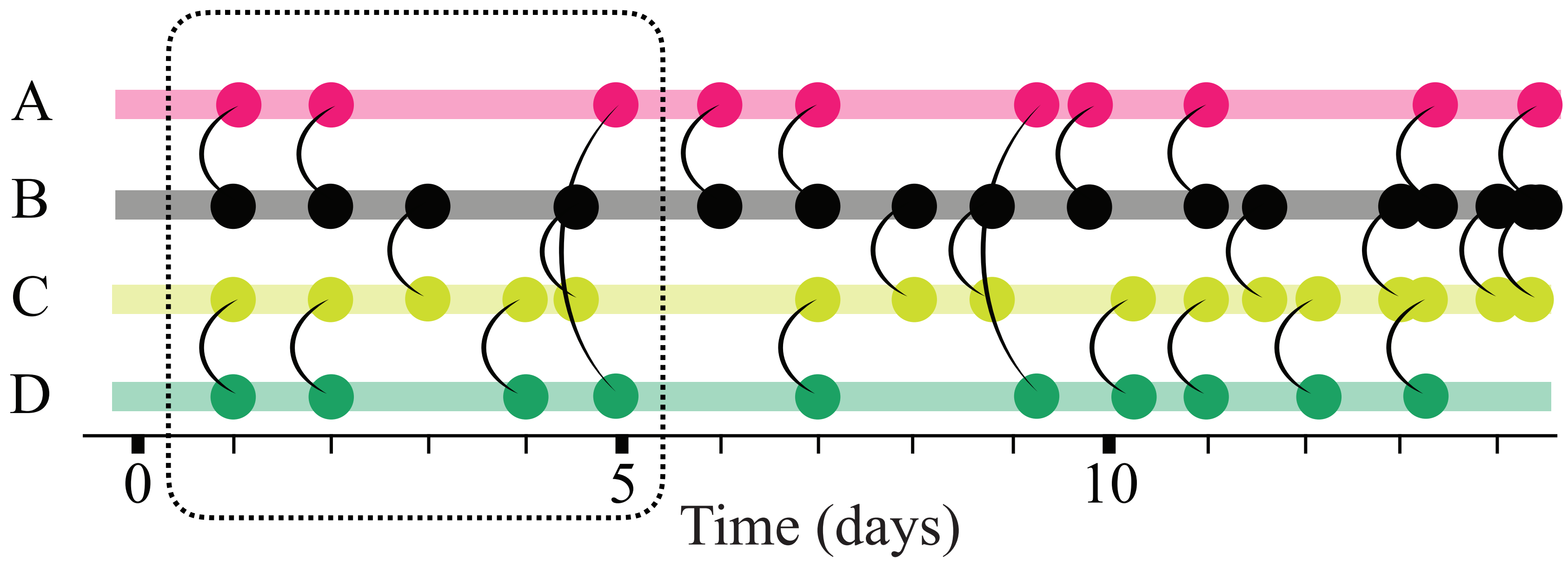
temporal dimension of networks



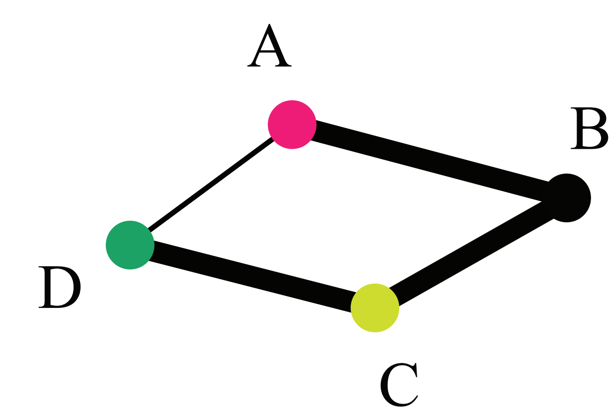
[Holme, Saramaki Phys. Rep. (2012)]

issue #1: visualisation

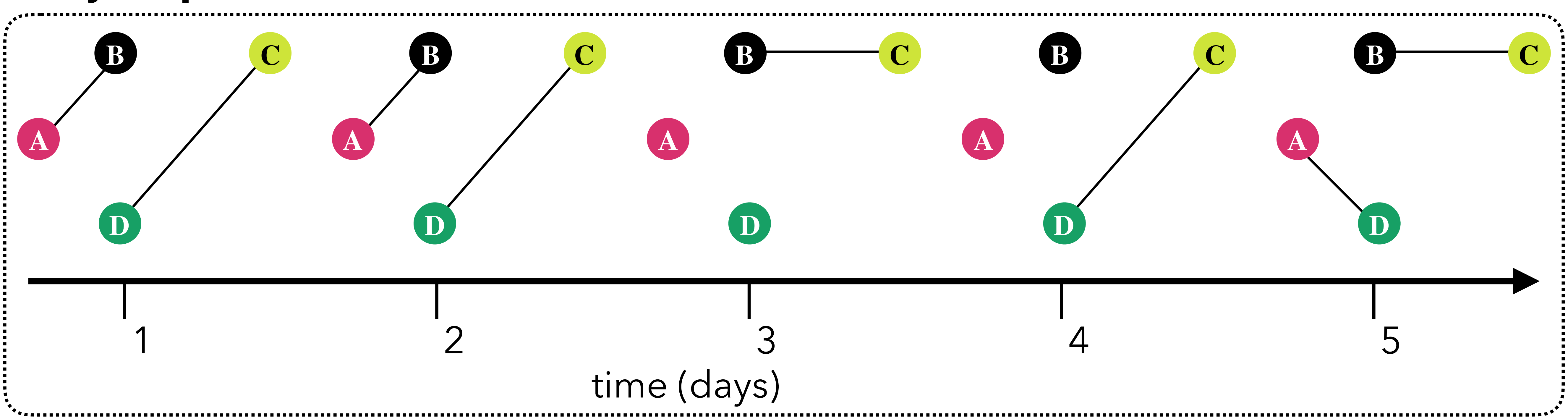
sequence of links: continuous time



weighted aggregated: static

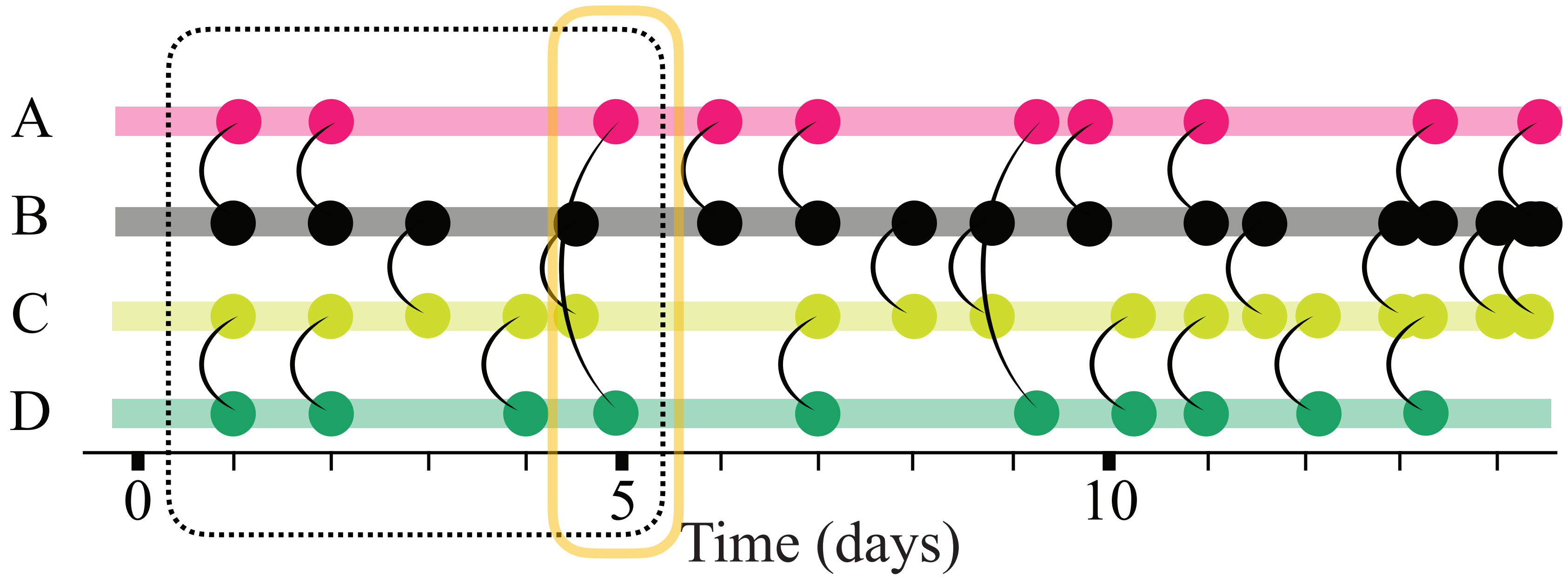


daily snapshot: discrete time

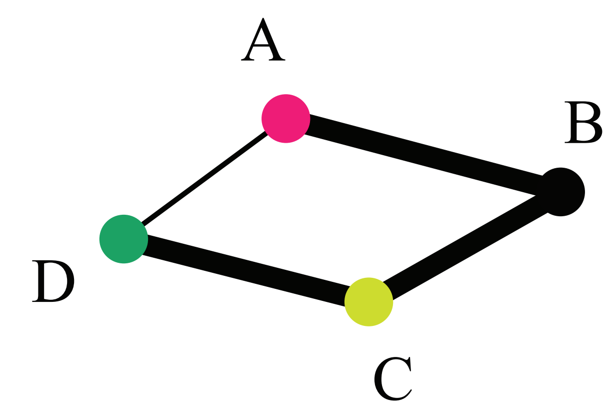


issue #1: visualisation

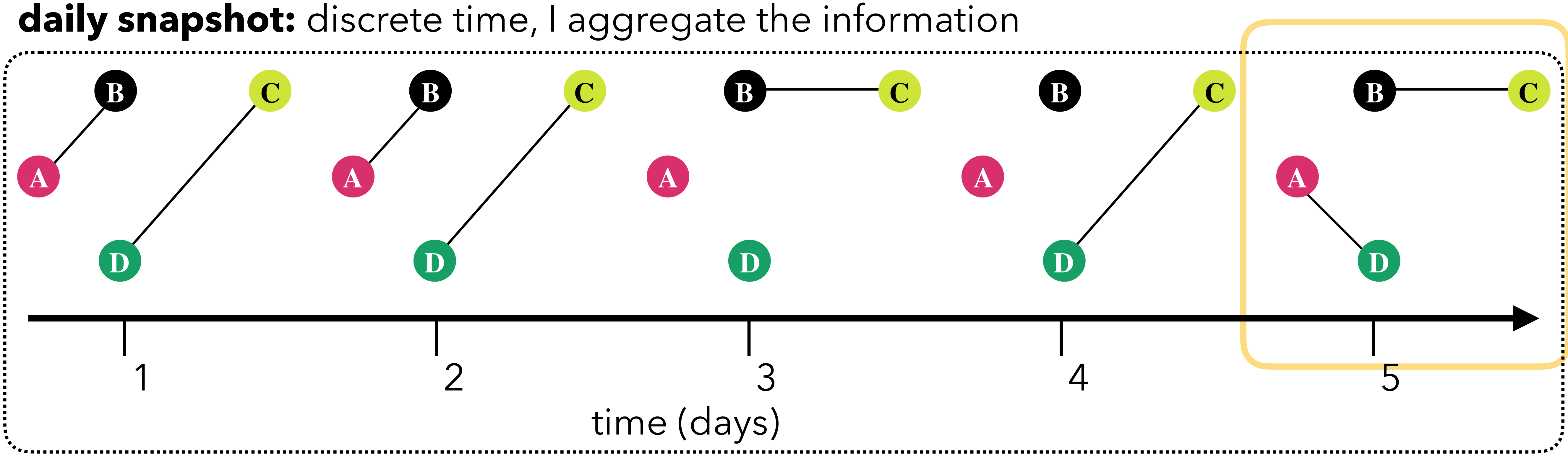
sequence of links: continuous time



weighted aggregated: static

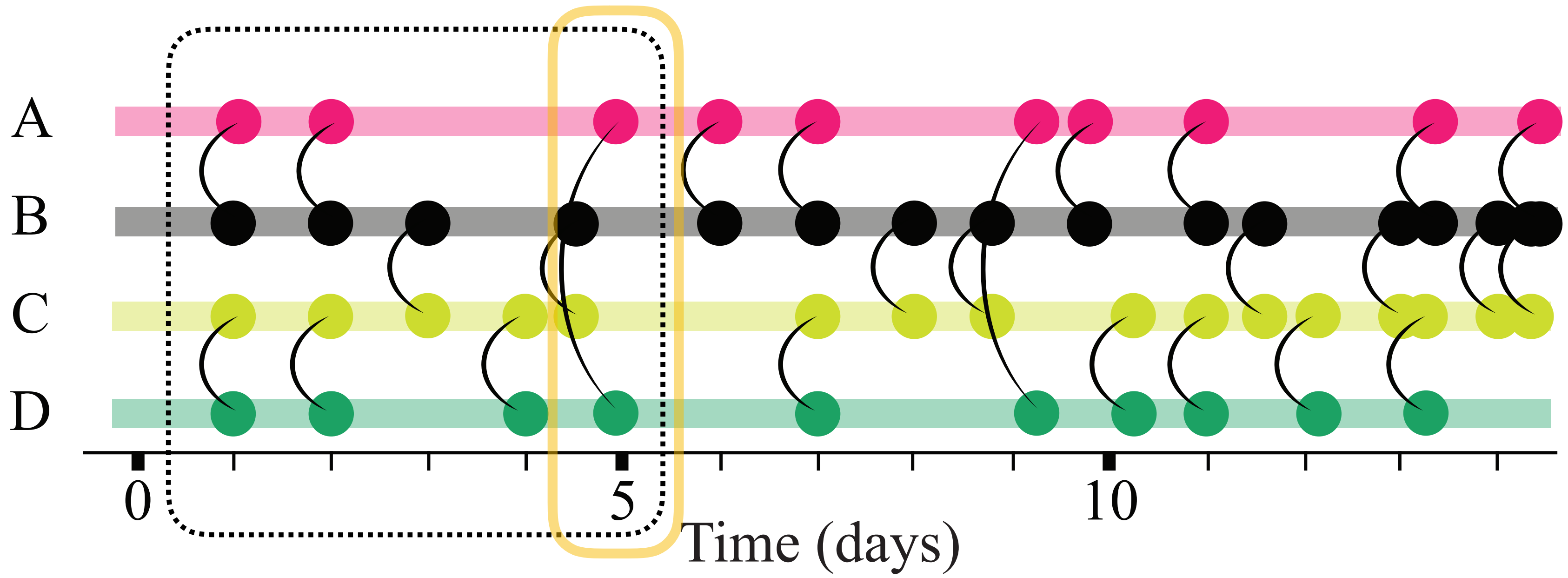


daily snapshot: discrete time, I aggregate the information

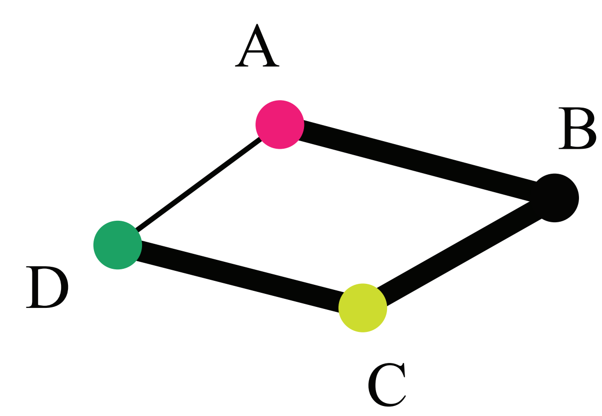


issue #1: visualisation

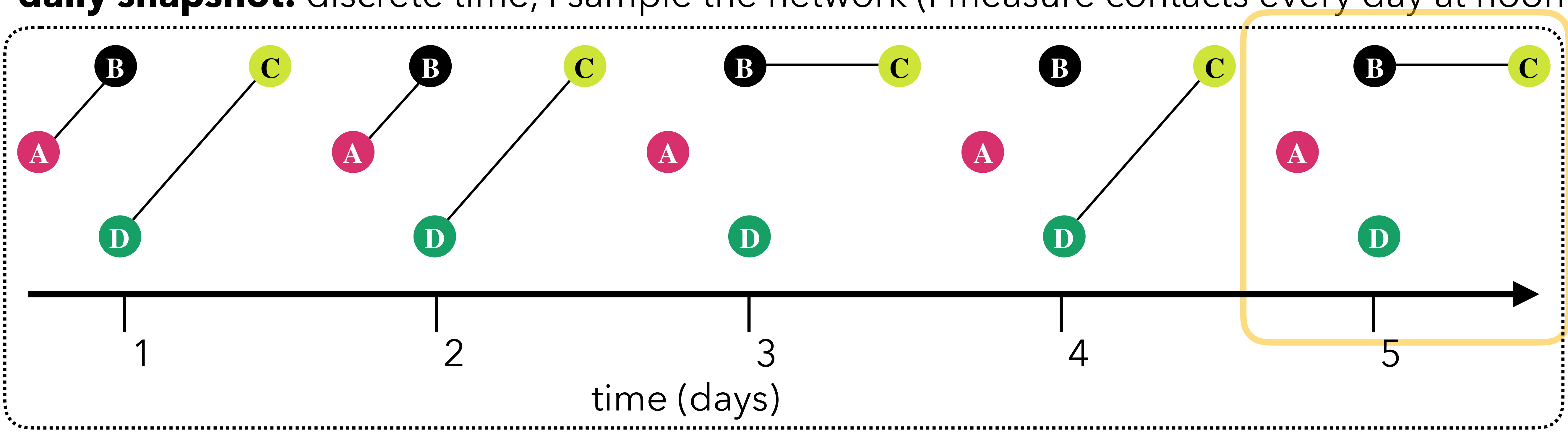
sequence of links: continuous time



weighted aggregated: static



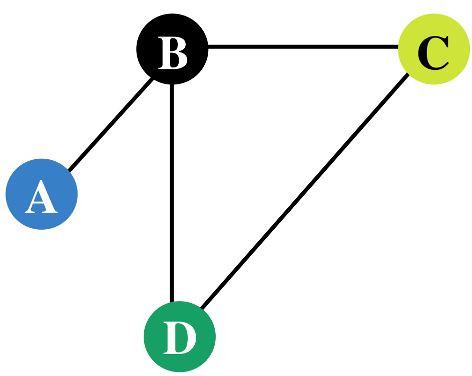
daily snapshot: discrete time, I sample the network (I measure contacts every day at noon)



issue #2: reachability

network reachability:

i is reachable from j if it exists a path from i to j

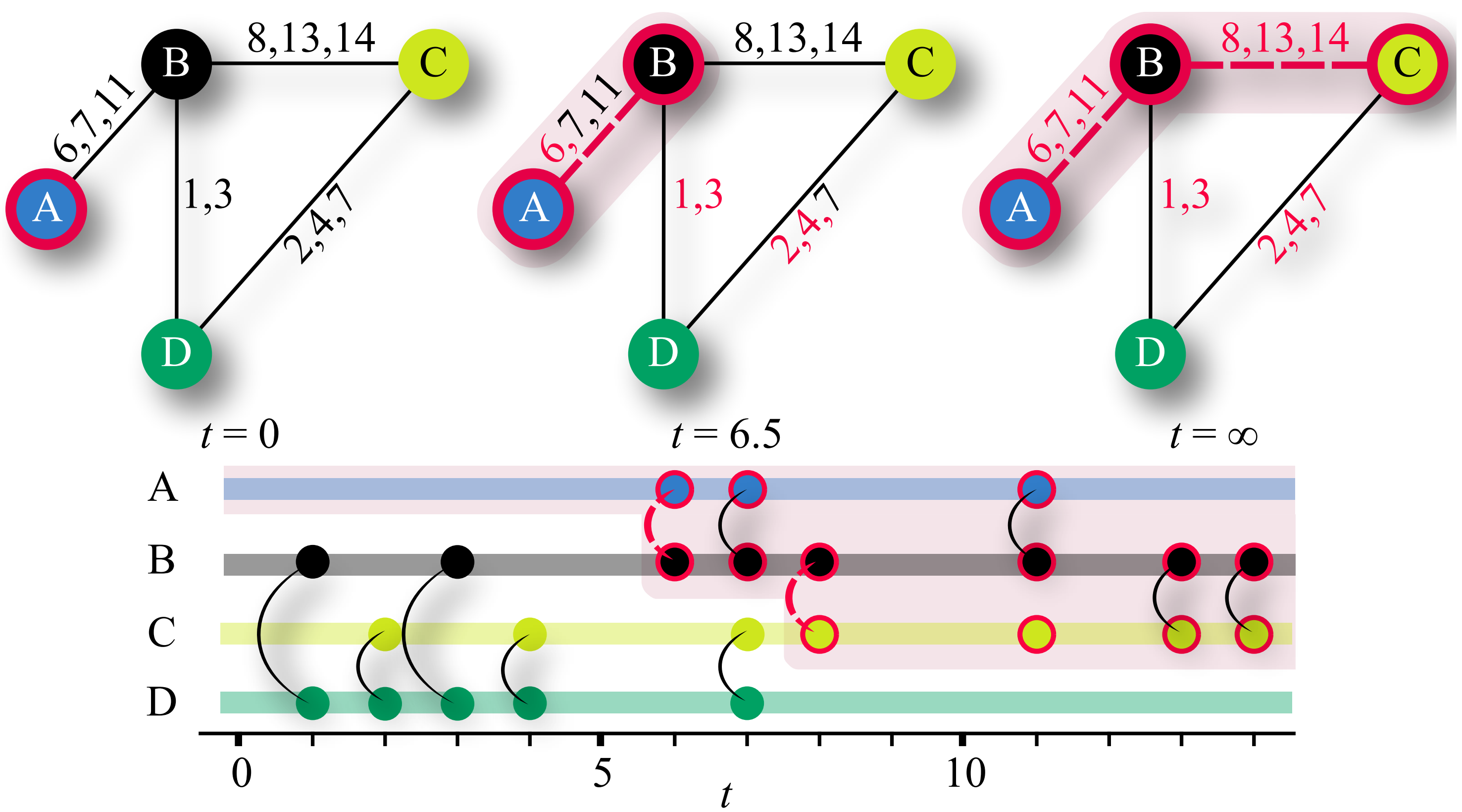


in an undirected static network every node is reachable from every node in its connected component

issue #2: reachability

network reachability:

i is reachable from j if it exists a path from i to j

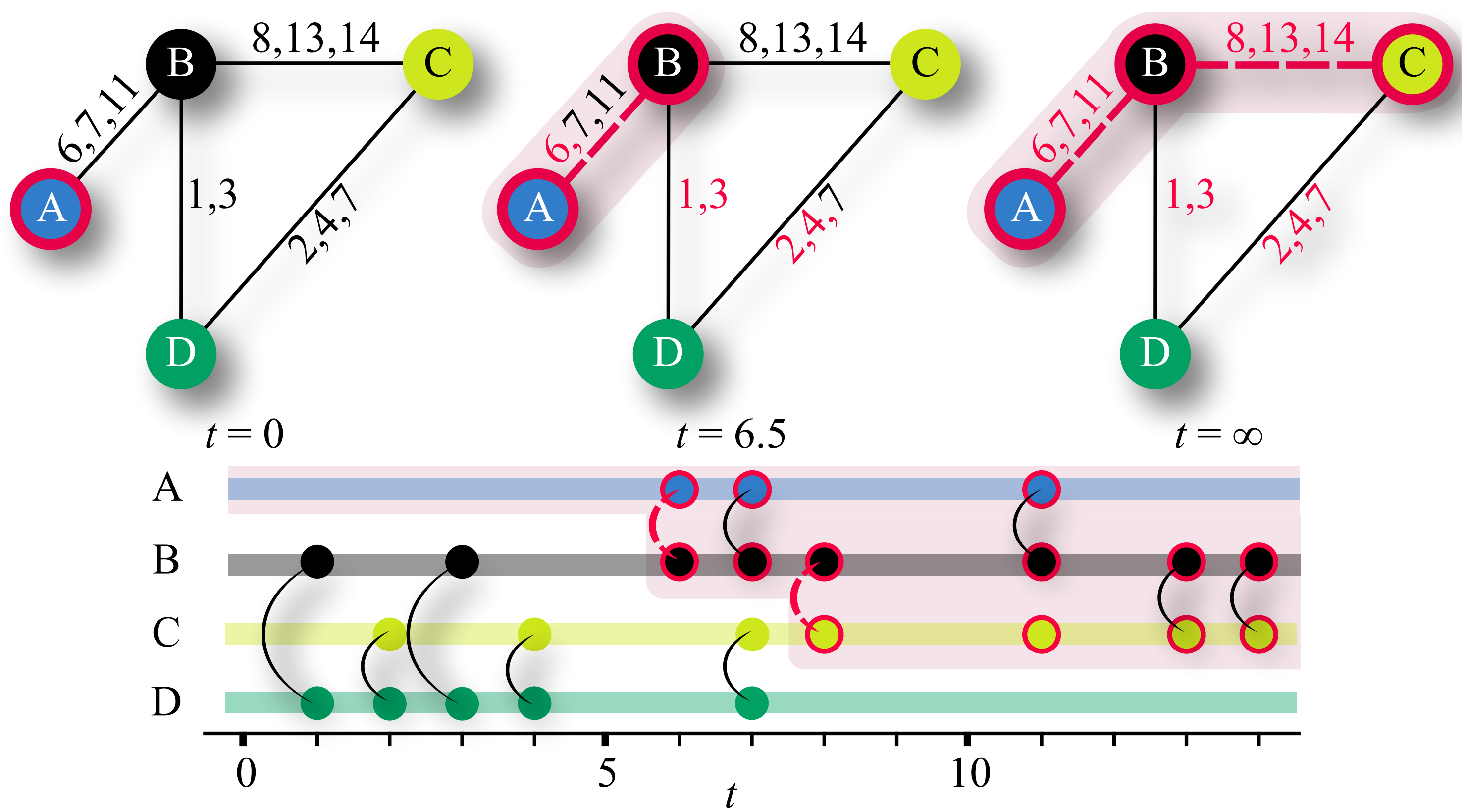


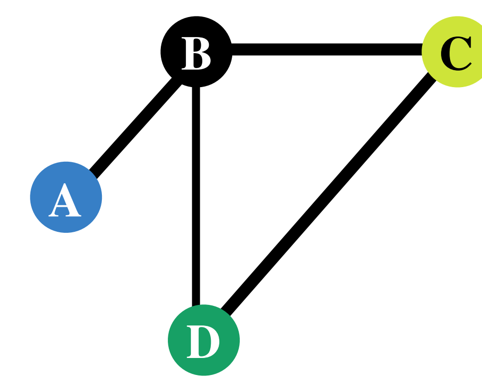
in a undirected temporal network, j is reachable from i only if there exists a **time respecting path** from i to j, i.e. a sequence of contacts that connect i and j with each contact in the path coming after the one before it in time

issue #2: reachability

network reachability:

i is reachable from j if it exists a path from i to j



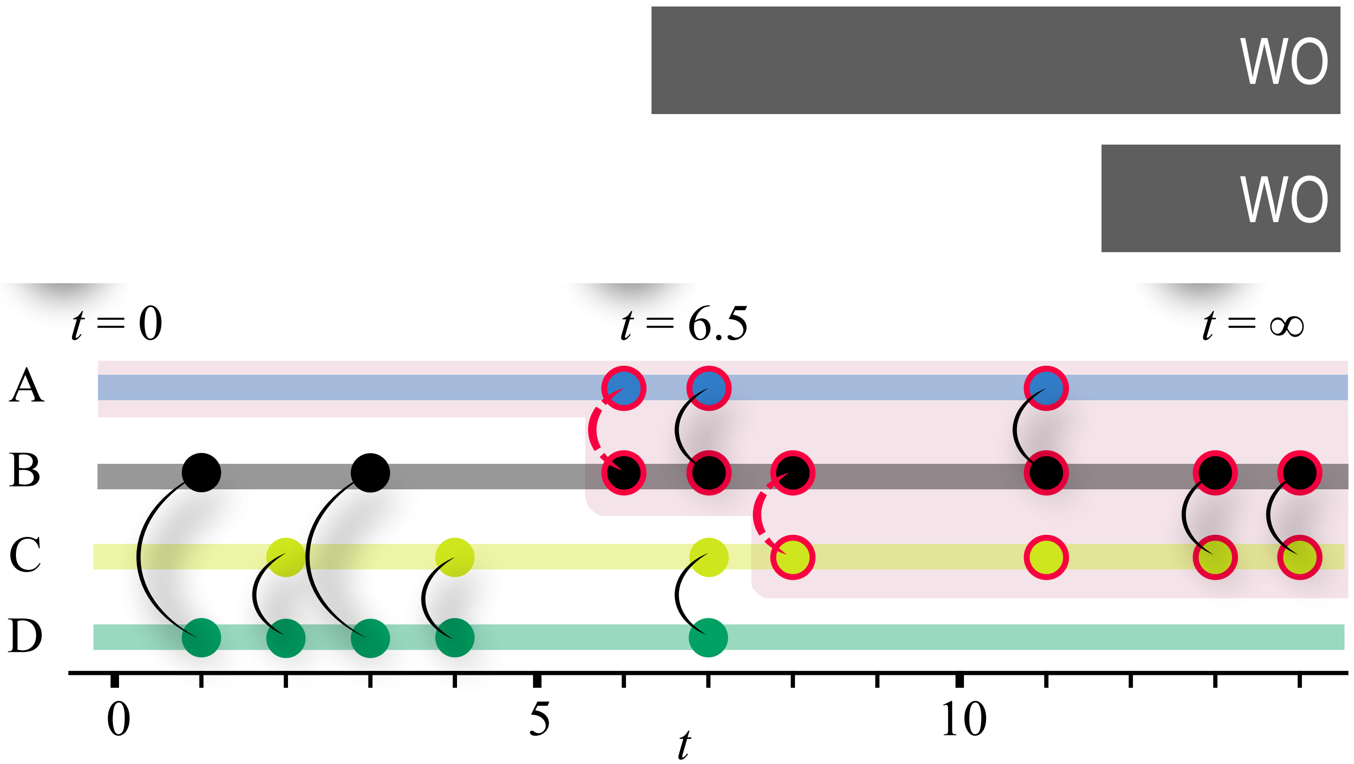


In the weighted aggregated network I lose a lot of information!

in a undirected temporal network, j is reachable from i only if there exists a **time respecting path** from i to j, i.e. a sequence of contacts that connect i and j with each contact in the path coming after the one before it in time

issue #2: reachability

The existence of a time respecting path depends on the window $[t, T]$ of observation



For $t = 6.5$ there is a path from A to C

For $t = 11.5$ there is no path from A to C

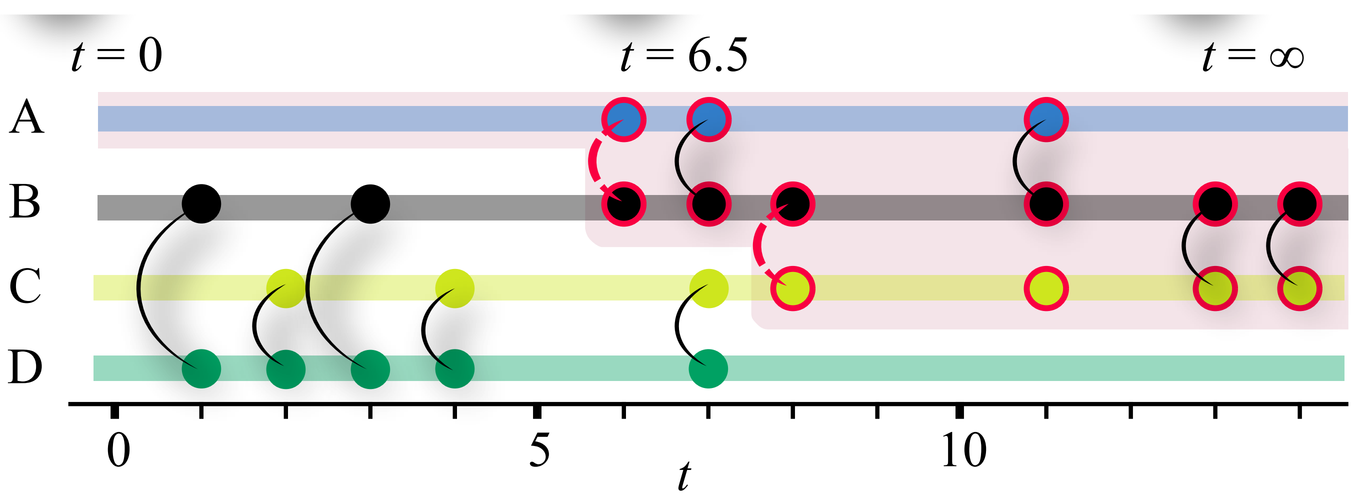
issue #2: reachability

In the window $[t, T]$ a path exist from i to j. Is i able to infect j?

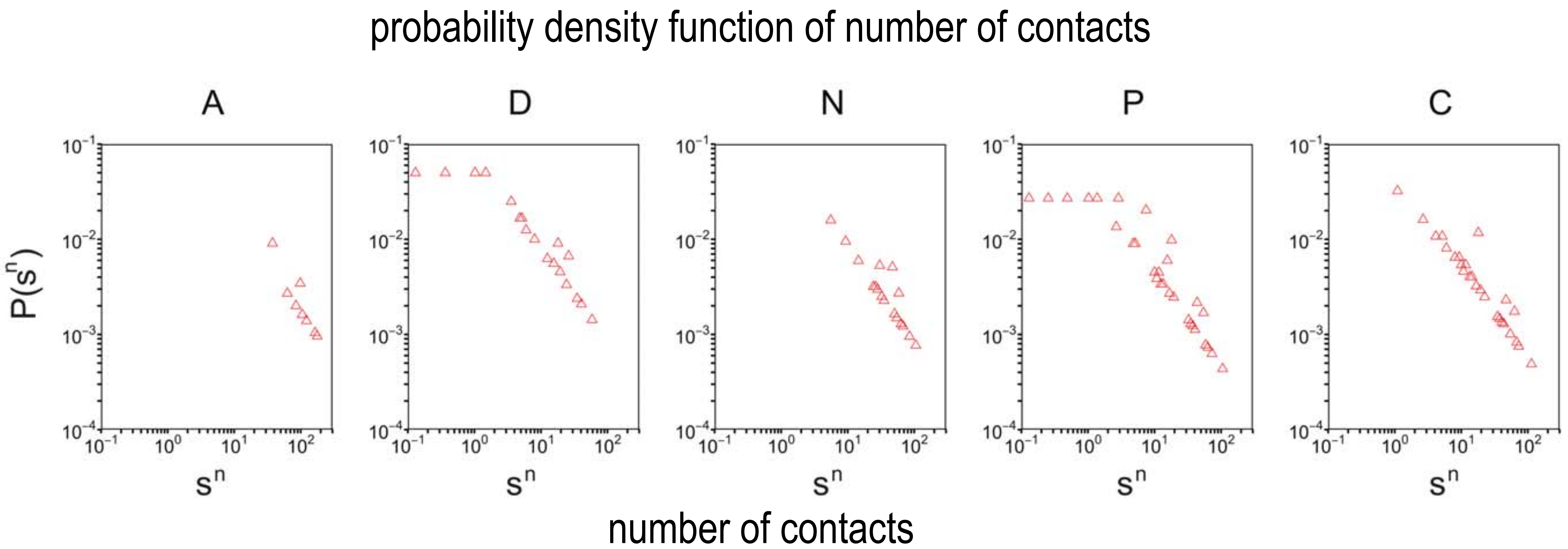
WO

For $\mu^{-1} = 3$ days YES

For $\mu^{-1} = 1$ days NO



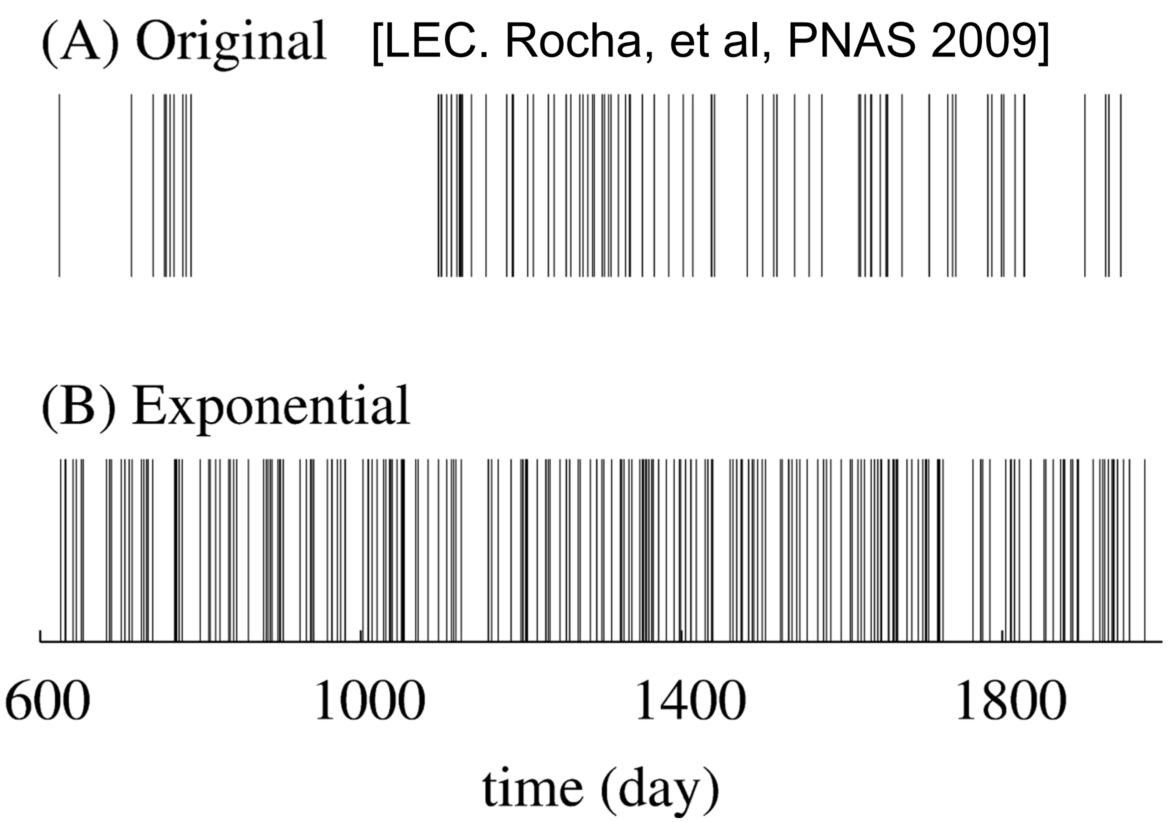
issue #3: contact heterogeneities



[Isella et al PLOS ONE 2011]

Cumulative number of contacts results from activation frequency and number of contacts made at each activation

issue #4: non homogenous activation



more realistic model: $P_E(\tau) = A\tau^{-\alpha}e^{-\tau/\tau_E}$

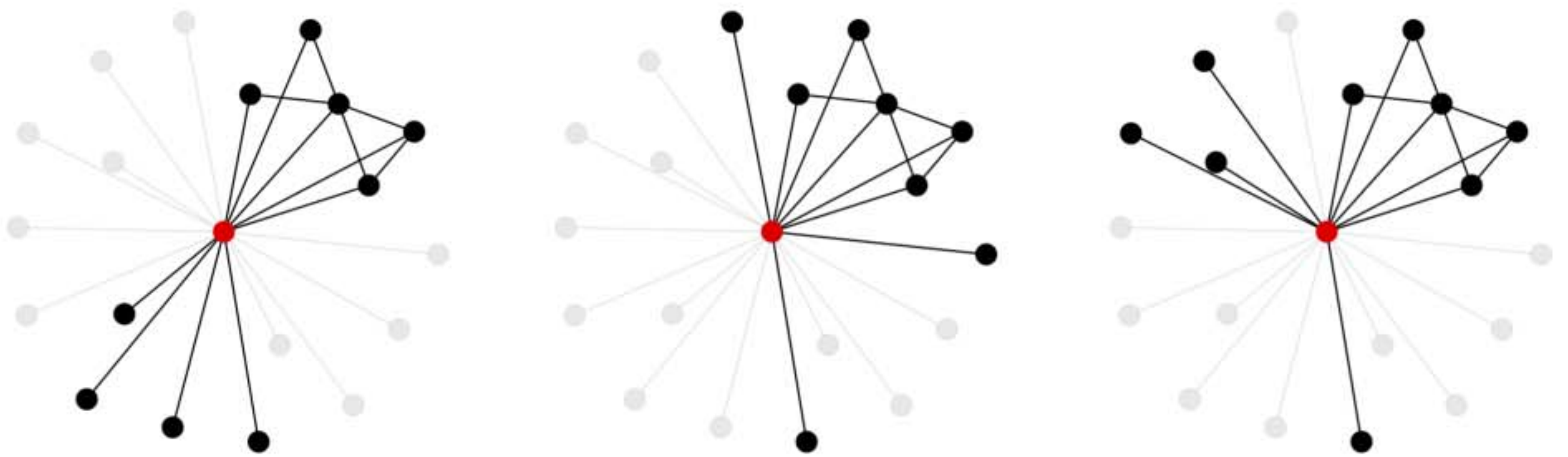
Poisson model: $P_P(\tau) = \frac{e^{-\tau/\langle\tau\rangle}}{\langle\tau\rangle}$

inter-contact time: time from two consecutive activations

human behaviour is bursty

burstiness: broader-than-expected distributions of inter-contact times

issue #5: temporal correlations



[Miritello, et al, Sci Rep 2013]

$k_{i,t}$= degree of i in the network aggregated over the interval $[t - \delta, t]$

$s_{i,t}$= weighted degree of i in the network aggregated over the interval $[t - \delta, t]$

social strategy: $\gamma_{i,t} = \frac{k_{i,t}}{s_{i,t}}$

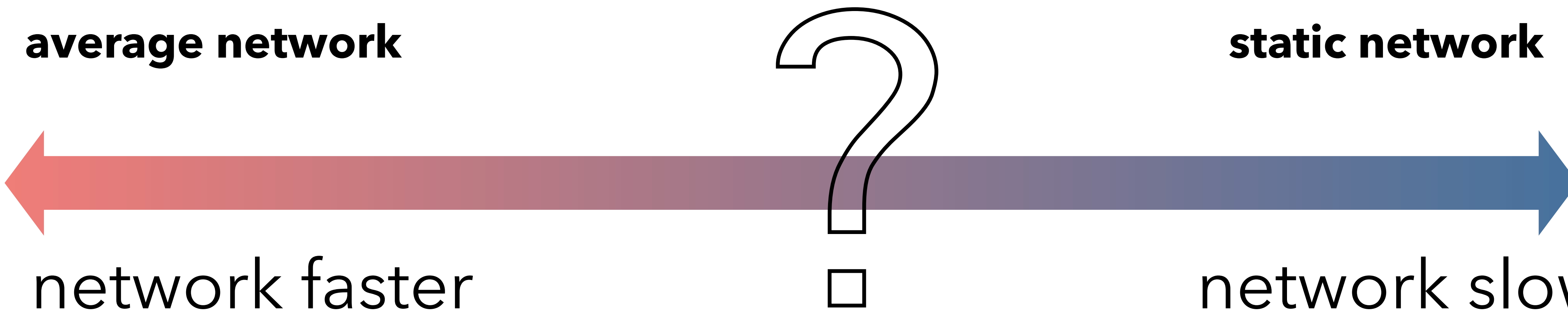
$\gamma \to 0$: memory-driven behavior (a node tends to make contacts always with the same nodes)
$\gamma \to 1$: memoryless behavior (a node shows a more socially exploratory behavior)

importance of contact dynamics

average infectious duration μ^{-1}

average inter contact time τ

average network **static network**

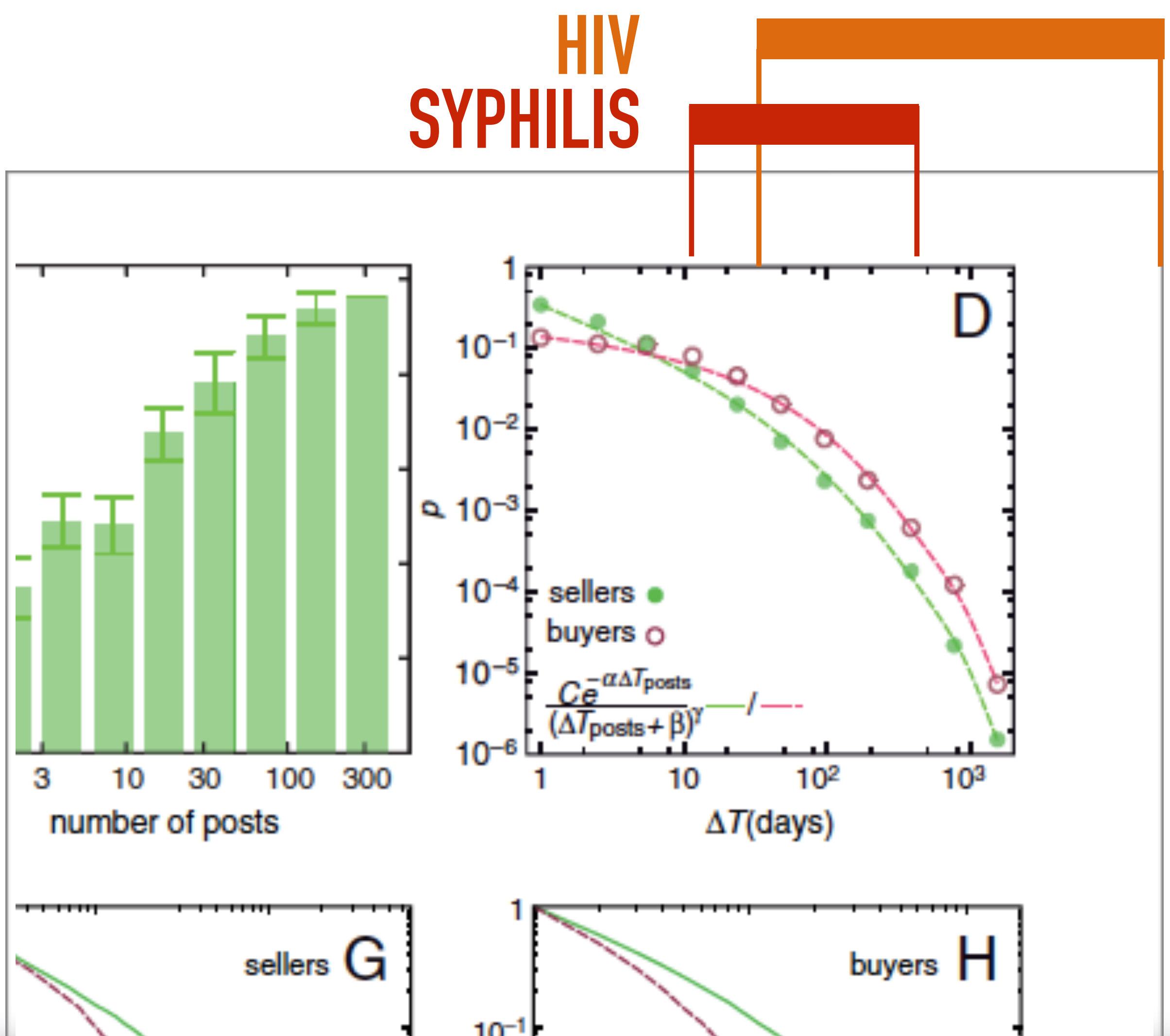


network faster than disease

network slower than disease

importance of contact dynamics

time scale separation not applicable in many cases



internet mediated prostitution

[LEC. Rocha, et al, PNAS 2009]

approaches to temporal network epidemiology

Botom-up: generative models

activity driven model, and its extensions

Top-down approaches: Randomised Reference Models

compare the epidemics on real data with the outcome in suitable null models

approaches to temporal network epidemiology

Botom-up: generative models

activity driven model, and its extensions

Top-down approaches: Randomised Reference Models

compare the epidemics on real data with the outcome in suitable null models

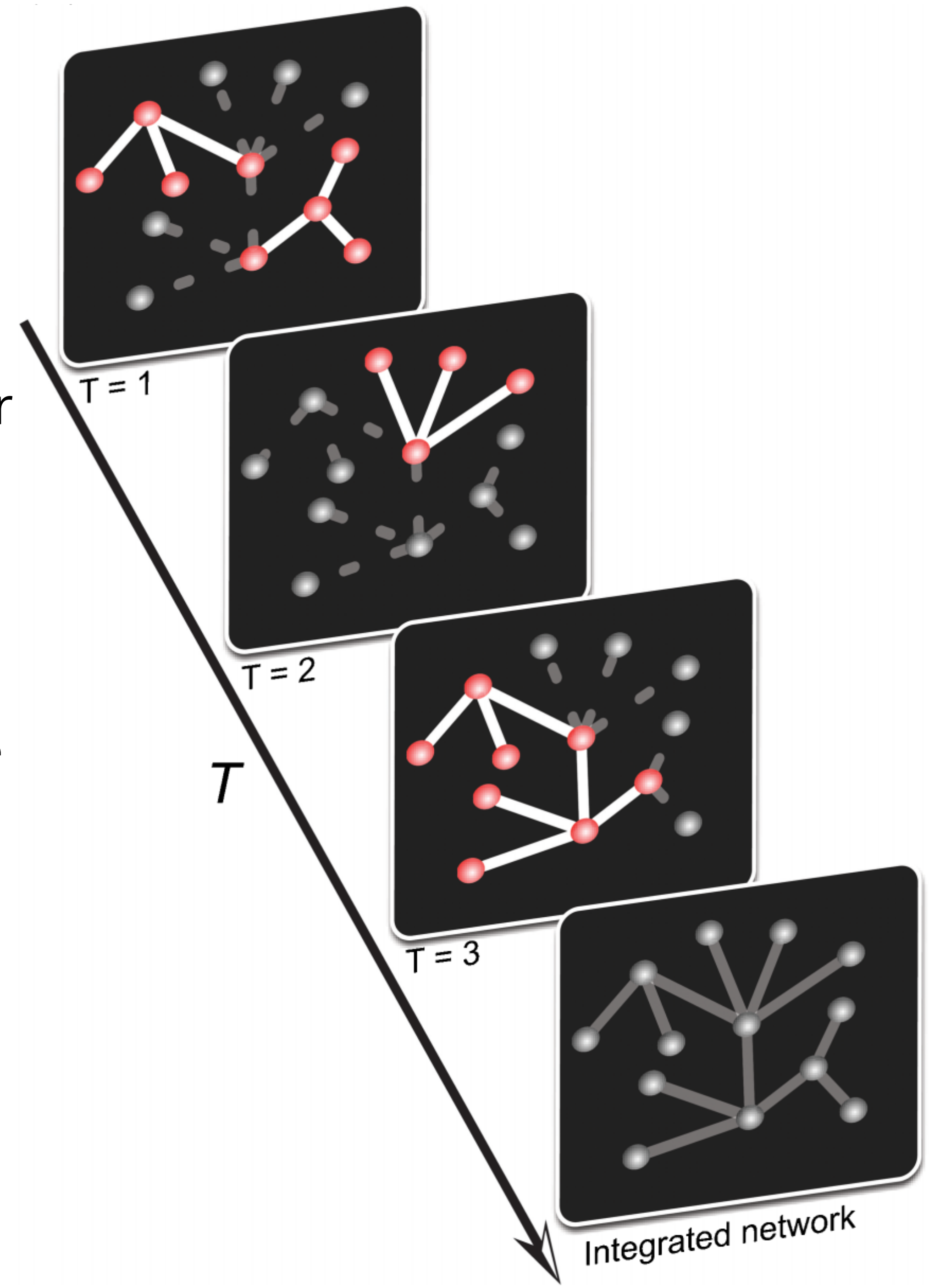
activity driven model

Model ingredients:

- discrete time (time step Δt)
- N: number of nodes
- x_i: activity potential, $\epsilon \leq x_i \leq 1$. This is e.g. number of activation of i during Δt normalised over the total number of activation.
- $F(x)$: distribution of activity potential
- $a_i = \eta x_i$: activation rate. Average number of active nodes $\tilde{N} = \eta \langle x \rangle N$
- m: number of connections made at each activation

At each time steps:

- number of edges $E_t = m\,\eta \langle x \rangle N$
- average degree $\langle k \rangle_t = \dfrac{2E_t}{N} = m\,\eta \langle x \rangle$
- <u>The network is homogeneous!</u>



[Perra et al, Sci Rep 2012]

activity driven model

Integrated network over a time window T

- degree of a node i in the aggregated network $k_T(i) = k_T^{\text{out}}(i) + k_T^{\text{in}}(i)$
- $k_T^{\text{out}}(i)$: i makes $Ta_i m$ links. How many different nodes does it connect to? (I don't count repeated links with the same node). Urn problem: # of different ball extracted from a urn with N balls after $Ta_i m$ extractions
  - prob each ball is extracted $p = 1 - \left[1 - \frac{1}{N}\right]^{Ta_i m}$
  - prob of extracting d balls is a Binomial $P(d) = \binom{N}{d} p^d (1-p)^{N-d}$
  - average # balls $k_T^{\text{out}} = Np = N\left[1 - e^{-Ta_i m/N}\right]$, if $N \to \infty$ and $\frac{T}{N} \to 0$

[Perra et al, Sci Rep 2012]

activity driven model

Integrated network over a time window T

- degree of a node i in the aggregated network $k_T(i) = k_T^{\text{out}}(i) + k_T^{\text{in}}(i)$
- $k_T^{\text{in}}(i)$: nodes that make connections with i among whose were not target by i (already counted in $k_T^{\text{out}}(i)$)
  - prob a node were not target by i : $\left[1 - \frac{1}{N}\right]^{Ta_i m} = e^{-Ta_i m/N}$
  - average number of links coming form these nodes $mN\langle a \rangle$
  - They connect to i with probability $\frac{1}{N}$
  - $k_T^{\text{in}}(i) = m\langle a \rangle e^{-Ta_i m/N}$

[Perra et al, Sci Rep 2012]

activity driven model

Integrated network over a time window T

- degree of a node i in the aggregated network $k_T(i) = k_T^{\text{out}}(i) + k_T^{\text{in}}(i)$

$$k_T(i) = N\left[1 - e^{-Ta_i m/N}\right] + m\langle a\rangle e^{-Ta_i m/N} \simeq N\left[1 - e^{-Ta_i m/N}\right] = N\left[1 - e^{-T\eta x_i m/N}\right]$$

if $N \to \infty$ and $\frac{T}{N} \to 0$

$$x(k) = -\frac{N}{\eta m T}\ln\left(1 - \frac{k}{N}\right)$$

$$P_T(k)dk \sim F(x)dx \qquad P_T(k) \sim F\left[x(k)\right]\frac{dx(k)}{dk} = \frac{1}{Tm\eta}\frac{1}{1-\frac{k}{N}}F\left[-\frac{N}{\eta m T}\ln\left(1-\frac{k}{N}\right)\right]$$

$\frac{k}{N} \to 0$, valid if $T \to 0$

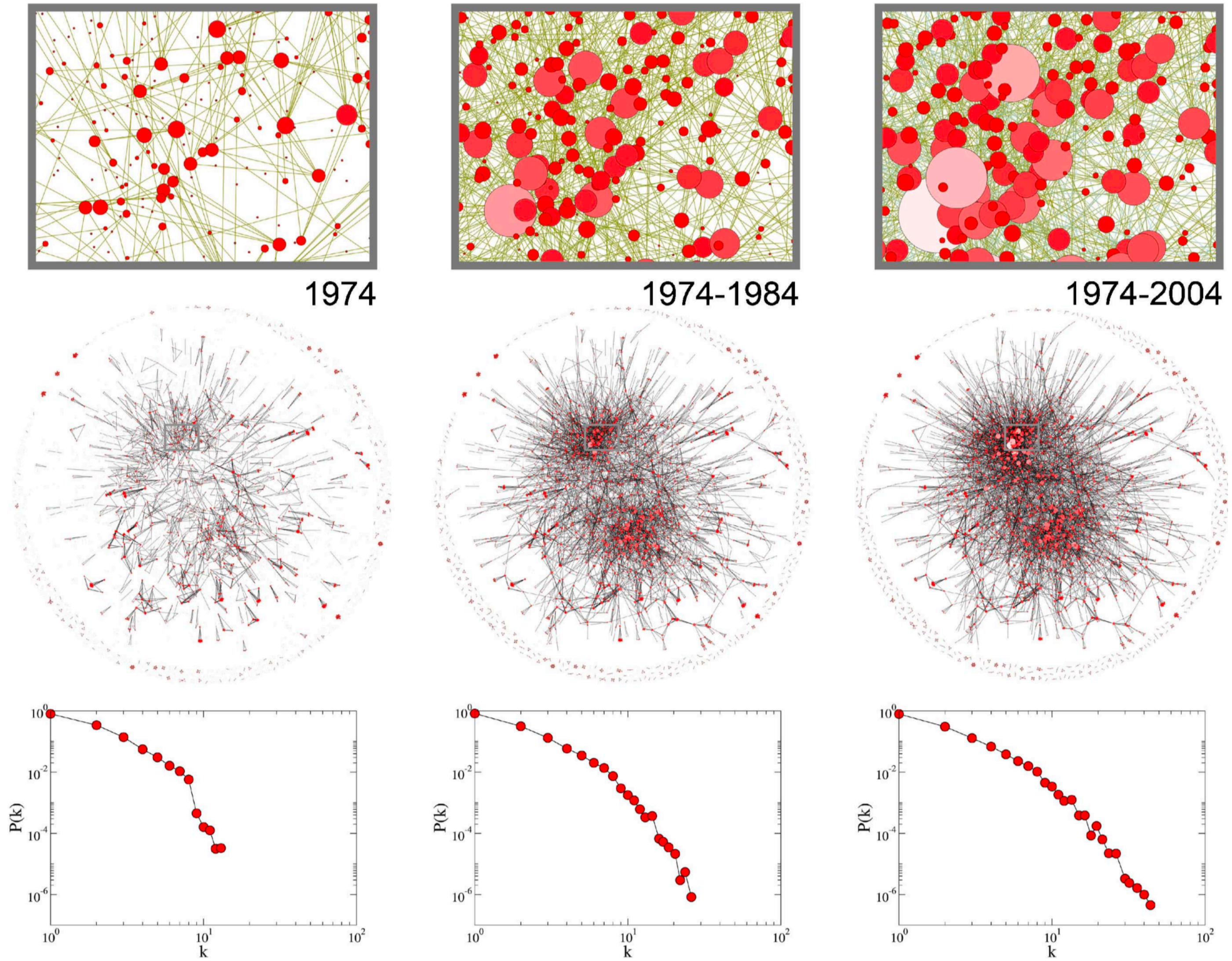
$$P_T(k) \sim \frac{1}{Tm\eta}F\left[\frac{k}{Tm\eta}\right]$$

[Perra et al, Sci Rep 2012]

activity driven model

$$P_T(k) \sim \frac{1}{Tm\eta} F\left[\frac{k}{Tm\eta}\right]$$

heterogenous topology in the aggregated network, over a window T, result from a heterogeneous activity potential



[Perra et al, Sci Rep 2012]

activity driven model

Effect of network dynamics on epidemic spreading

- activity block approximation
- SIR dynamics
- probability of transmission per contact λ
- for simplicity let's assume $m = 1$

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

$$\int da I_a^{t+\Delta t} = I^{t+\Delta t} = I^t - \mu \Delta t I^t + \lambda \langle a \rangle I^t \Delta t + \lambda \theta^t \Delta t$$

$$\theta^{t+\Delta t} = \theta^t - \mu \theta^t \Delta t + \lambda \langle a^2 \rangle I^t \Delta t + \lambda \langle a \rangle \theta^t \Delta t$$

$$\begin{aligned} \partial_t I &= -\mu I + \lambda \langle a \rangle I + \lambda \theta, \\ \partial_t \theta &= -\mu \theta + \lambda \langle a^2 \rangle I + \lambda \langle a \rangle \theta \end{aligned}$$

[Perra et al, Sci Rep 2012]

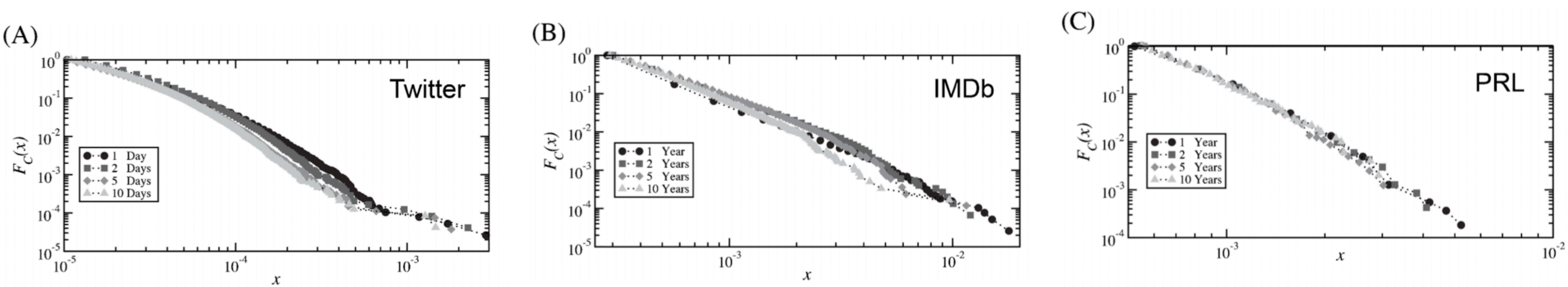
activity driven model

$$\partial_t I = -\mu I + \lambda \langle a \rangle I + \lambda \theta,$$
$$\partial_t \theta = -\mu \theta + \lambda \langle a^2 \rangle I + \lambda \langle a \rangle \theta$$

$$J = \begin{pmatrix} -\mu + \lambda\langle a \rangle & \lambda \\ \lambda\langle a^2 \rangle & -\mu + \lambda\langle a \rangle \end{pmatrix} \qquad \Lambda_{(1,2)} = \lambda\langle a \rangle - \mu \pm \lambda\sqrt{\langle a^2 \rangle}$$

$$\frac{\lambda}{\mu} > \frac{1}{\langle a \rangle + \sqrt{\langle a^2 \rangle}} + \mathcal{O}\left(\frac{1}{N}\right)$$

heterogeneities in the activation rate lower the epidemic threshold



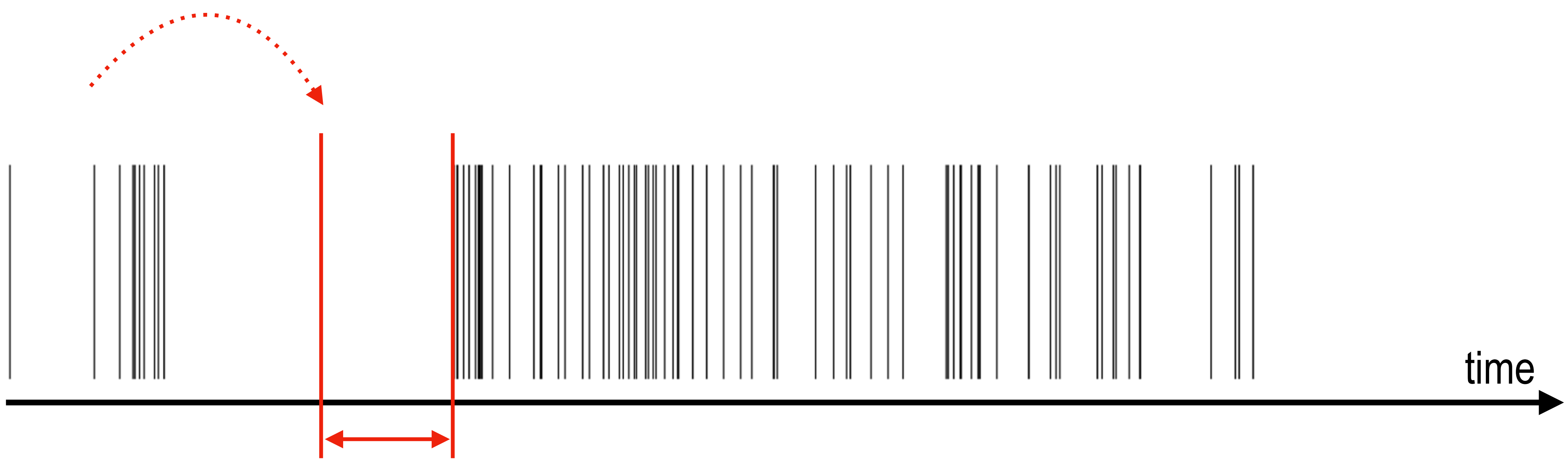
[Perra et al, Sci Rep 2012]

activity driven model

- A model that captures a realistic property of human behaviour (face-to-face, sexual contacts, phone call, email, tweets)
- human have heterogeneous activity rate
- the contact network at a certain instant of time is sparse, with homogeneous degree
- the aggregated network over a certain window is well connected with heterogeneous degree
- pattern of activation unfolds at the same time scale of the spreading process
- calculation possible in the activity-block approximation (same scheme as the degree block approximation)
- contact heterogeneities lower the epidemic threshold

bustiness & spreading

spread of computer viruses I receive an email. As soon as I open it the email is automatically sent to my contacts



- inter-contact time $P(\tau)$
- time from arrival of the email to the moment in which I open it: residual waiting time

$$g(\tau) = \frac{1}{\langle \tau \rangle} \int_{\tau}^{\infty} dx P(x)$$

[Vazquez et al PRL 2007]

bustiness & spreading

Spread of computer viruses: I receive an email. As soon as I open it the email is automatically sent to my contacts

- inter-contact time $P(\tau)$
- time from arrival of the email to the moment in which I open it: residual waiting time

$$g(\tau) = \frac{1}{\langle\tau\rangle}\int_{\tau}^{\infty} dx P(x)$$

- number of infected users in time: $n(t) = \sum_{d=1}^{D} z_d g^{*d}(t)$
    - $g^{*d}(t)$: g order convolution of $g(\tau)$, $g^{*d}(t) = \int_0^t d\tau g(\tau) g^{*d}(t-\tau)$ for $d > 1$
    - z^d average # of users d email contacts away from the first user

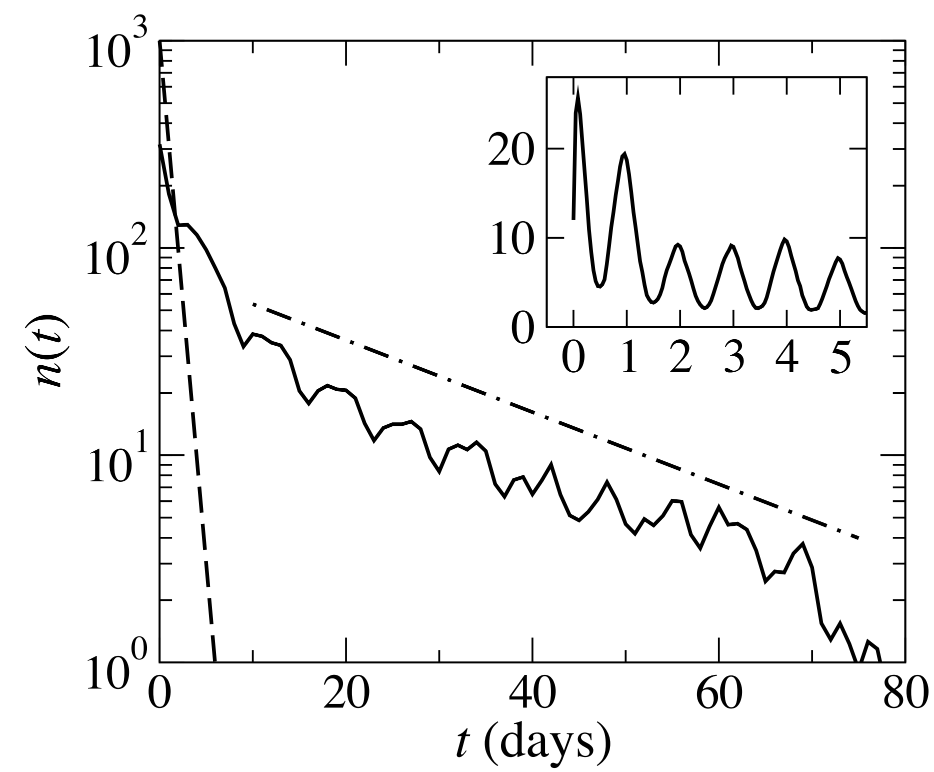
[Vazquez et al PRL 2007]

burstiness & spreading

- number of infected users in time: $n(t) = \sum_{d=1}^{D} z_d g^{*d}(t)$
    - $g^{*d}(t)$: g order convolution of $g(\tau)$, $g^{*d}(t) = \int_0^t d\tau g(\tau) g^{*d}(t-\tau)$ for $d > 1$
    - z^d average # of users d email contacts away from the first user
- $n(t) = F(t)\exp\left(-\frac{t}{\tau_0}\right)$; Poisson distribution $\tau_0 = \langle\tau\rangle$; Power low distribution $\tau_0 = \tau_E$

$\tau_E \gg \langle\tau\rangle \Rightarrow$ long time decay in incidence

burstiness slow down spreading



approaches to temporal network epidemiology

Botom-up: generative models

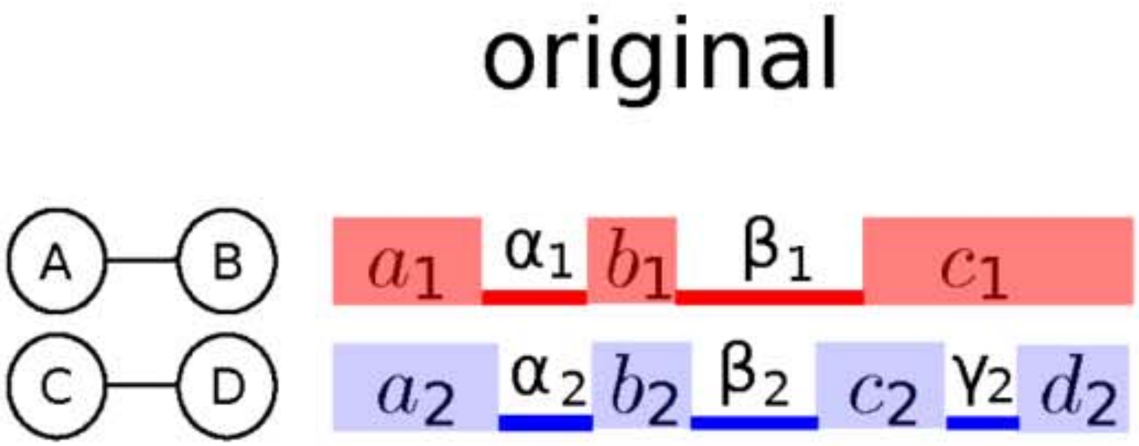
activity driven model, and its extensions

Top-down approaches: Randomised Reference Models

compare the epidemics on real data with the outcome in suitable null models

Randomised reference models (RRM)

- $P(\tau)$: inter-contact time distribution
- ω_{AB}: cumulated contact durations of an arbitrary link
- $P(\omega)$: distribution of the cumulated contacts duration
- n_{AB}: number of contacts per link of an arbitrary link
- $P(n)$: distribution of the number of contacts per link



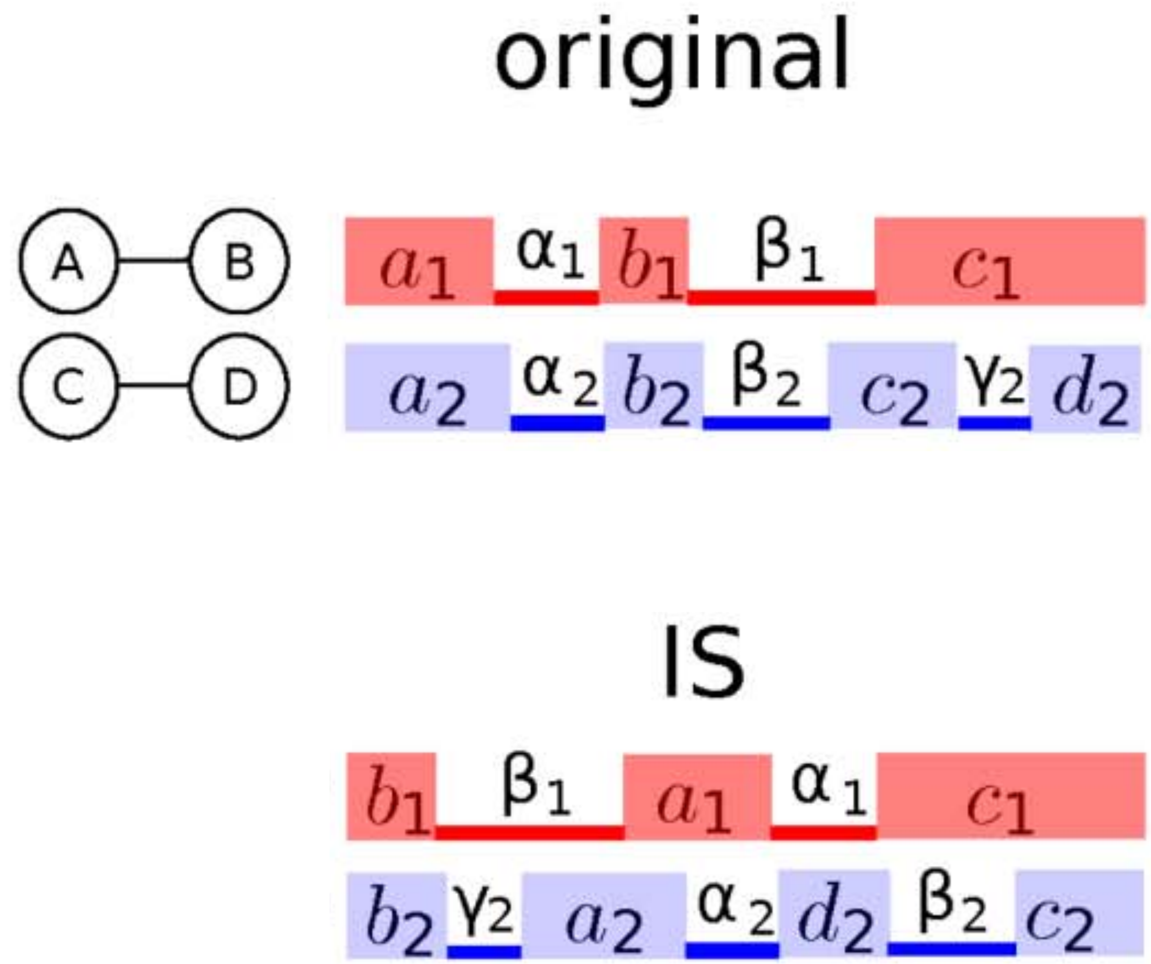
RRM	Topology	Causality	$P(\tau)$	ω_{AB}	$P(\omega)$	n_{AB}	$P(n)$

[Gauvin et al Sci Rep 2013]

Randomised reference models (RRM)

- $P(\tau)$: inter-contact time distribution
- ω_{AB}: cumulated contact durations of an arbitrary link
- $P(\omega)$: distribution of the cumulated contacts duration
- n_{AB}: number of contacts per link of an arbitrary link
- $P(n)$: distribution of the number of contacts per link

RRM	Topology	Causality	$P(\tau)$	ω_{AB}	$P(\omega)$	n_{AB}	$P(n)$
IS							

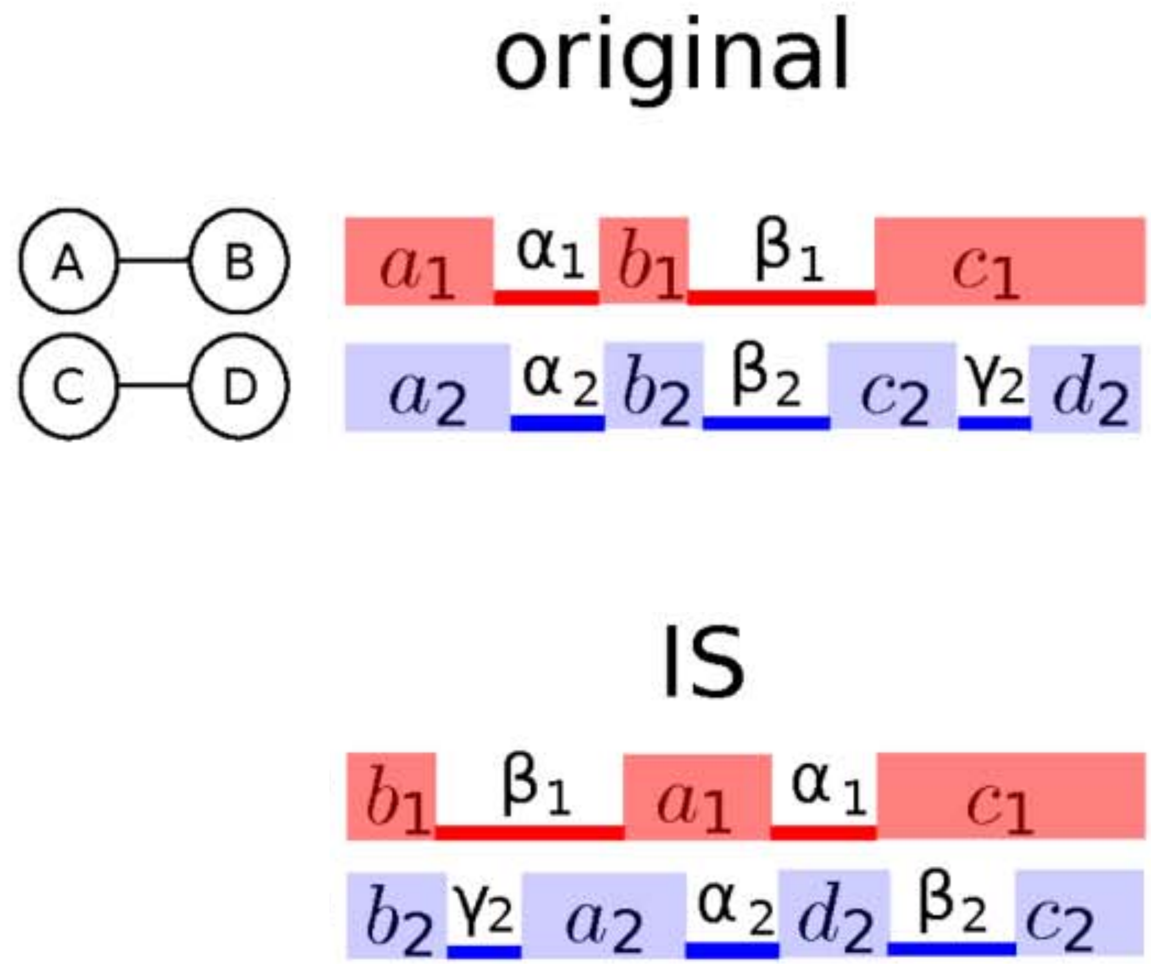


interval shuffling (IS): the sequences of contact and inter-contact durations are reshuffled for each link separately

[Gauvin et al Sci Rep 2013]

Randomised reference models (RRM)

- $P(\tau)$: inter-contact time distribution
- ω_{AB}: cumulated contact durations of an arbitrary link
- $P(\omega)$: distribution of the cumulated contacts duration
- n_{AB}: number of contacts per link of an arbitrary link
- $P(n)$: distribution of the number of contacts per link



RRM	Topology	Causality	$P(\tau)$	ω_{AB}	$P(\omega)$	n_{AB}	$P(n)$
IS	V	X	V	V	V	V	V

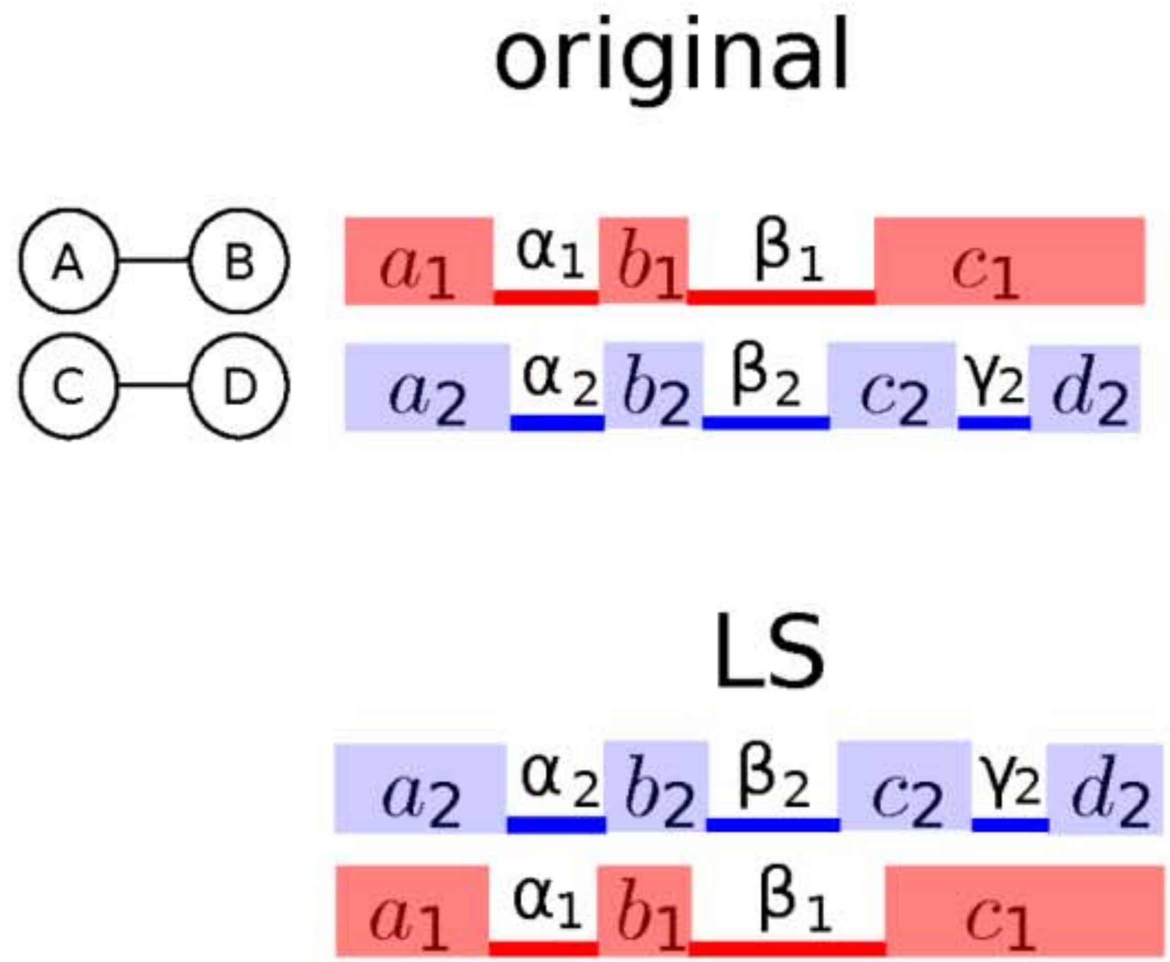
interval shuffling (IS): the sequences of contact and inter-contact durations are reshuffled for each link separately

[Gauvin et al Sci Rep 2013]

Randomised reference models (RRM)

- $P(\tau)$: inter-contact time distribution
- ω_{AB}: cumulated contact durations of an arbitrary link
- $P(\omega)$: distribution of the cumulated contacts duration
- n_{AB}: number of contacts per link of an arbitrary link
- $P(n)$: distribution of the number of contacts per link

RRM	Topology	Causality	$P(\tau)$	ω_{AB}	$P(\omega)$	n_{AB}	$P(n)$
IS	V	X	V	V	V	V	V
LS							



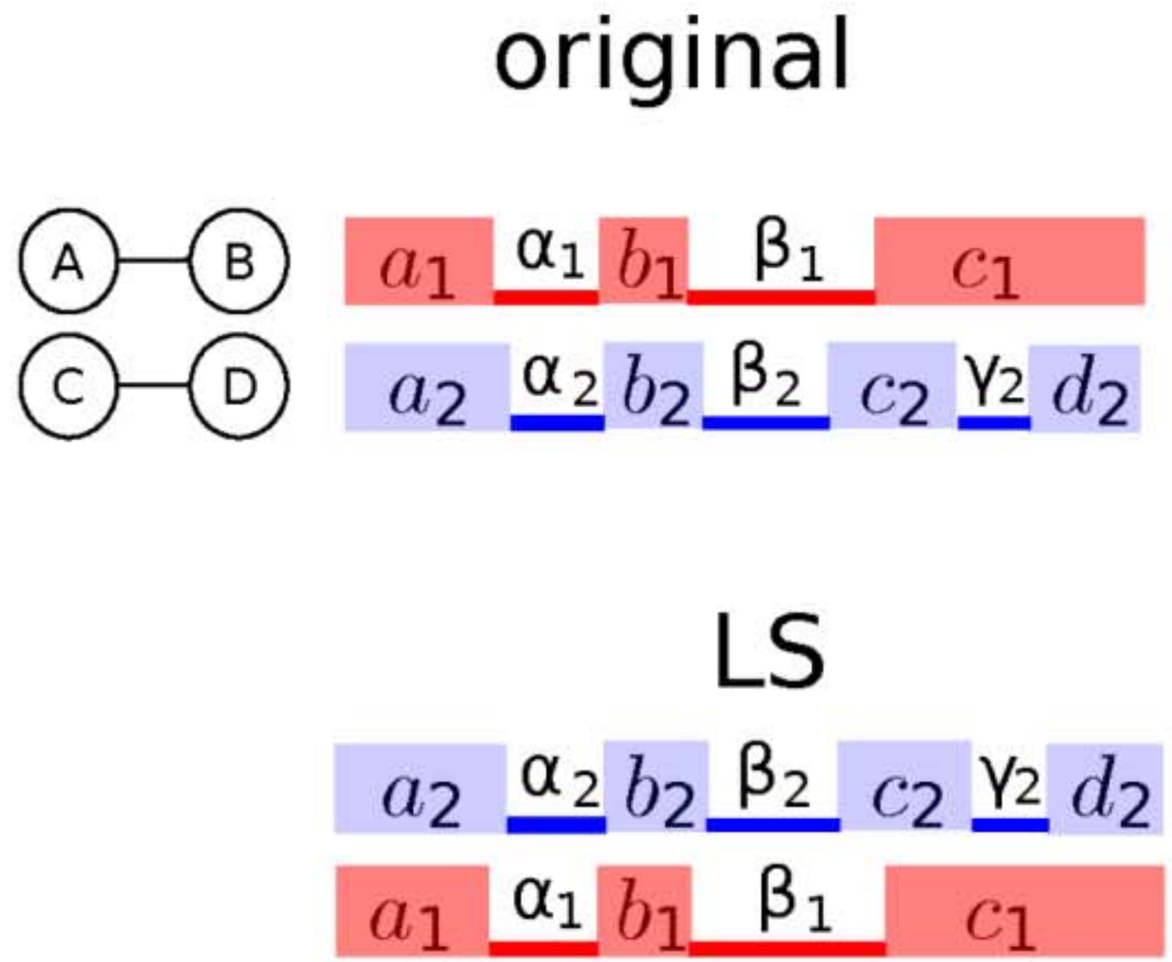
link shuffling (LS): the unaltered sequences of events are swapped between link pairs

[Gauvin et al Sci Rep 2013]

Randomised reference models (RRM)

- $P(\tau)$: inter-contact time distribution
- ω_{AB}: cumulated contact durations of an arbitrary link
- $P(\omega)$: distribution of the cumulated contacts duration
- n_{AB}: number of contacts per link of an arbitrary link
- $P(n)$: distribution of the number of contacts per link

RRM	Topology	Causality	$P(\tau)$	ω_{AB}	$P(\omega)$	n_{AB}	$P(n)$
IS	V	X	V	V	V	V	V
LS	V	X	V	X	V	X	V

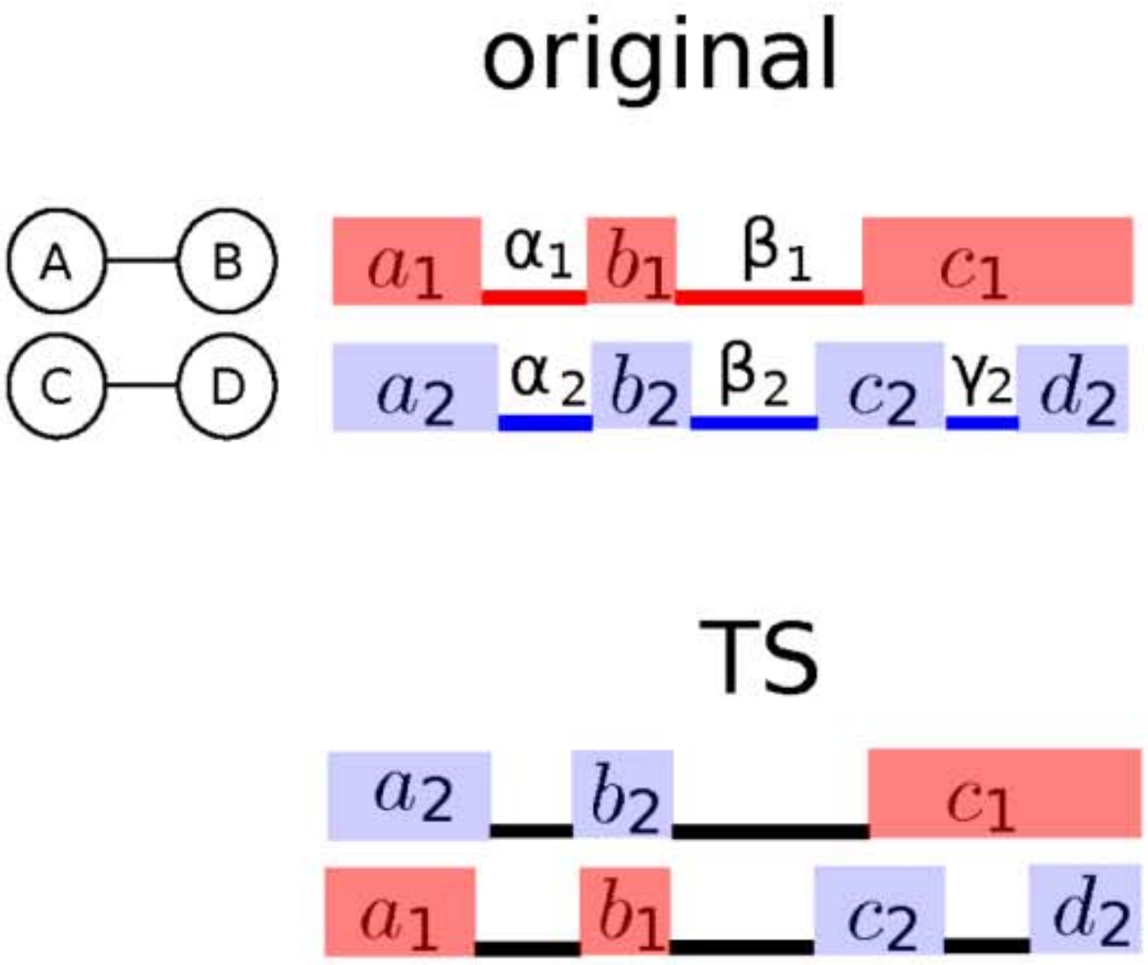


link shuffling (LS): the unaltered sequences of events are swapped between link pairs

[Gauvin et al Sci Rep 2013]

Randomised reference models (RRM)

- $P(\tau)$: inter-contact time distribution
- ω_{AB}: cumulated contact durations of an arbitrary link
- $P(\omega)$: distribution of the cumulated contacts duration
- n_{AB}: number of contacts per link of an arbitrary link
- $P(n)$: distribution of the number of contacts per link

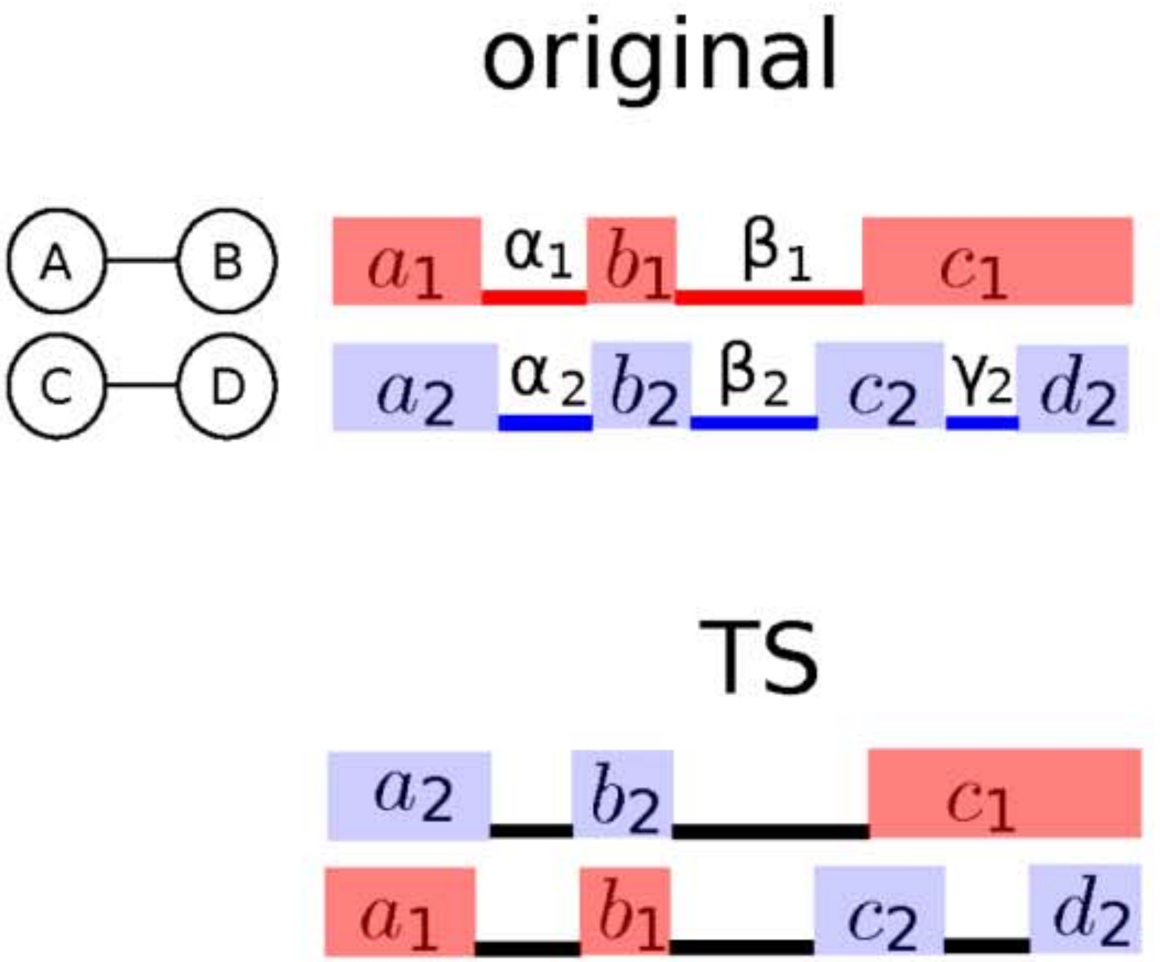


RRM	Topology	Causality	$P(\tau)$	ω_{AB}	$P(\omega)$	n_{AB}	$P(n)$
IS	V	X	V	V	V	V	V
LS	V	X	V	X	V	X	V
TS							

global time shuffling (TS): build a global list of the contact durations. For each link, generate a synthetic activity timeline by sampling with replacement the global list according to the original number of contacts for that link

Randomised reference models (RRM)

- $P(\tau)$: inter-contact time distribution
- ω_{AB}: cumulated contact durations of an arbitrary link
- $P(\omega)$: distribution of the cumulated contacts duration
- n_{AB}: number of contacts per link of an arbitrary link
- $P(n)$: distribution of the number of contacts per link



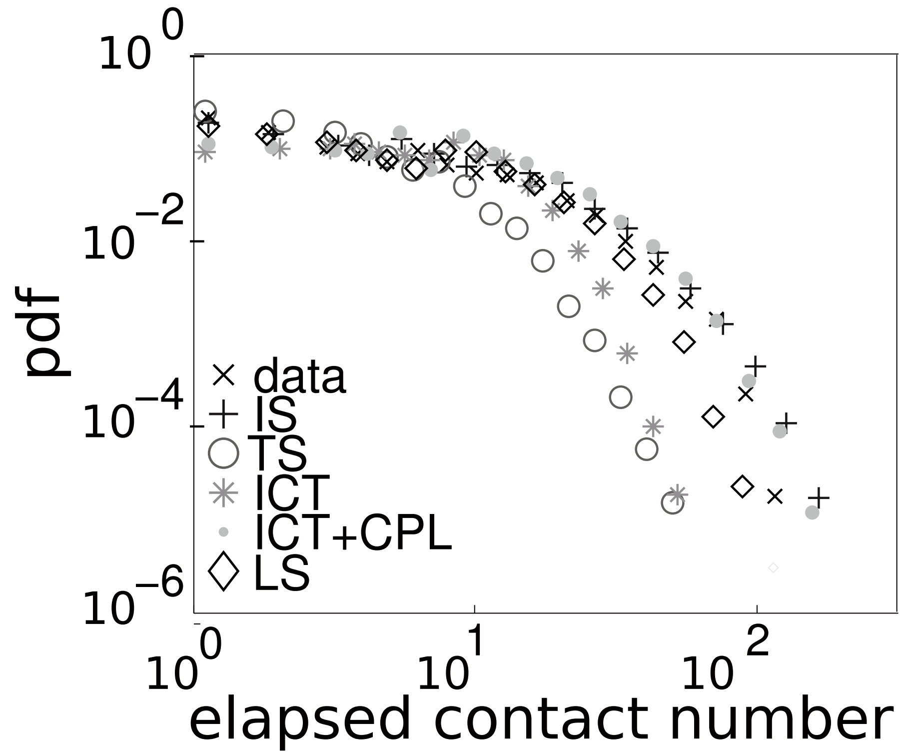
RRM	Topology	Causality	$P(\tau)$	ω_{AB}	$P(\omega)$	n_{AB}	$P(n)$
IS	V	X	V	V	V	V	V
LS	V	X	V	X	V	X	V
TS	V	X	X	X	X	V	V

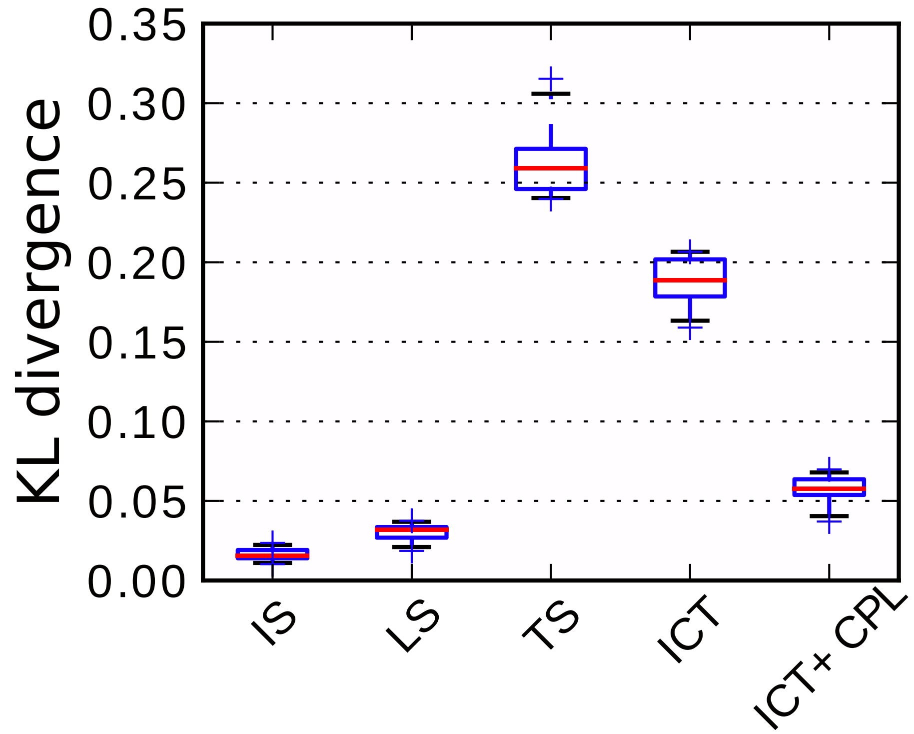
global time shuffling (TS): build a global list of the contact durations. For each link, generate a synthetic activity timeline by sampling with replacement the global list according to the original number of contacts for that link

Randomised reference models (RRM)

- $P(\tau)$: inter-contact time distribution
- ω_{AB}: cumulated contact durations of an arbitrary link
- $P(\omega)$: distribution of the cumulated contacts duration
- n_{AB}: number of contacts per link of an arbitrary link
- $P(n)$: distribution of the number of contacts per link

RRM	Topology	Causality	$P(\tau)$	ω_{AB}	$P(\omega)$	n_{AB}	$P(n)$
IS	V	X	V	V	V	V	V
LS	V	X	V	X	V	X	V
TS	V	X	X	X	X	V	V





[Gauvin et al Sci Rep 2013]

Randomised reference models (RRM)

[paper of Christian]