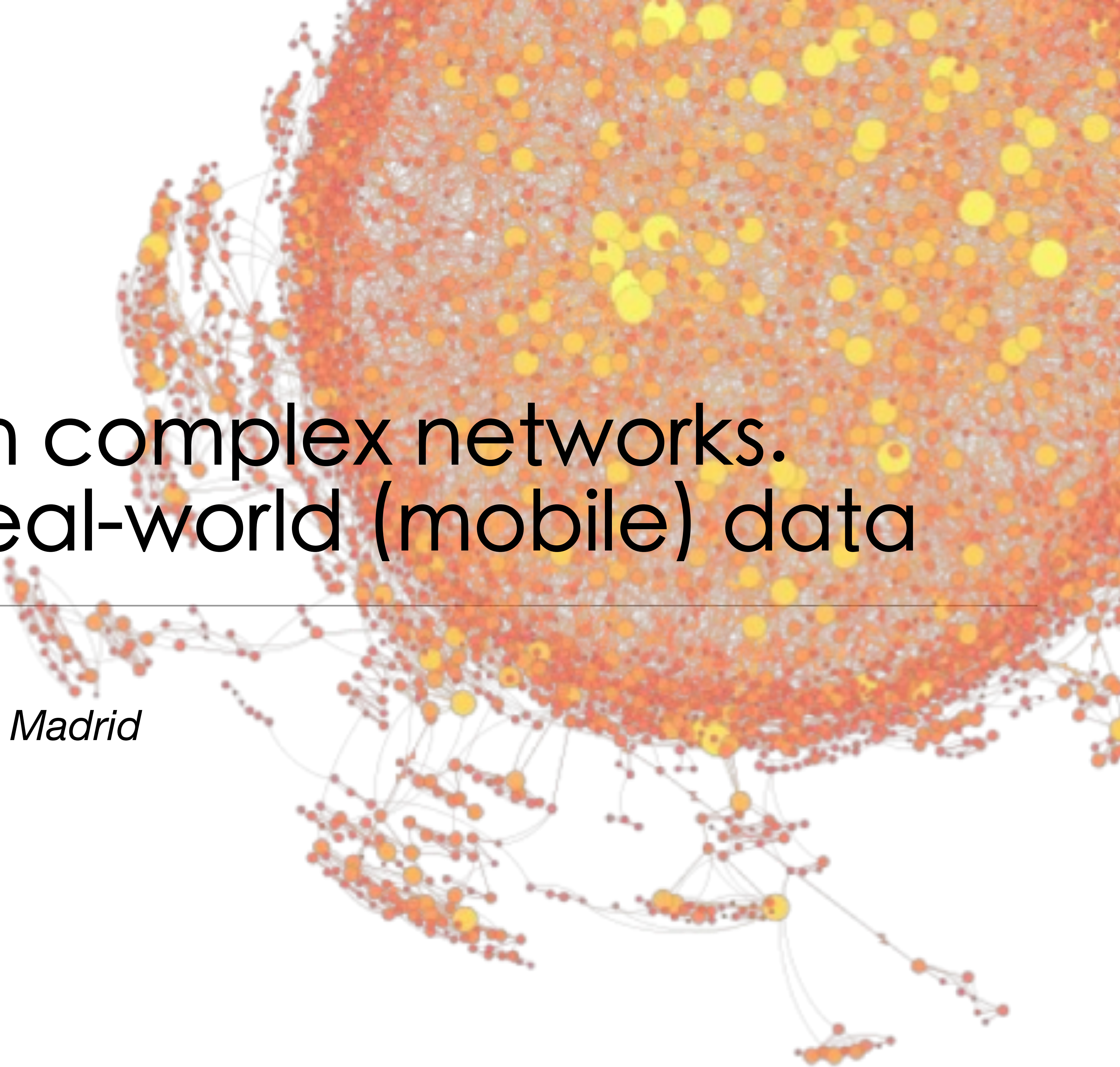
Dynamics in complex networks. Analysing real-world (mobile) data



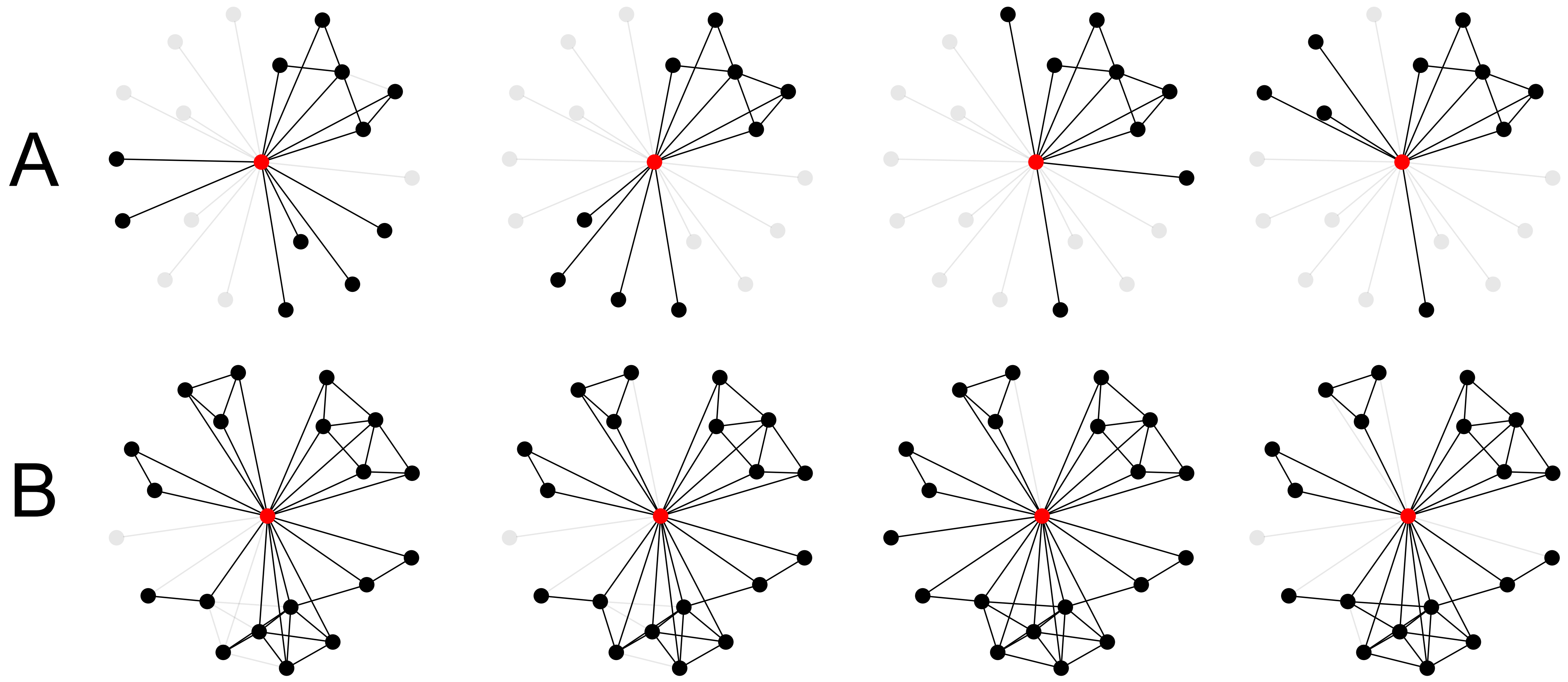
Esteban Moro
Universidad Carlos III de Madrid

Universidad Carlos III de Madrid

DPCN 2016, Wrocław

Summary

1. Motivation
2. Social dynamical processes
    2.1. Individual dynamics
    2.2. Tie dynamics
        2.2.1. Tie interaction activity
        2.2.2. Tie formation/decay dynamics
    2.3. Geographical dynamics
3. Impact on information diffusion
4. Applications to real-world problems

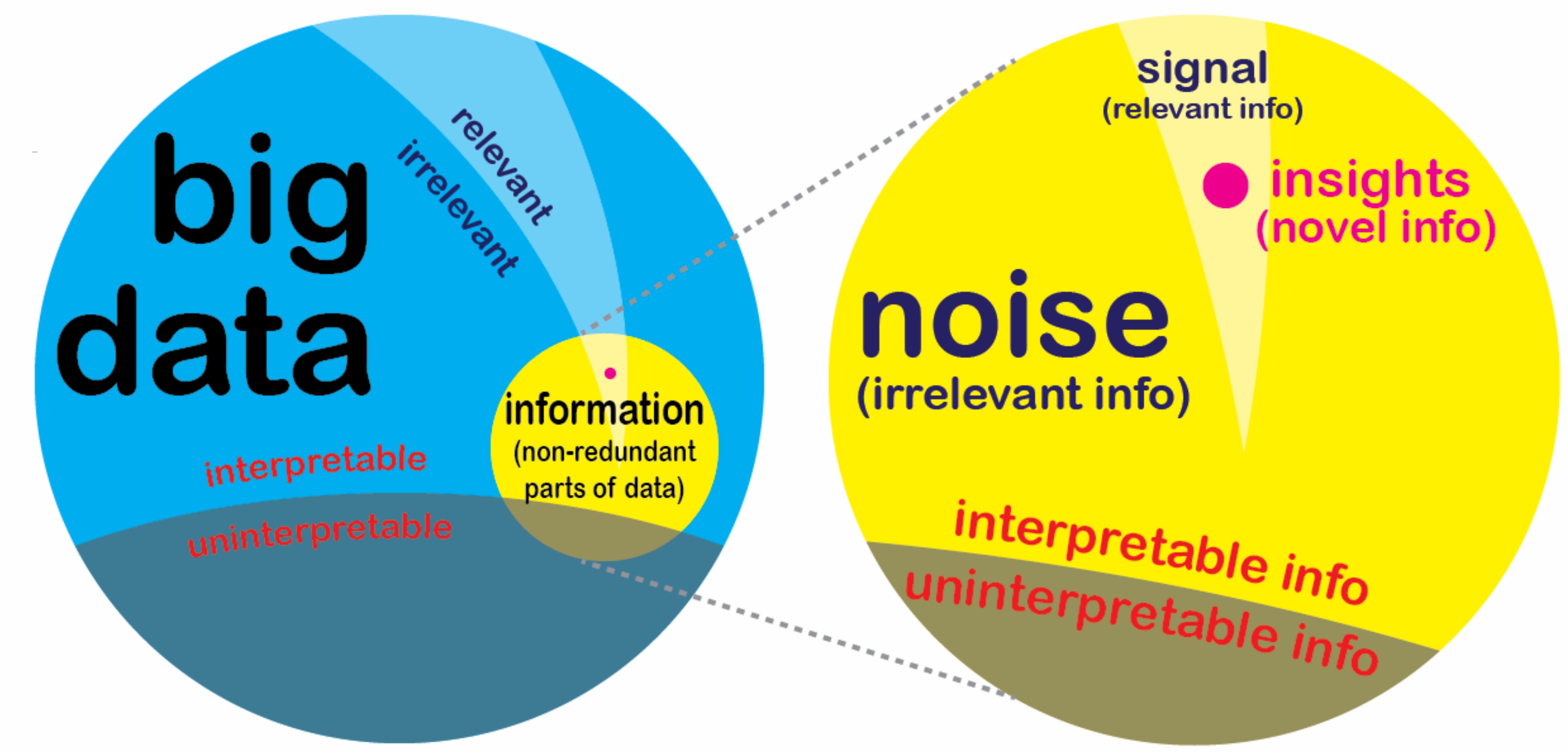


Motivation 1

Motivation

stebanmoro

Motivation


big
data
relevant
irrelevant
information
(non-redundant
parts of data)
interpretable
uninterpretable
signal
(relevant info)
insights
(novel info)
noise
(irrelevant info)
interpretable info
uninterpretable info

Motivation

You are what *you repeatedly do [Aristóteles]*

Using BigData to infer behavior or society situation

Situation
- Demographics
- Health
- Economy
- Unemployment
- Transportation
- Geography
- Politics

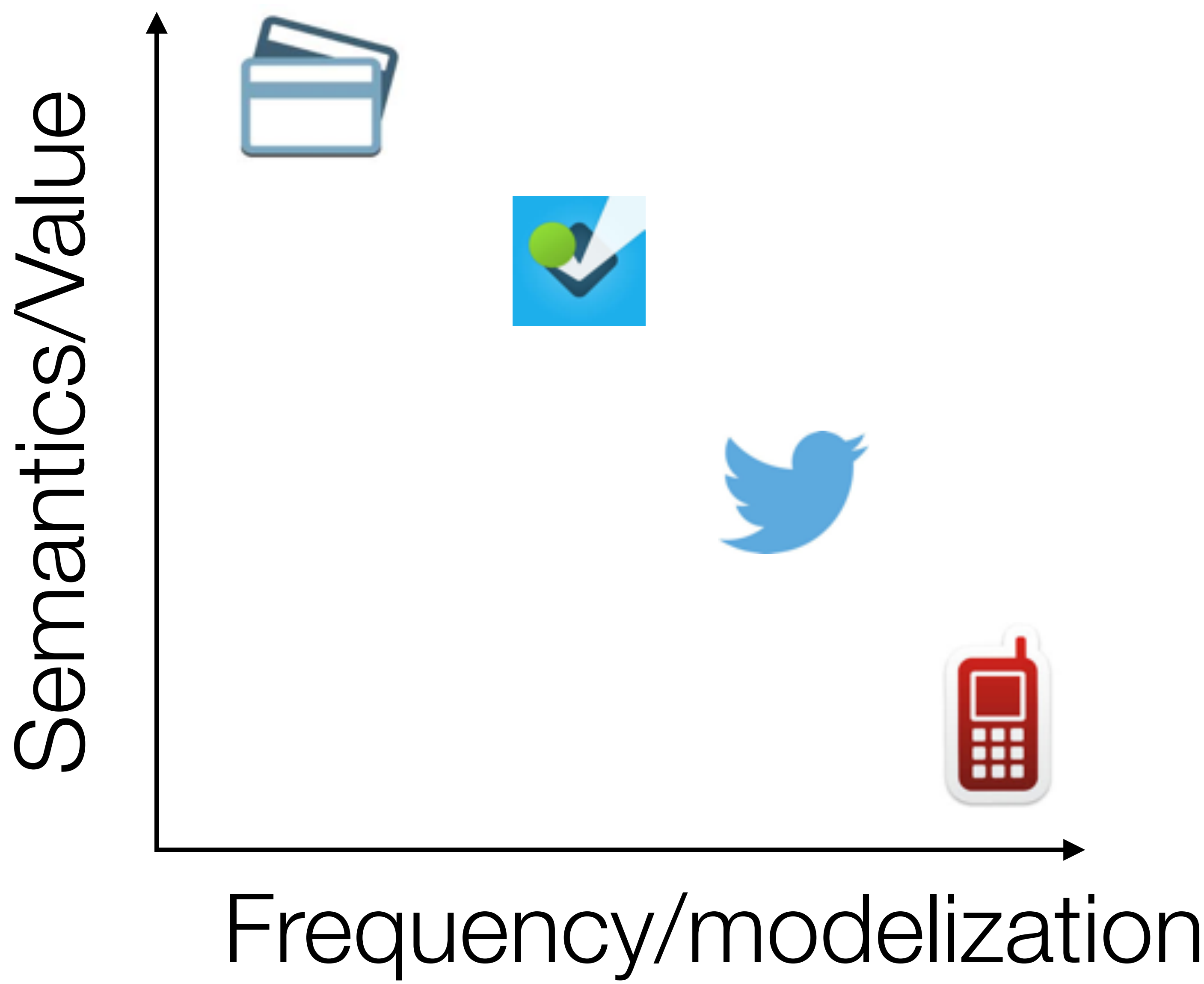
Behavior
- Social
- Mobility
- Activity
- Content

Observation
- Surveys
- Credit card
- Mobile phone
- Social media
- Searches
- …

Individual - Group - City

Motivation

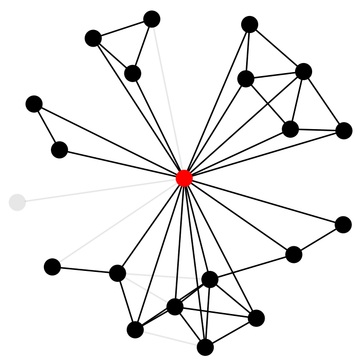
Sources of BigData



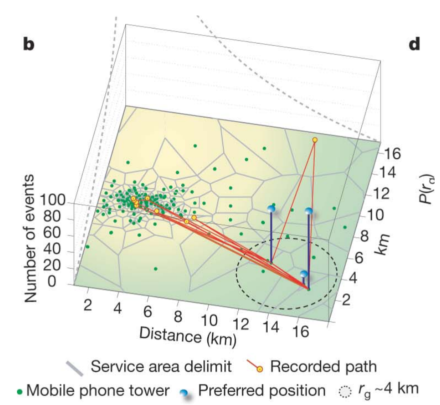
Motivation

Sources of BigData

netmob.org



Social interactions



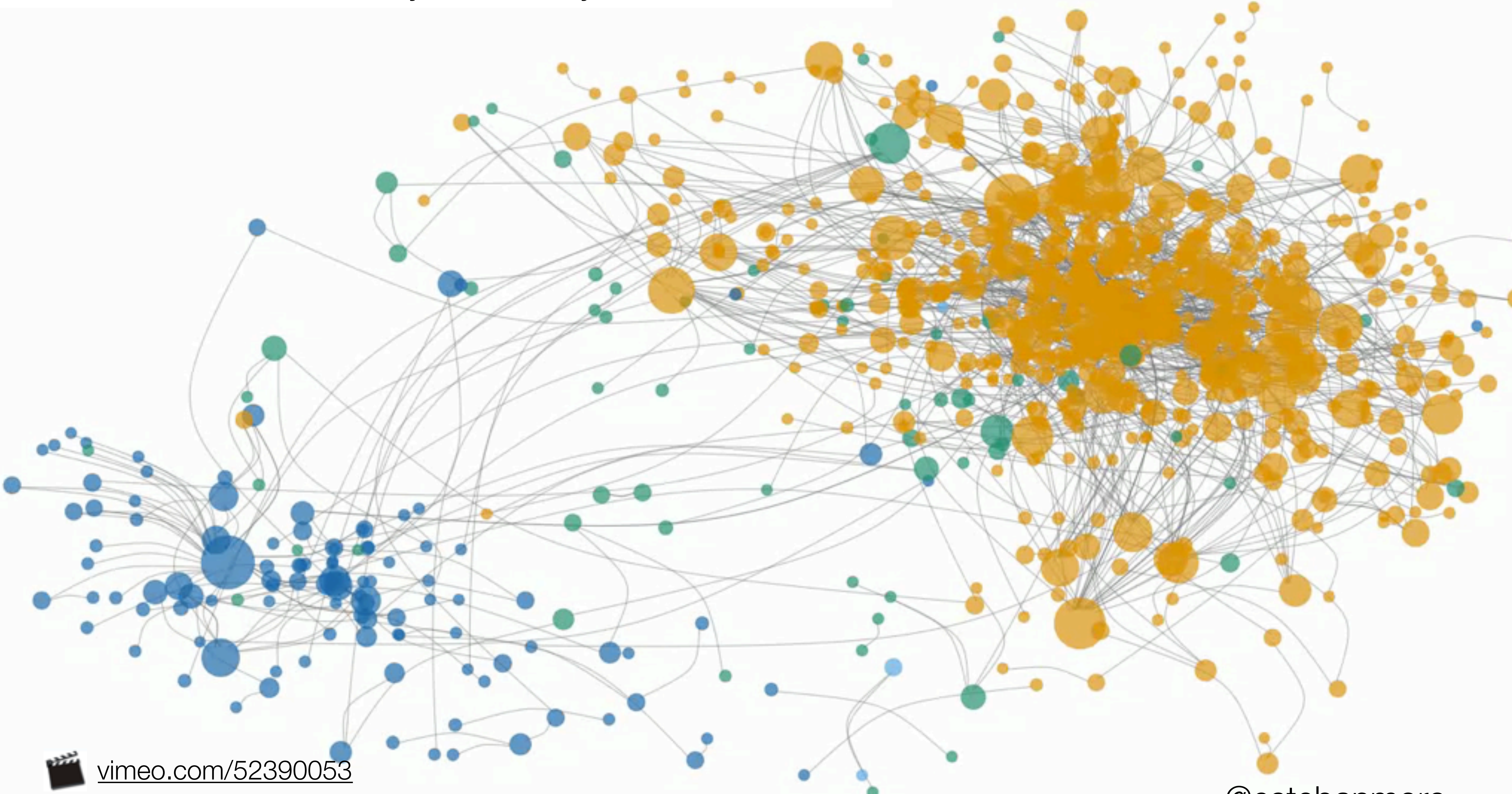
Mobility

Transport
Marketing
Development
Food Consumption
Economy
Energy
Retail Analytics
Unemployment
Smart Cities
Geography

- Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. EPJ Data Science, 4(1), 10. http://doi.org/10.1140/epjds/s13688-015-0046-0
- MOBILE PHONE NETWORK DATA FOR DEVELOPMENT. (2013). UN Global Pulse
- Saramaki, J., & Moro, E. (2015). From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. The European Physical Journal B, 88(6). http://doi.org/10.1140/epjb/e2015-60106-6
- Naboulsi, D., Fiore, M., Ribot, S., & Stanica, R. (n.d.). Large-scale Mobile Traffic Analysis: a Survey. IEEE Communications Surveys & Tutorials, 1–1. http://doi.org/10.1109/COMST.2015.2491361

Motivation

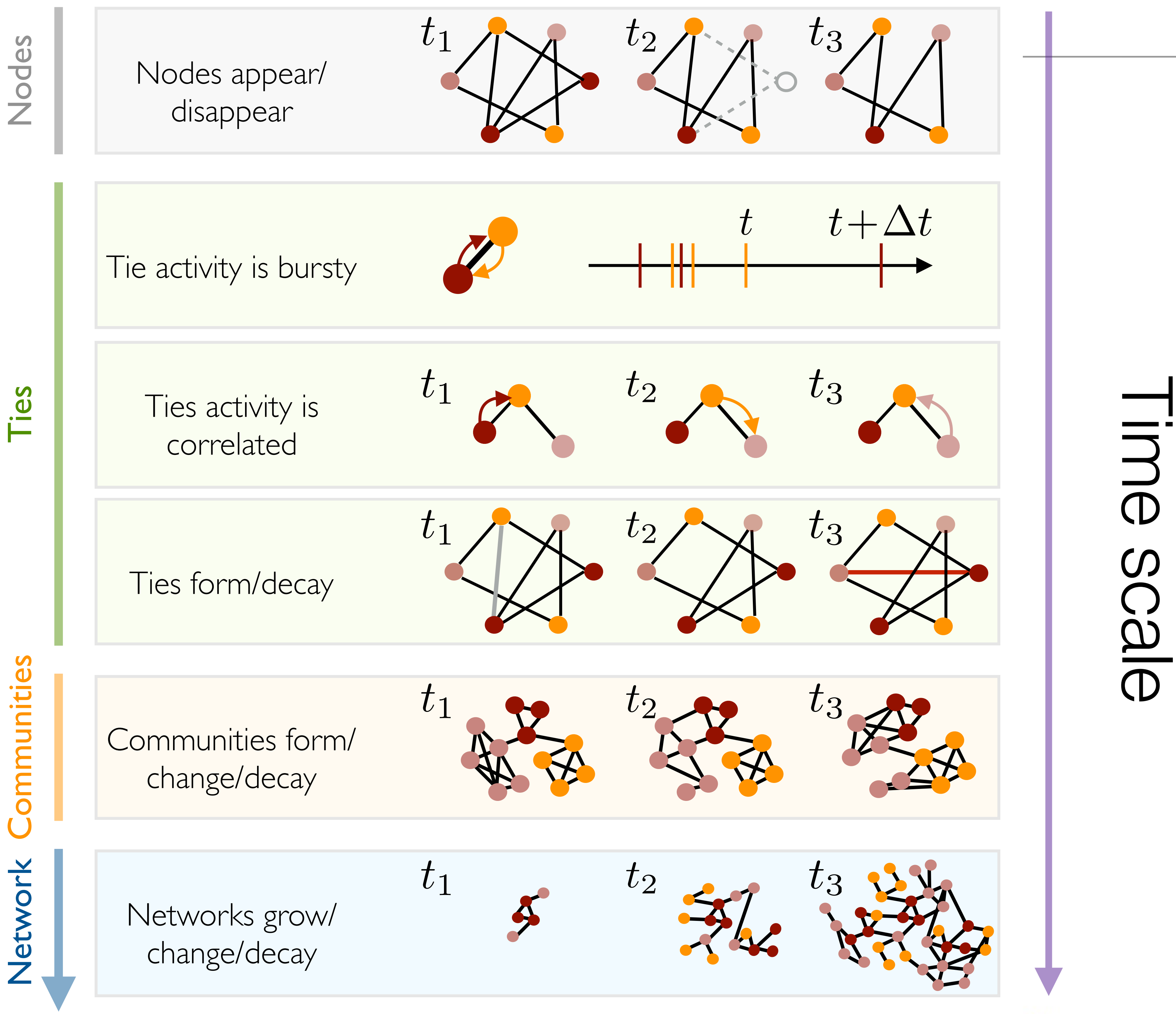
Social networks are dynamical by nature



vimeo.com/52390053

@estebanmoro

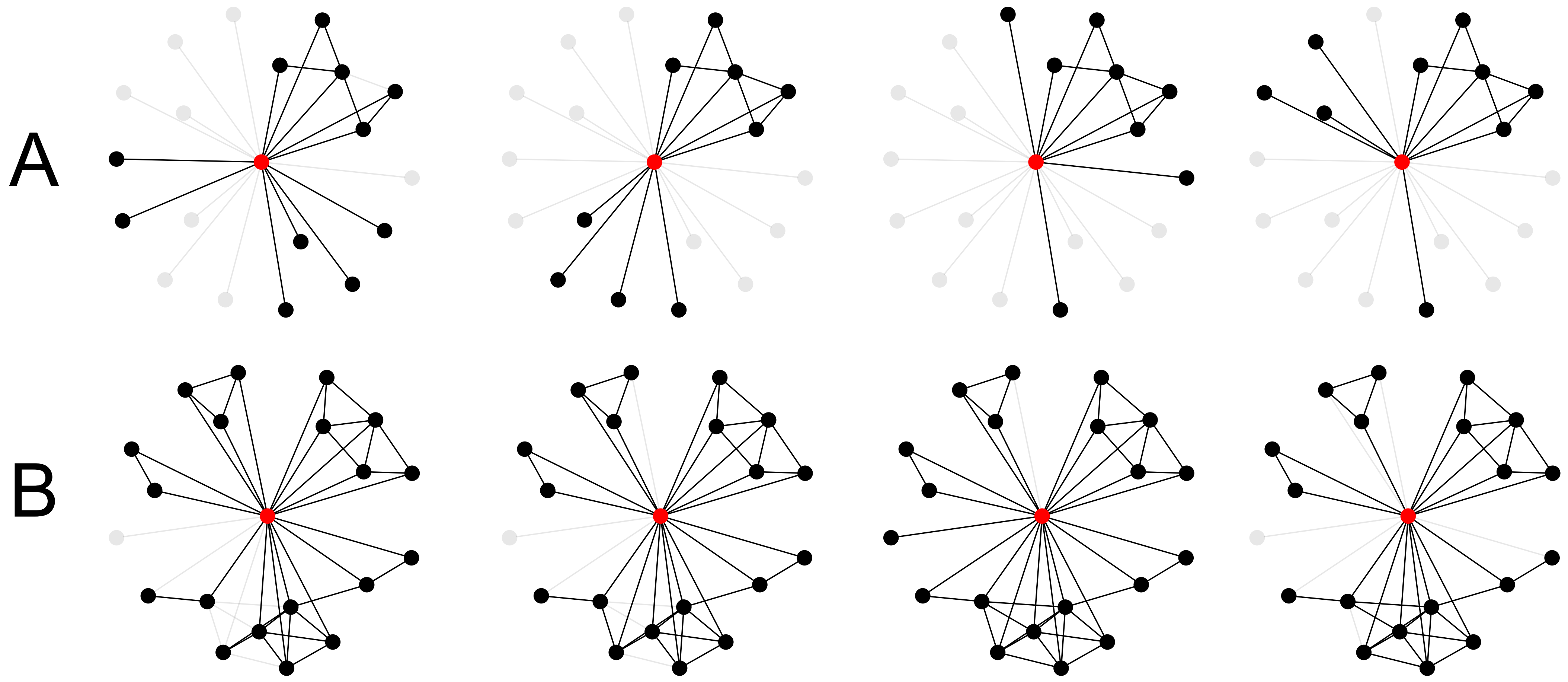
Complex dynamics of real networks



Universidad Carlos III de Madrid

@estebanmoro

i

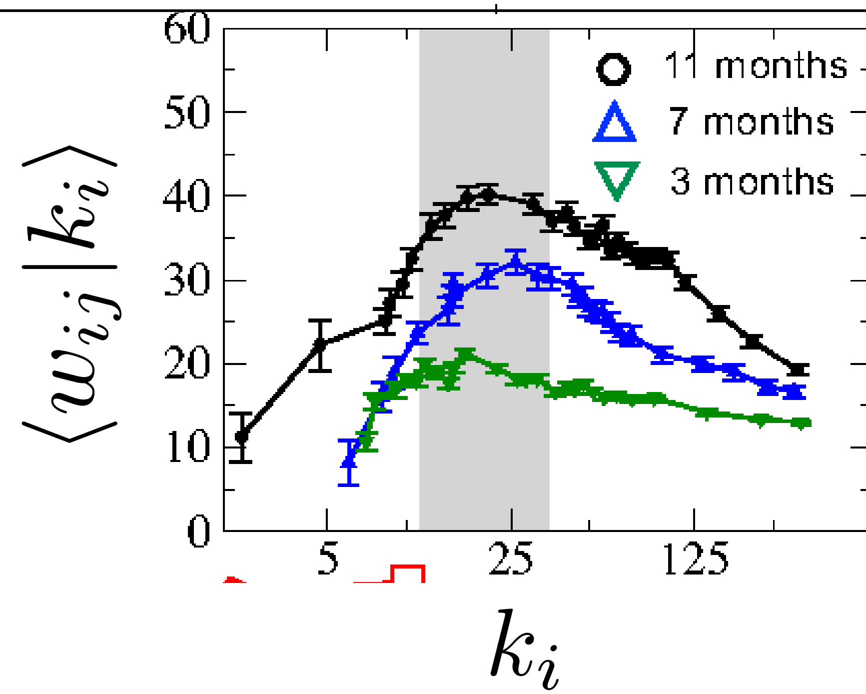


Social dynamical process

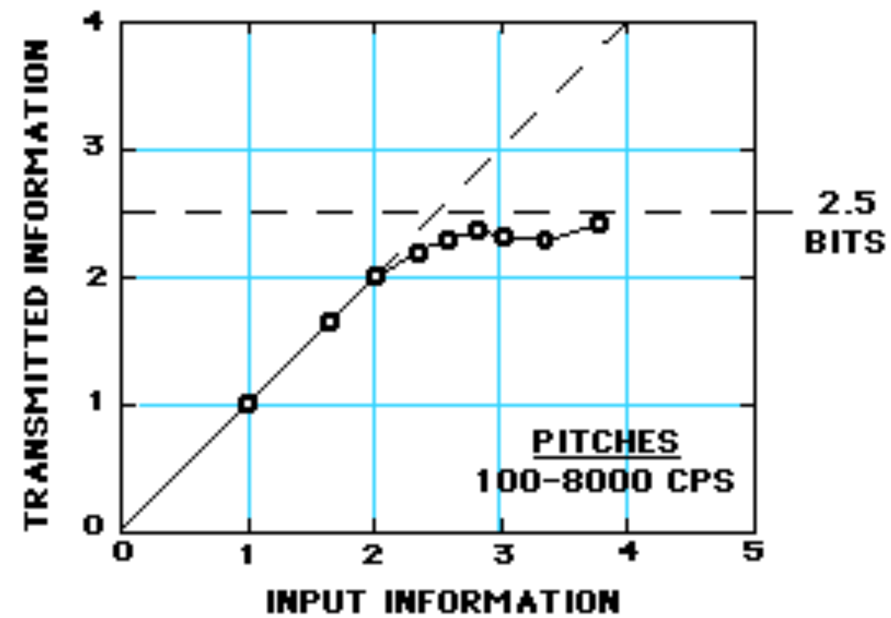
2

Social dynamical process

Miritello, G. et al., 2013. Time as a limited resource: Communication strategy in mobile phone networks. *Social Networks*.

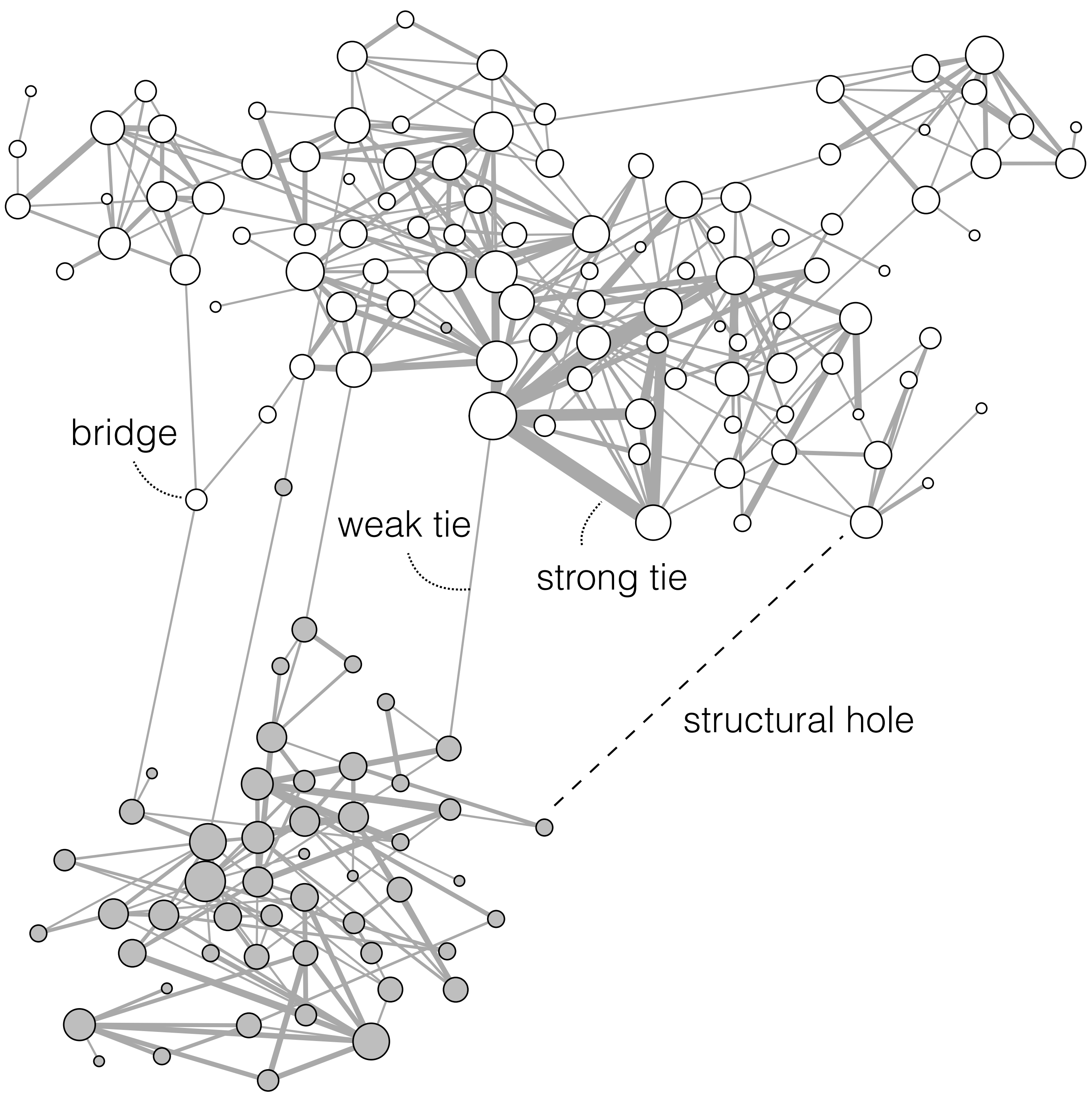


- **Cognitive limits**
  - Dunbar's number
    - There is a cognitive limit to the number of people with whom one can maintain stable social relationships. (Dunbar 1992)
  - The magical number Seven Plus Minus Two
    - The number of objects an average human can hold in working memory is 7 ± 2 (Miller '56)



Social dynamical processes

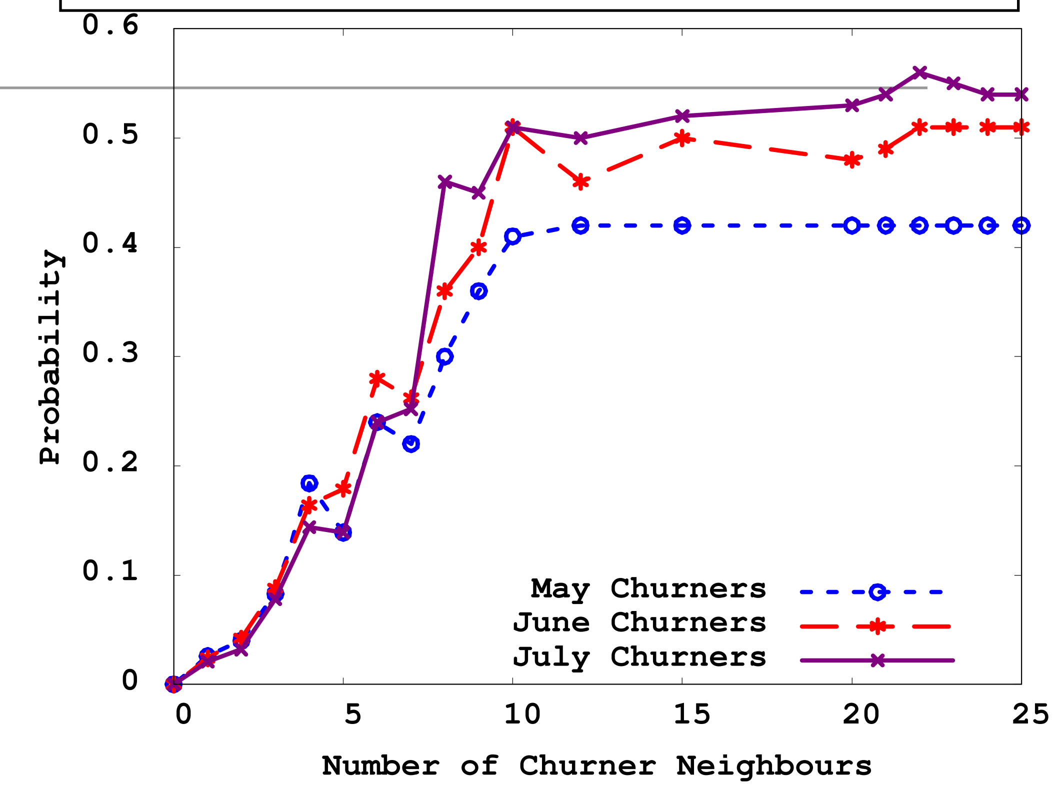
- **Embeddedness / clustering / triadic closure / weak ties**
  - *Embeddedness, clustering:* People who spend time with a third are likely to encounter each other (triadic closure). Minimizes conflict, maximizes trusts,…
  - *Bridges, structural holes* (Burt): Bridges have structural advantages since they have access to non-redundant information
  - *Weak ties* (Granovetter): weak ties tend to connect different areas of the network (they are more likely to be sources of novel information)



Rivera, M.T., Soderstrom, S.B. & Uzzi, B., 2010. Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. *Annual Review of Sociology*, 36(1), pp.91–115.

Social dynamical processes

Dasgupta, K. et al., 2008. Social ties and their relevance to churn in mobile telecom networks.

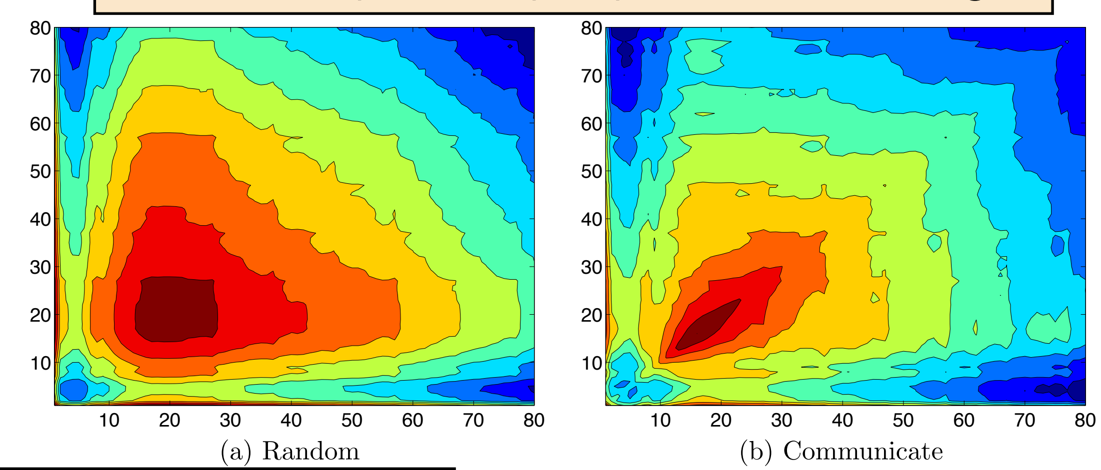


- ***Contagion***
  - Human behaviors spread on the network
  - Dynamics too
- ***Homophily***
  - The greater the similarity between individuals the more likely they are to establish a connection

Attribute	Random	Communicate
Age	-0.0001	0.297
Gender	0.0001	-0.032
ZIP	-0.0003	0.557
County	0.0005	0.704
Language	-0.0001	0.694

Correlation coefficient

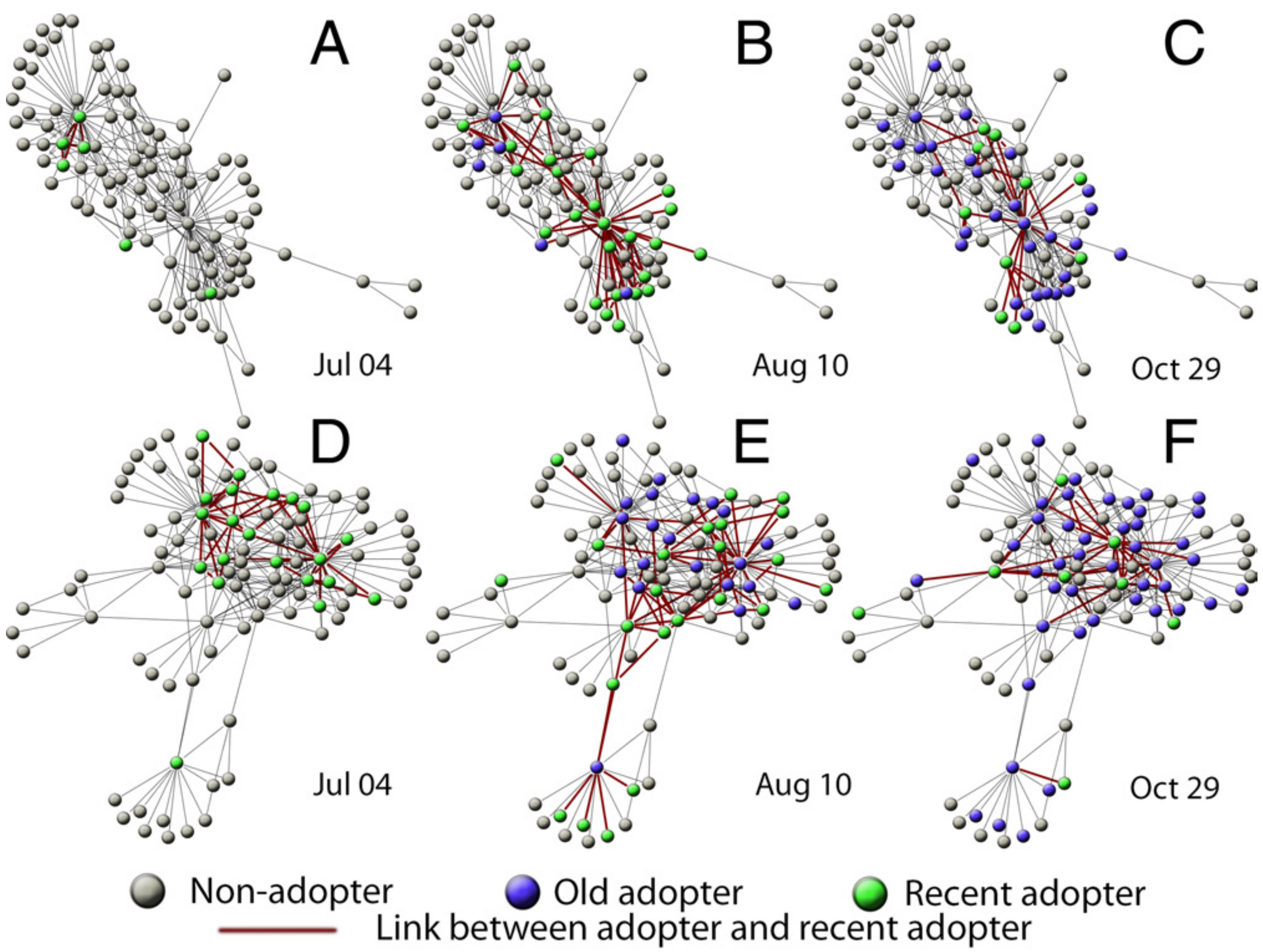
Number of pairs of people at different ages

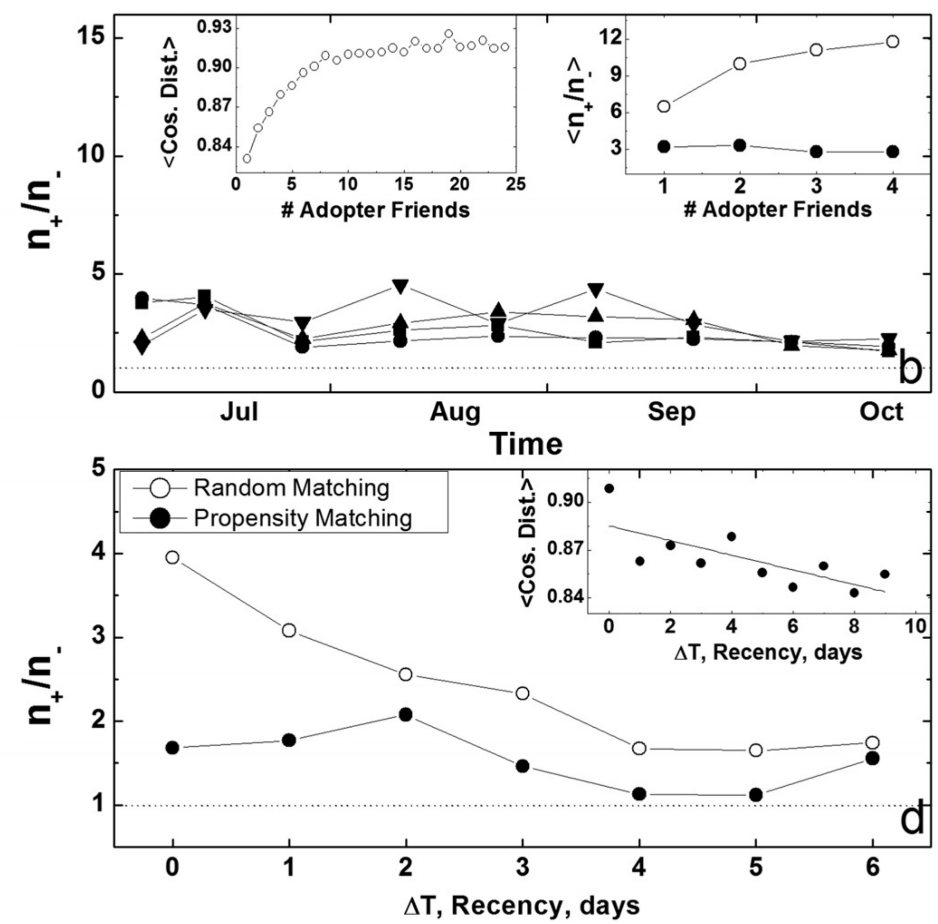


Leskovec, J. & Horvitz, E., 2008. Planetary-scale views on a large instant-messaging network. pp.915–924.

Social dynamical processes

- ***Contagion = Homophily?***
  - *Influence and homophily are usually confounded in observational social network studies*

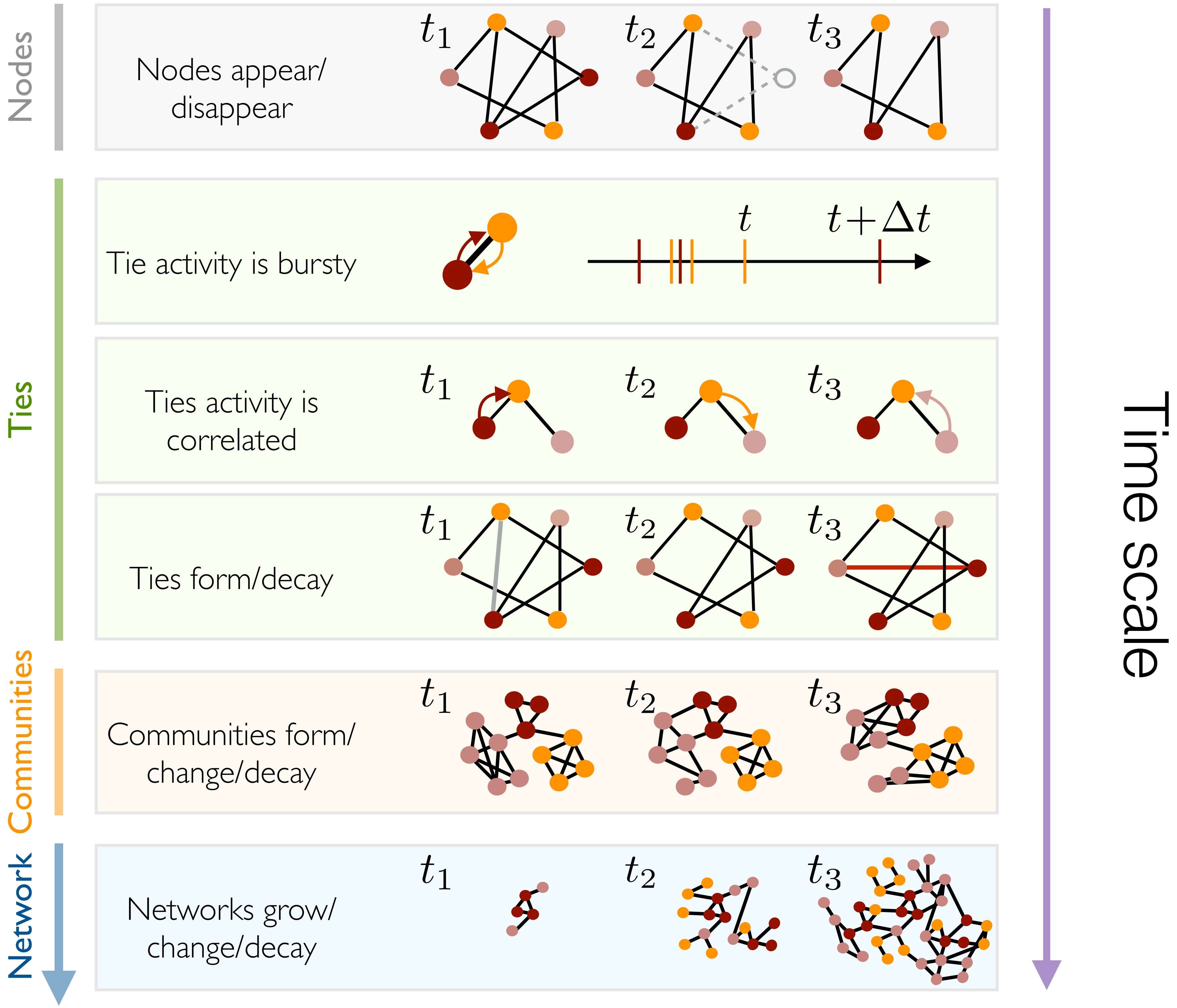




Aral, S., et al. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51), p.21544.

Universidad Carlos III de Madrid

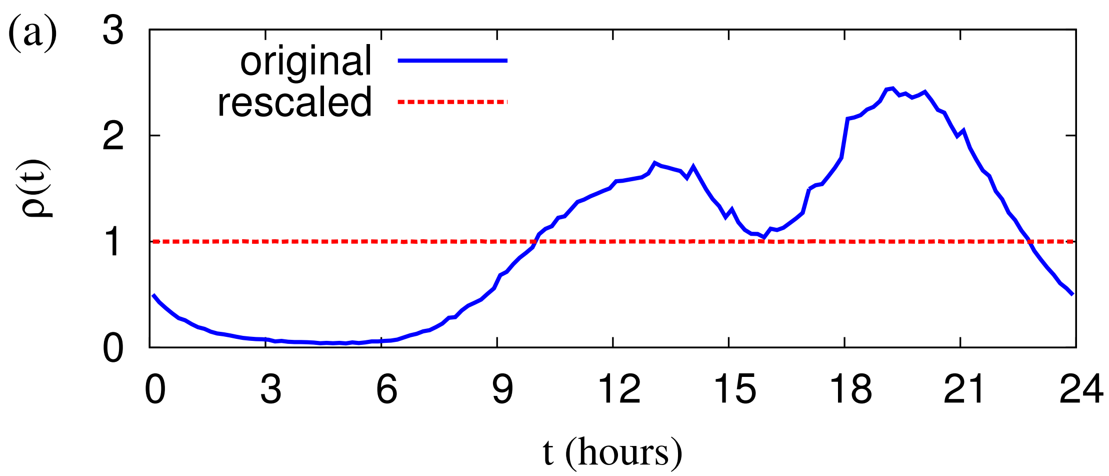
@estebanmoro

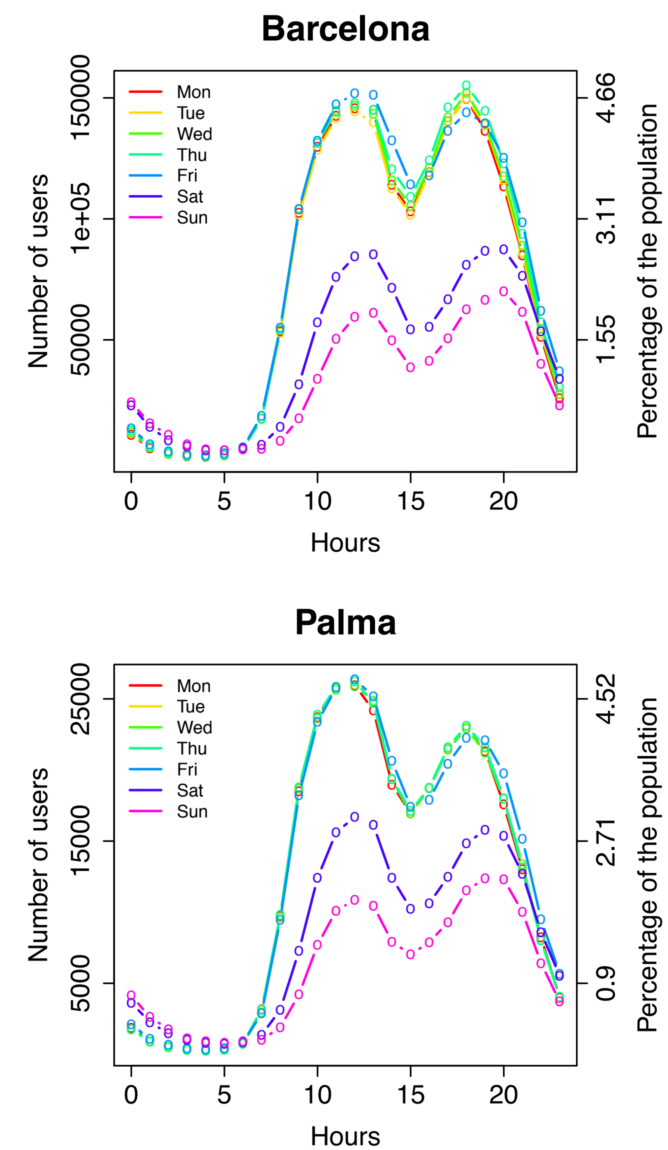


Individual dynamics 2.1

Individual dynamics

- Humans distribute their time differently along the day (circadian rhythms)



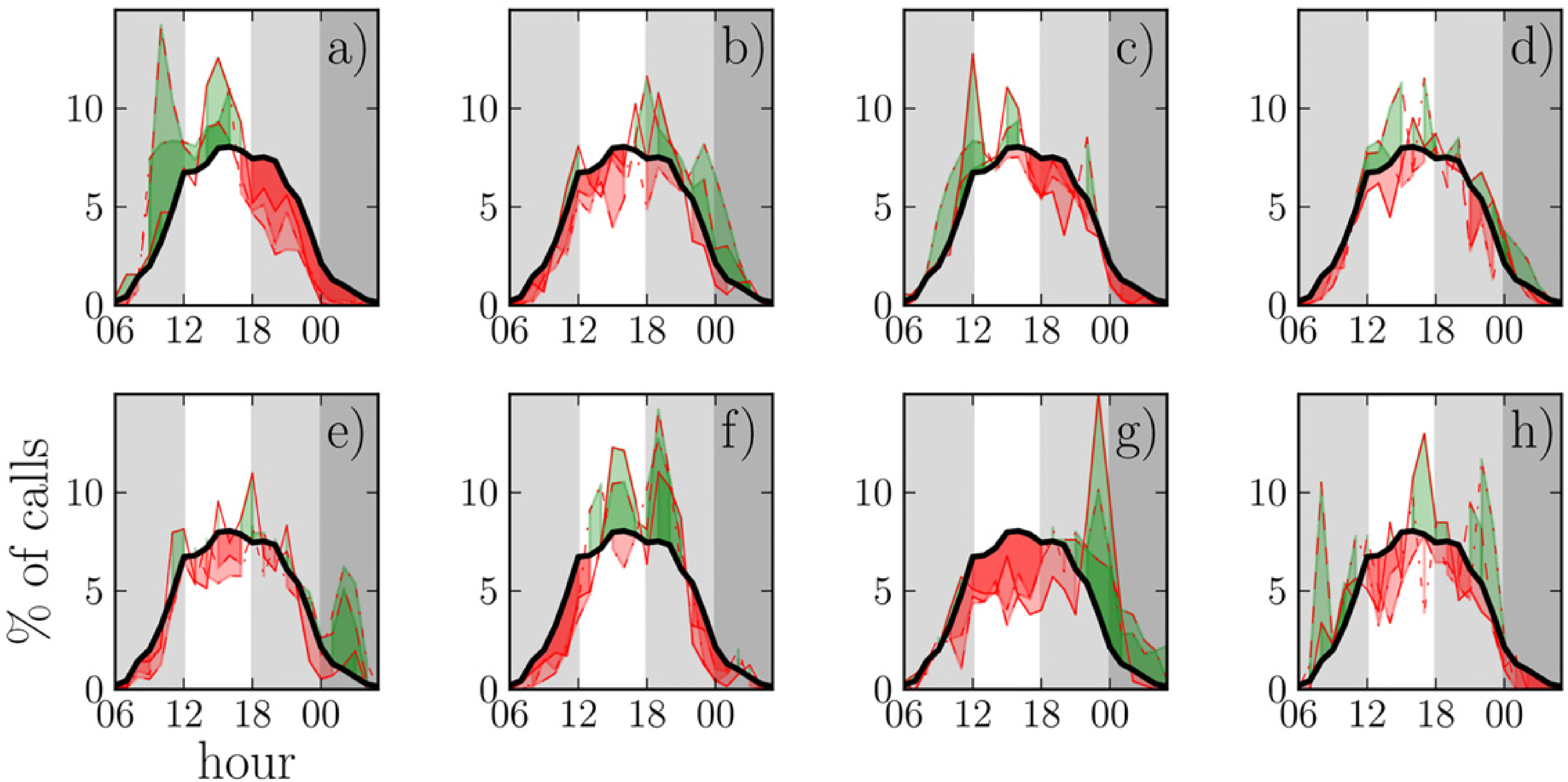


Jo, H.-H., Karsai, M., Kertesz, J., & Kaski, K. (2012). Circadian pattern and burstiness in mobile phone communication. *New Journal of Physics*, *14*(1), 013055. http://doi.org/10.1088/1367-2630/14/1/013055

Louail, T., Lenormand, M., Cantú, O. G., Picornell, M., Herranz, R., Frias-Martinez, E., et al. (2014). From mobile phone data to the spatial structure of cities. *Scientific Reports*, *4*. http://doi.org/10.1038/srep05276

Individual dynamics

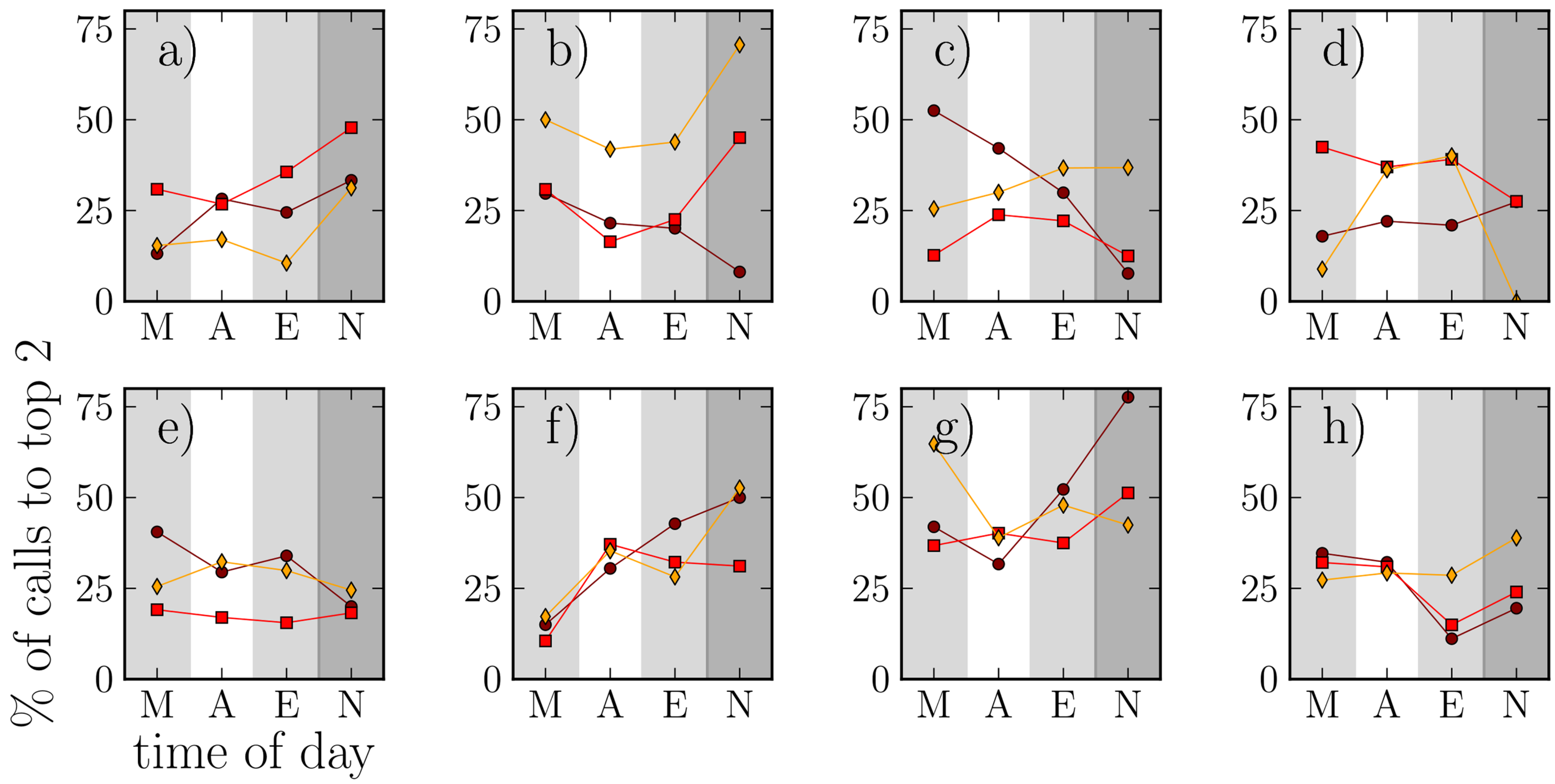
- Individual heterogeneity is significant and persistent.

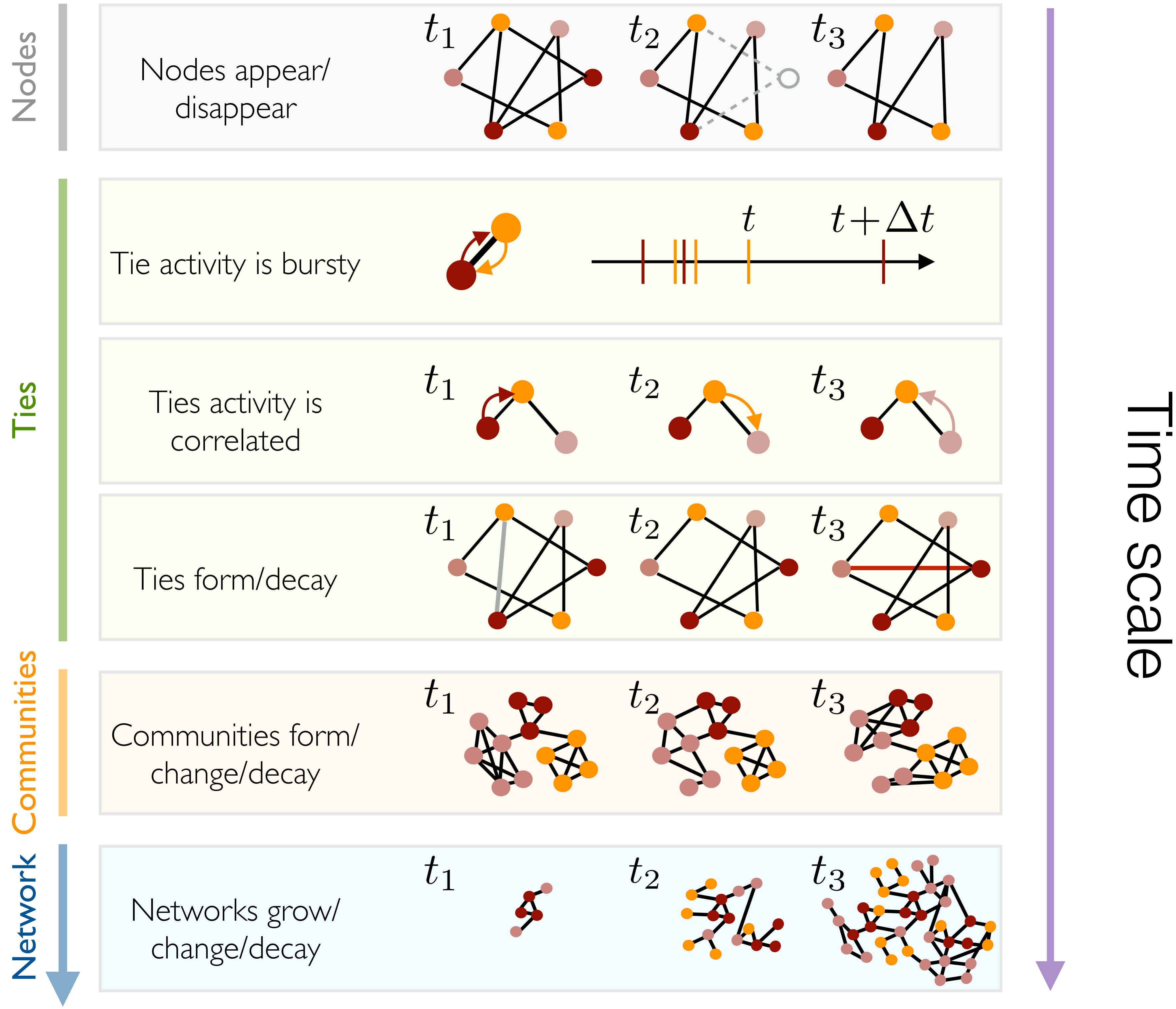


Aledavood, T., López, E., Roberts, S. G. B., Reed-Tsochas, F., Moro, E., Dunbar, R. I. M., & Saramaki, J. (2015). Daily Rhythms in Mobile Telephone Communication. *PLoS ONE*, *10*(9), e0138098. http://doi.org/10.1371/journal.pone.0138098

Individual dynamics

- Calling patterns are different at different times of the day
  - At mornings we call a lot of new people
  - At nights we call less people and those are the more significant ones

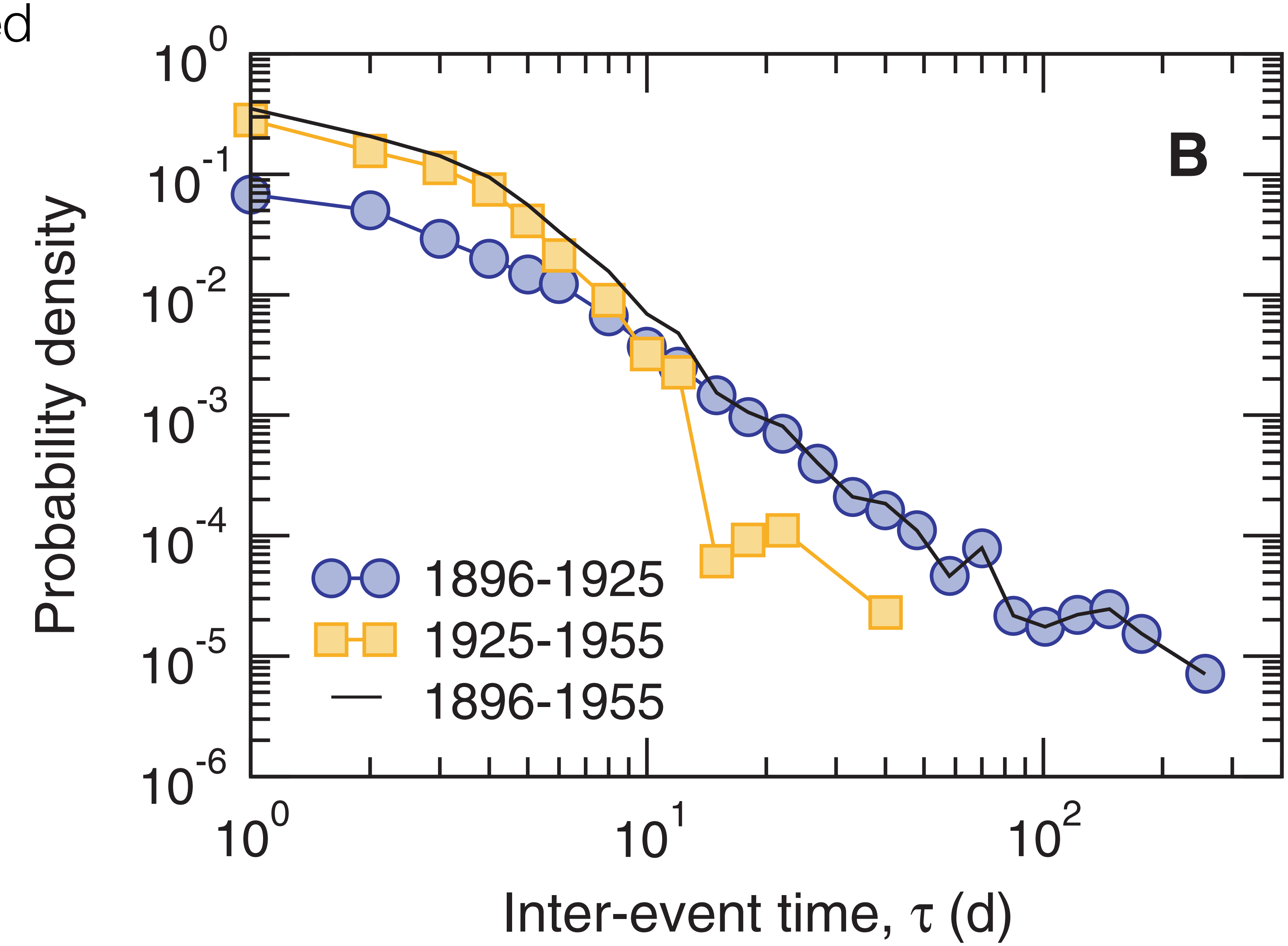




Tie dynamics 2.2

Contact dynamics

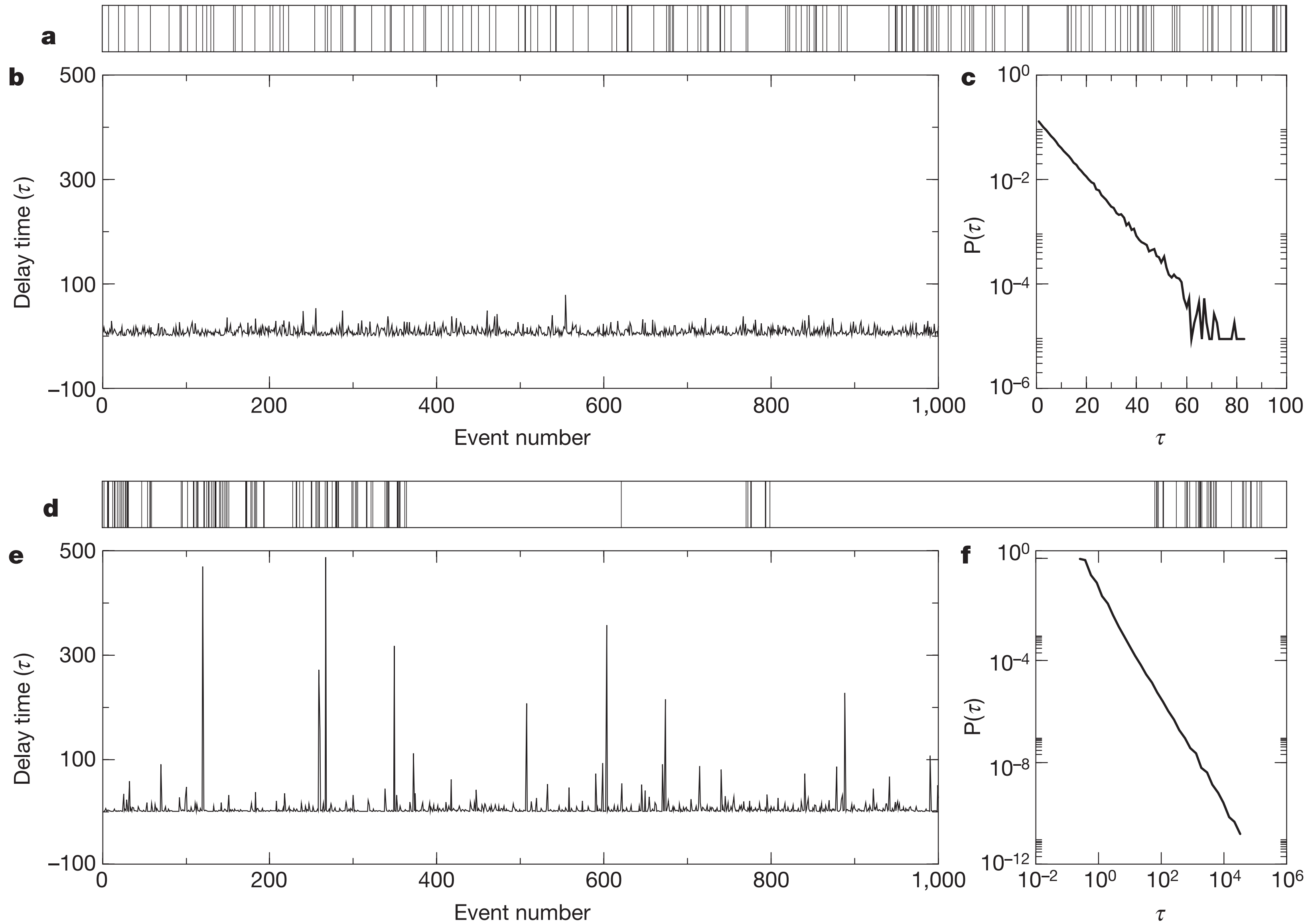
- Bursty human dynamics: inter-event time between activities is heavy-tailed distributed



Malmgren, R. et al., 2009. On universality in human correspondence activity. *Science*, 325(5948), p.1696.

Contact dynamics

- Bursty human dynamics: inter-event time between activities is heavy-tailed distributed



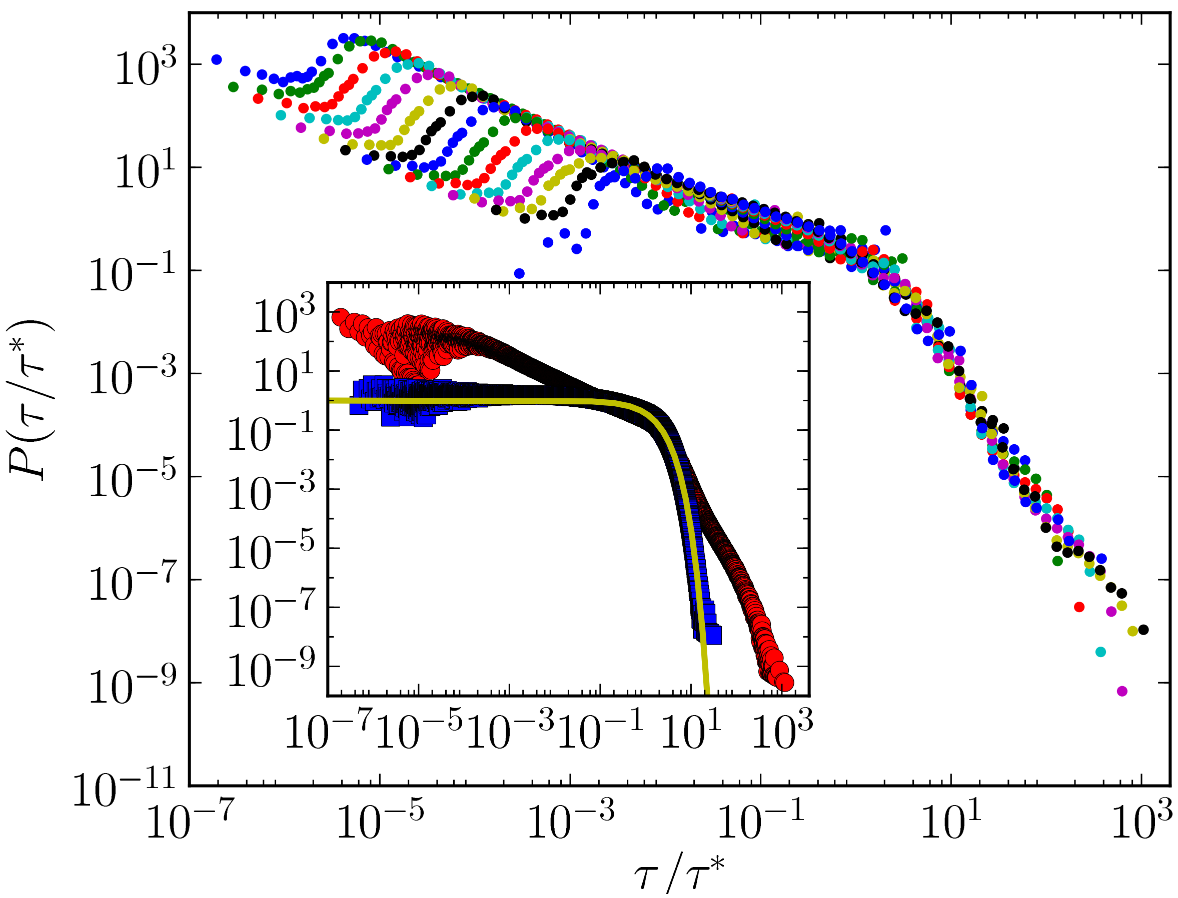
Barabasi, A.-L., 2005. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039), pp.207–211.

Universidad Carlos III de Madrid

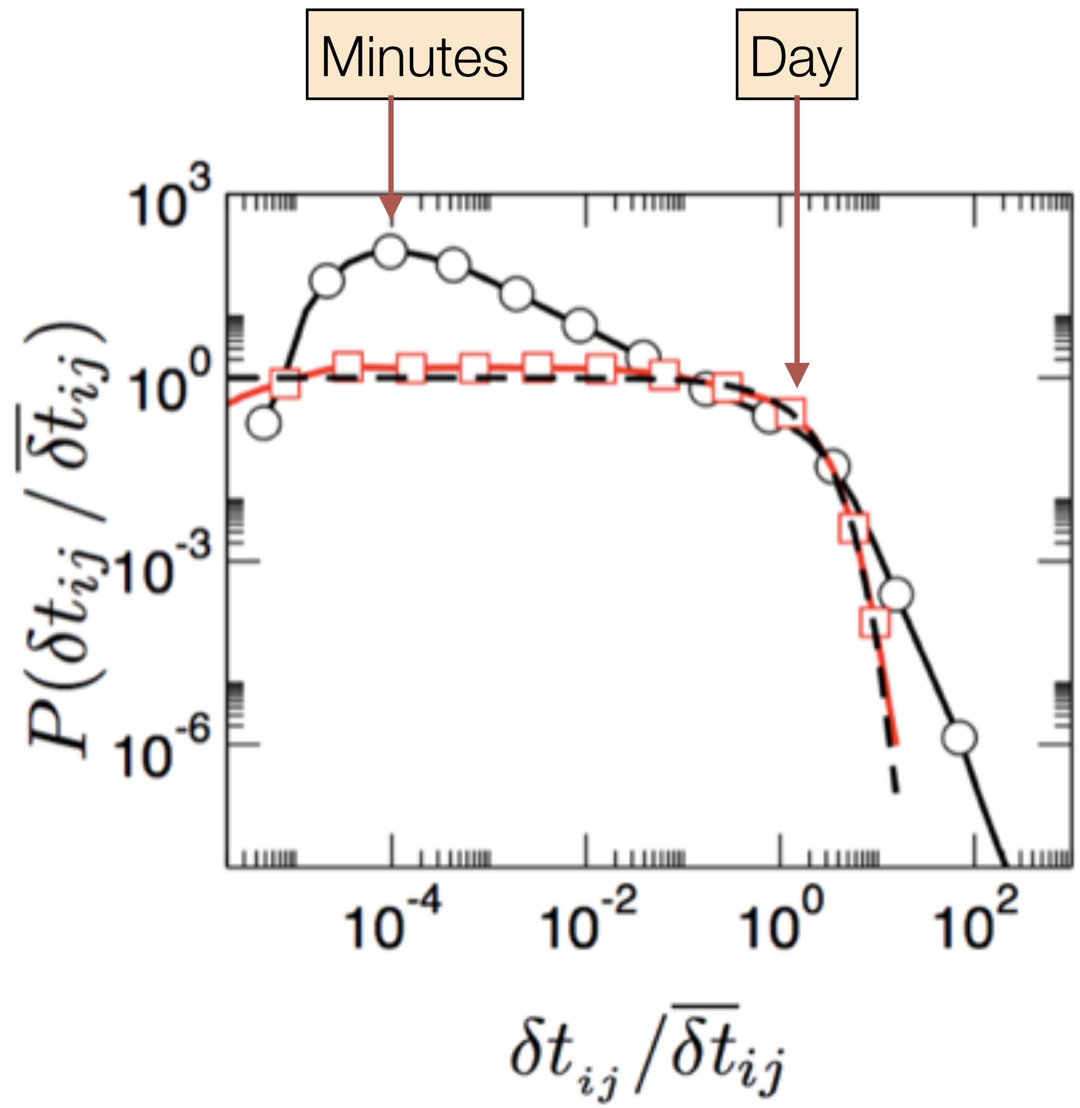
@estebanmoro

Tie activity

- Bursty contacts: inter-event times on ties are also heavy-tailed distributed



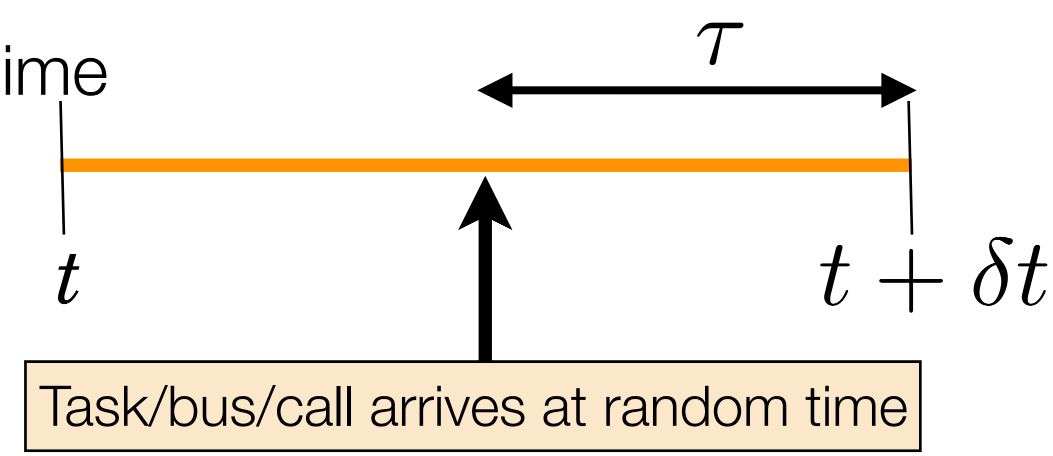
Karsai, M. et al., 2011. Small But Slow World: How Network Topology and Burstiness Slow Down Spreading. *Physical Review E*, 83(2), p.025102.



Miritello, G., Moro, E. & Lara, R., 2011. Dynamical strength of social ties in information spreading. *Physical Review E*, 83(4), p.045102.

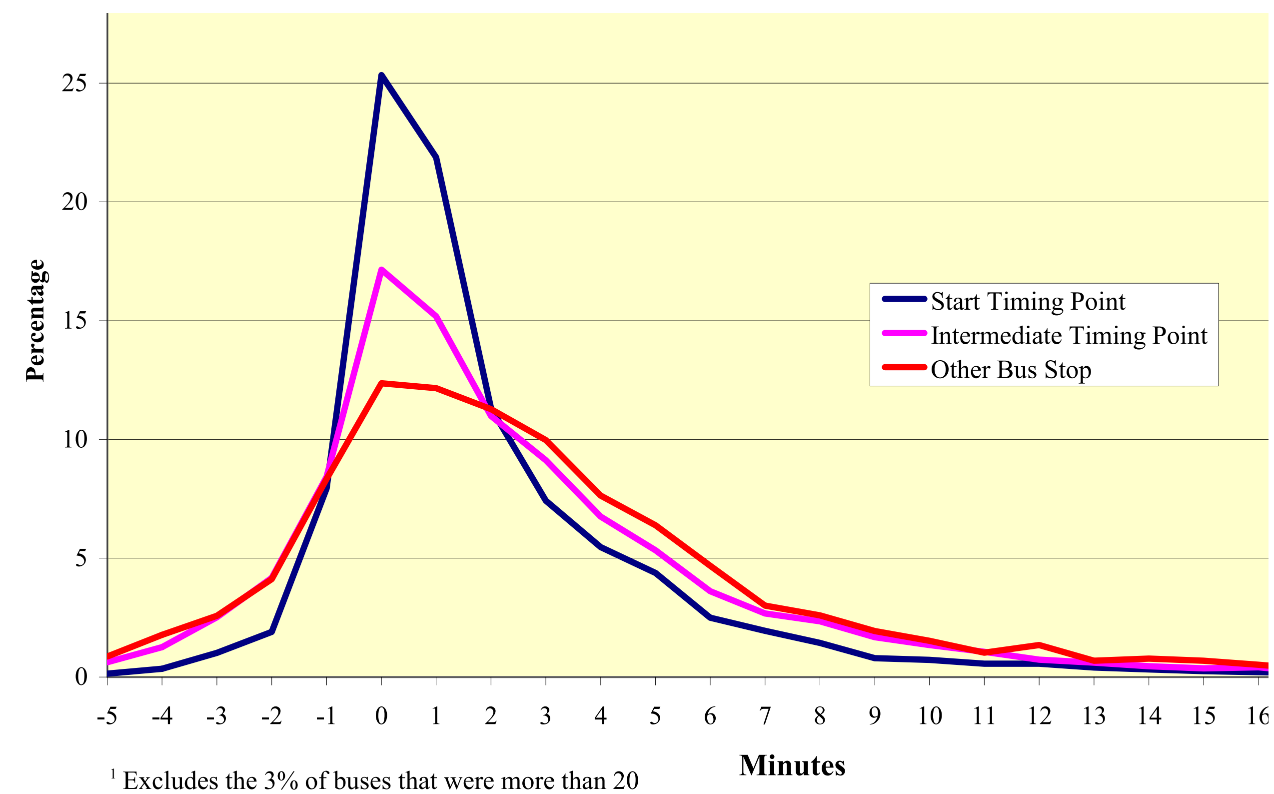
Tie activity

- Bursty contacts: impact on the waiting/response time
- When should I wait next call from a friend?
- When is the next bus coming?
    - Given $P(\delta t)$, calculate $P(\tau)$



$$P(\tau) = \int_{\tau}^{\infty} d\delta t \, \frac{\delta t P(\delta t)}{\overline{\delta t}} \, \frac{1}{\delta t}$$

$$\overline{\tau} = \frac{\overline{\delta t}}{2}\left(1 + \frac{\sigma_{\delta t}^2}{\overline{\delta t}^2}\right)$$



Bus Punctuality Statistics GB 2007. Dept. of Transport

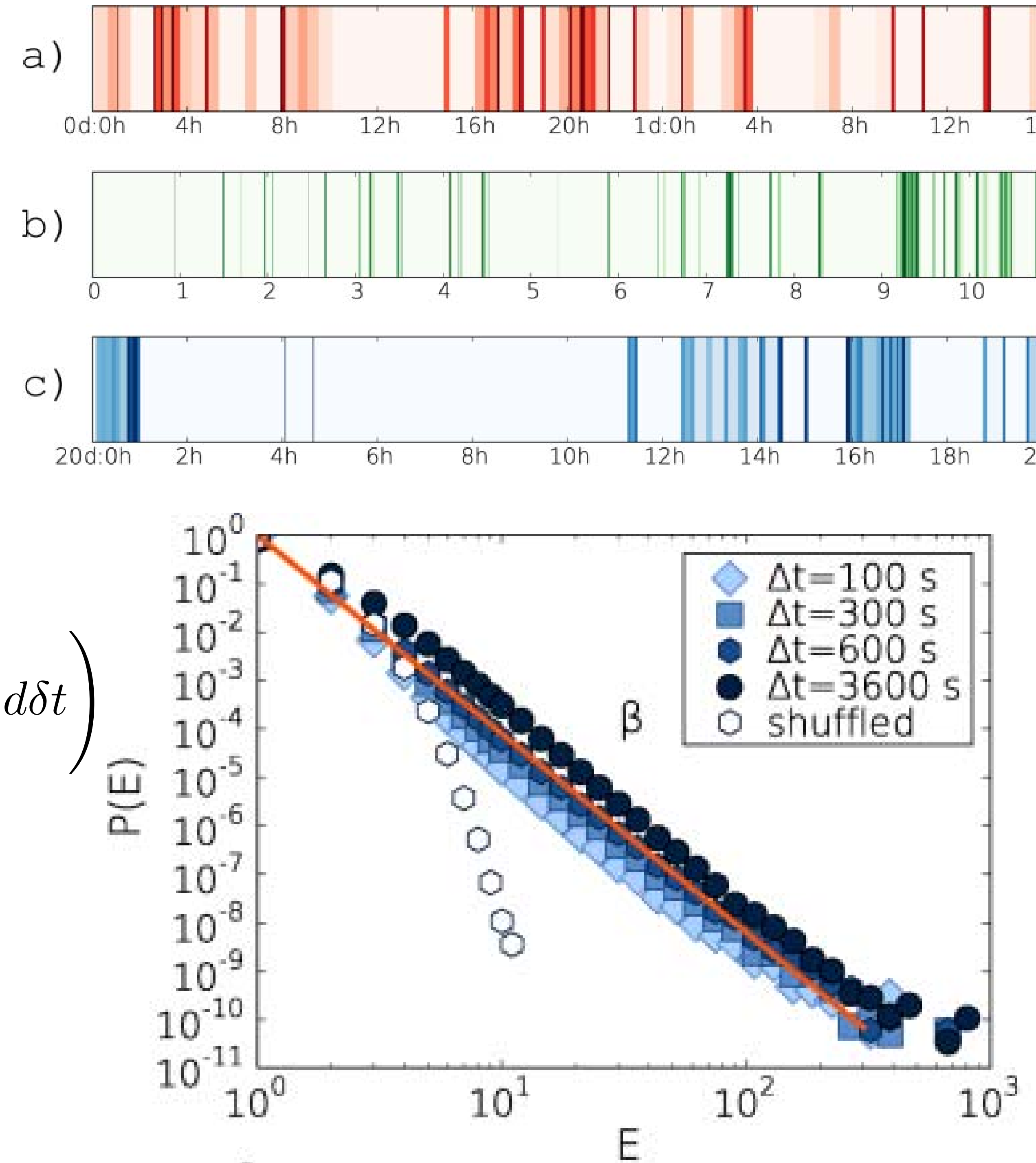
Tie activity

Karsai, M. et al., 2012. Universal features of correlated bursty behaviour. *Scientific Reports*, 2.

- Is that all? Nope: bursts are correlated in time
- To find correlation, detect sequence of events with $\delta t < \Delta t$
- If activity is a renewal process, the probability that we find *n* of such events in a row is

$$P(E=n) = \left(\int_0^{\Delta t} P(\delta t) d\delta t\right)^{n-1} \left(1 - \int_0^{\Delta t} P(\delta t) d\delta t\right)$$

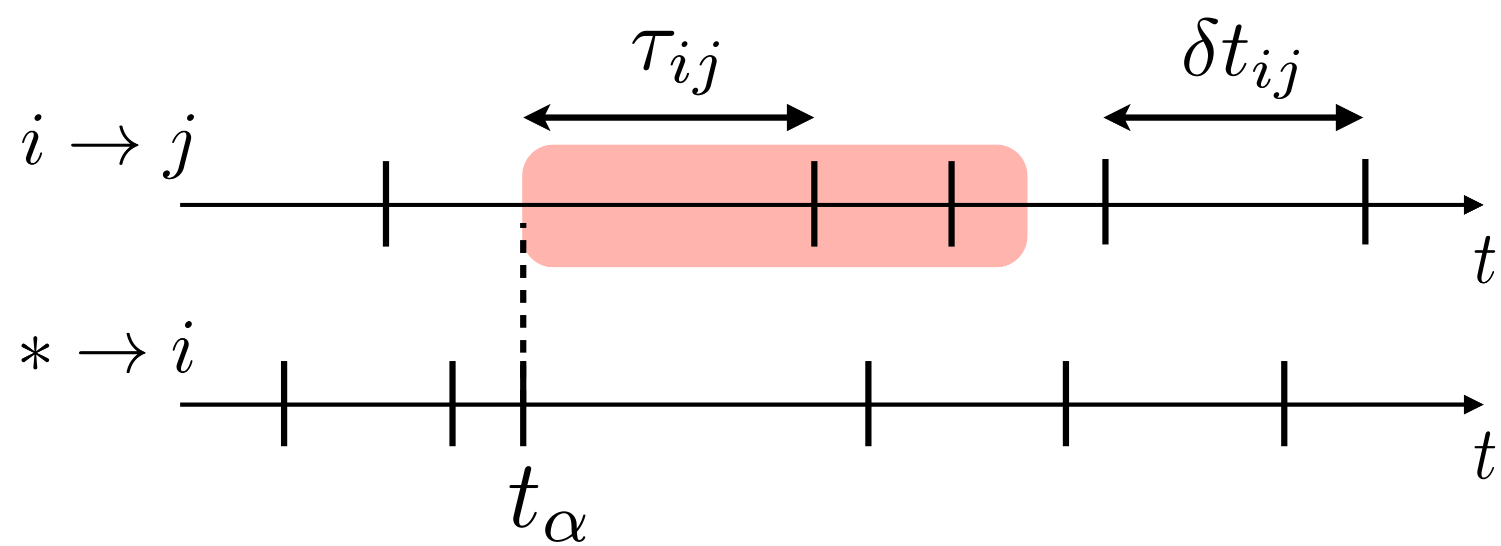
- $P(E)$ decays exponentially
- However, in real data it decays like a power-law

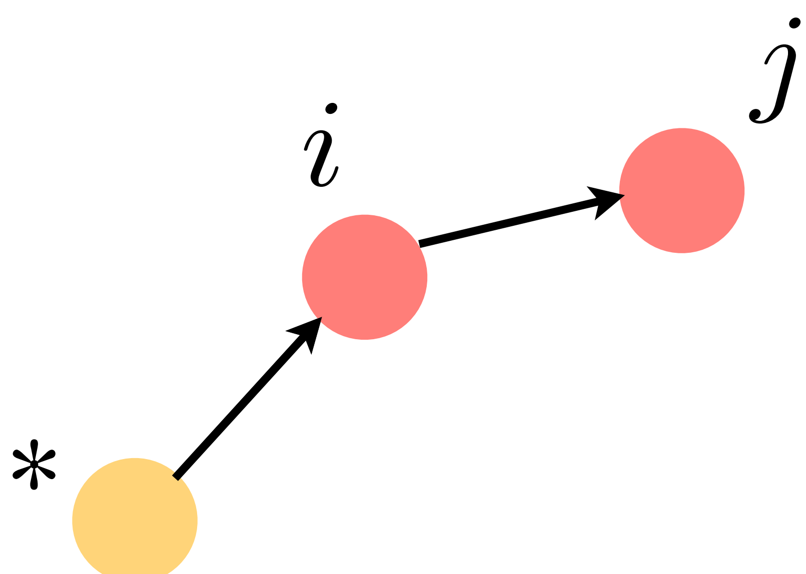


Tie activity

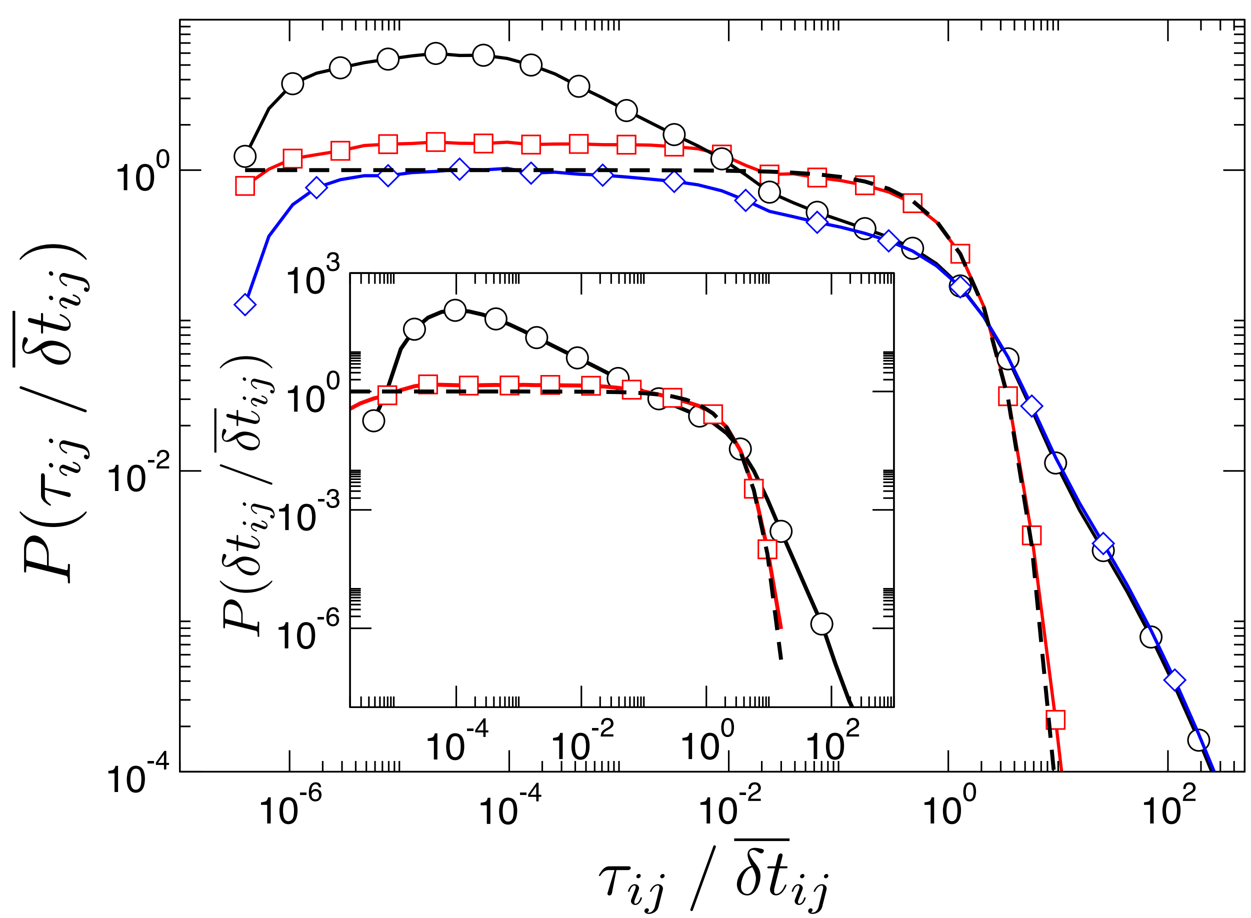
Miritello, G., Moro, E. & Lara, R., 2011. Dynamical strength of social ties in information spreading. *Physical Review E*, 83(4), p.045102.

- Is that all? Nope
  - Adjacent tie contacts are correlated in time



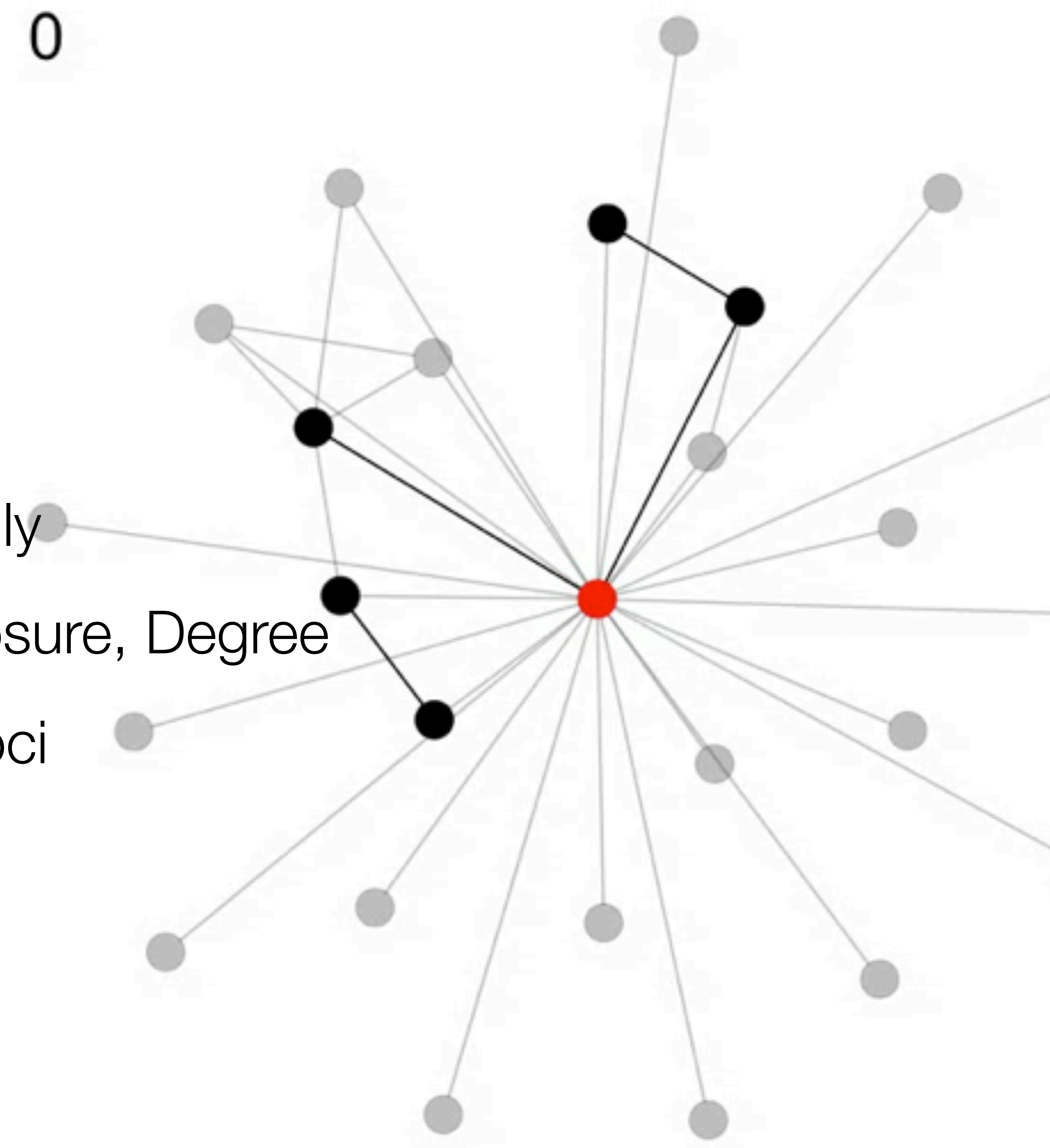


$$P(\tau) = \int_\tau^\infty d\delta t \, \frac{\delta t P(\delta t)}{\overline{\delta t}} \, \frac{1}{\delta t}$$



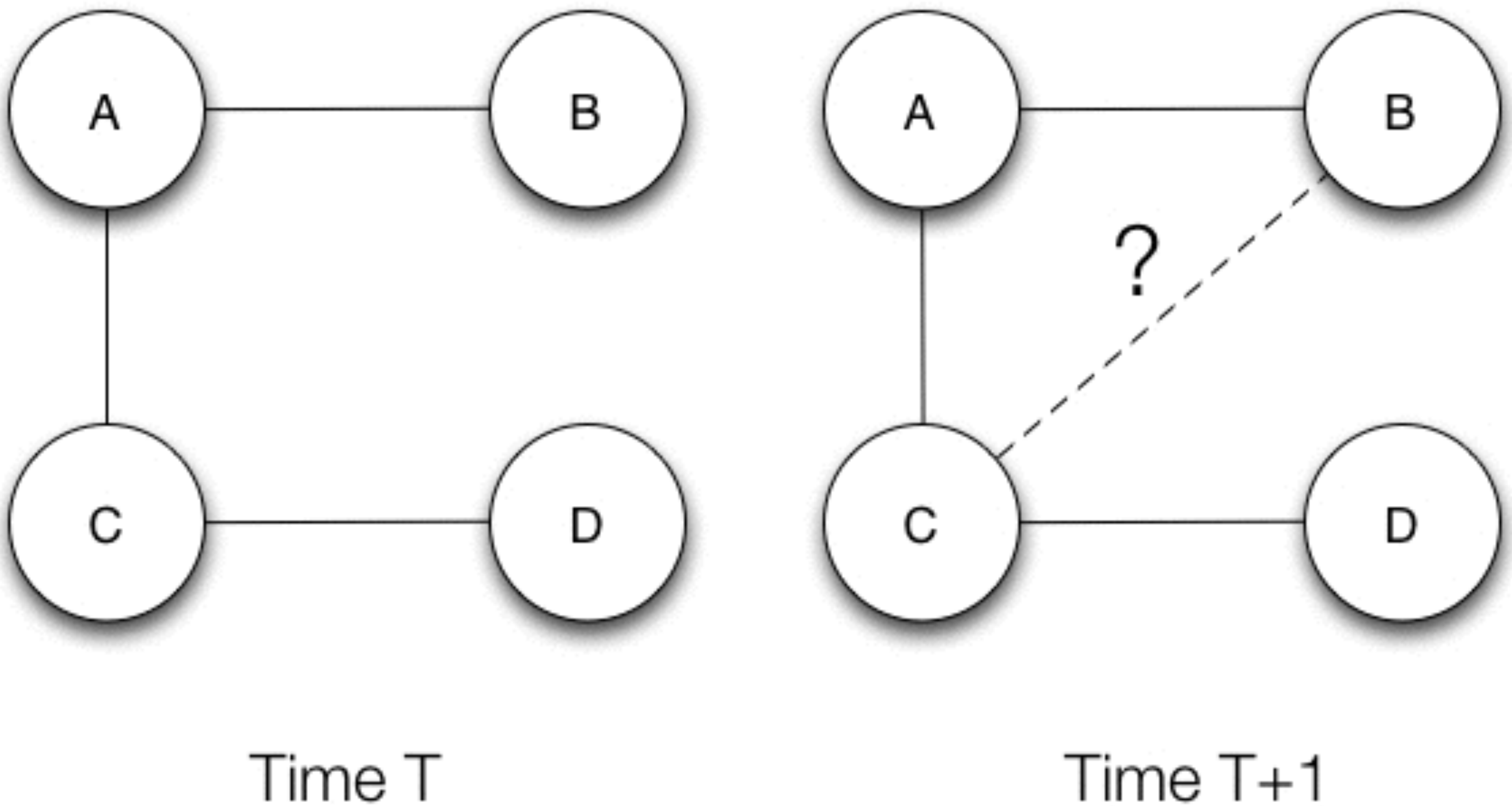
Tie dynamics

- Ties are formed and decay
  - **Why**?
    - *Creation*
      - Node creation/decay
      - **Assortative**: Homophily/Heterophily
      - **Relational**: Reciprocity, Triadic Closure, Degree
      - **Proximity**: Proximity and Social Foci
    - *Decay*
      - **Idem**
  - **How**?
    - Social limitations / strategies



Tie dynamics

- Tie formation: Relational predictors



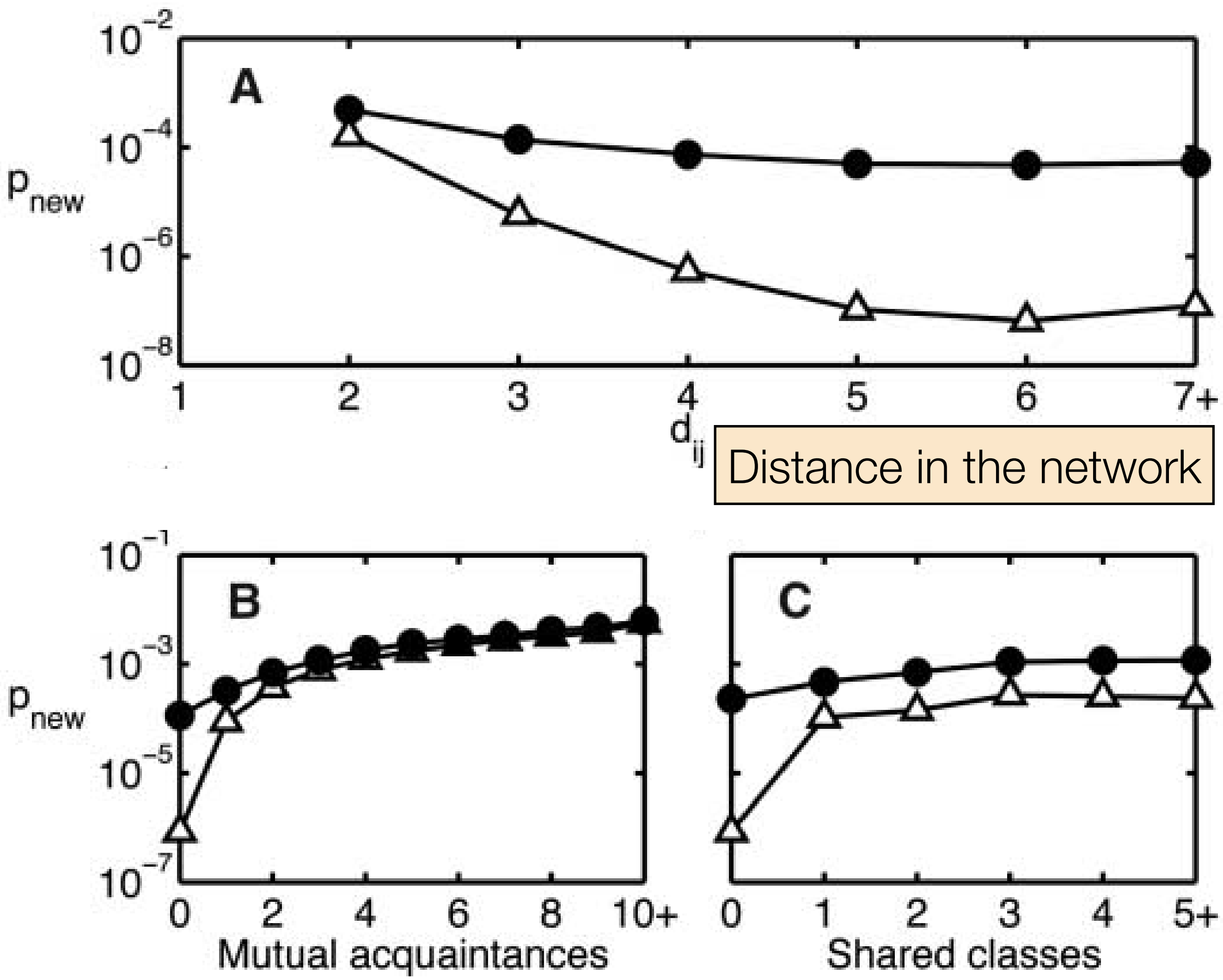
- Triadic closure
- Mutual acquaintances / embeddedness

Tie dynamics

Liben Nowell, D. & Kleinberg, J., 2007. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7), pp.1019–1031.

- Tie formation: predictors

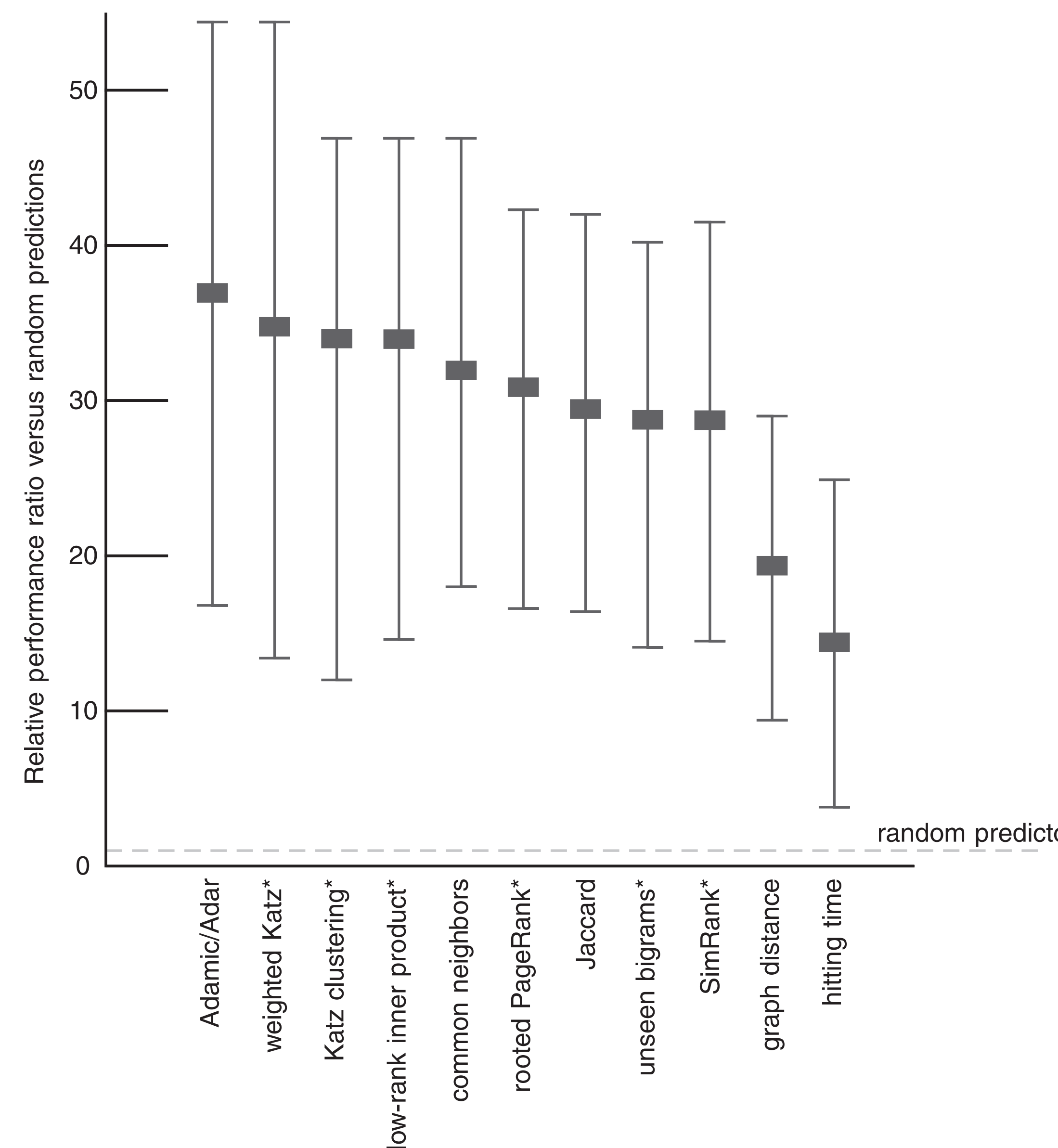
Probability of a new tie



Distance in the network

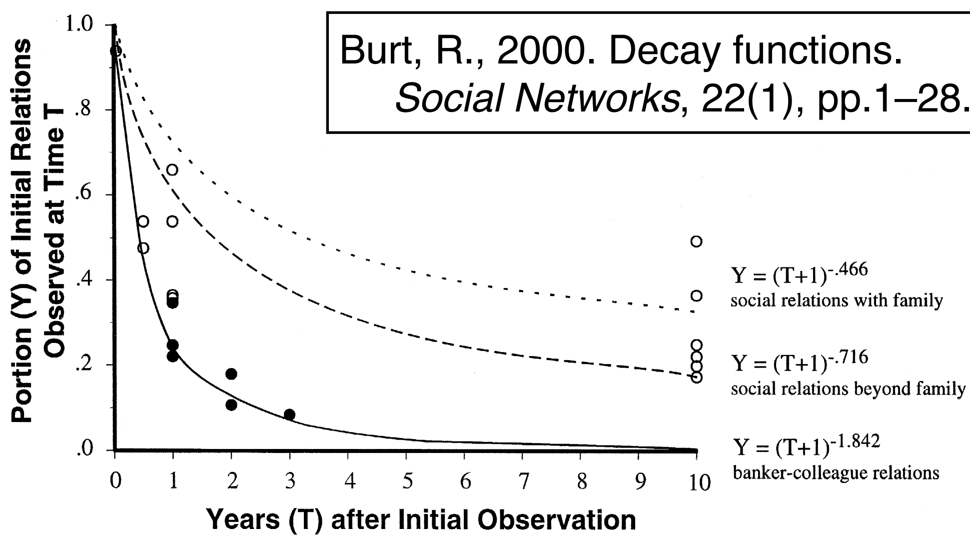
Kossinets, G. & Watts, D.J., 2006. Empirical analysis of an evolving social network. *Science*, 311(5757), pp.88–90.

$$\sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{\log|\Gamma(z)|}$$



Tie dynamics

- Tie decay: predictors



Burt, R., 2000. Decay functions. *Social Networks*, 22(1), pp.1–28.

Hidalgo, C. & Rodriguez-Sickert, C., 2008. The dynamics of a mobile phone network. *Physica A*, 387(12), pp.3017–3024.

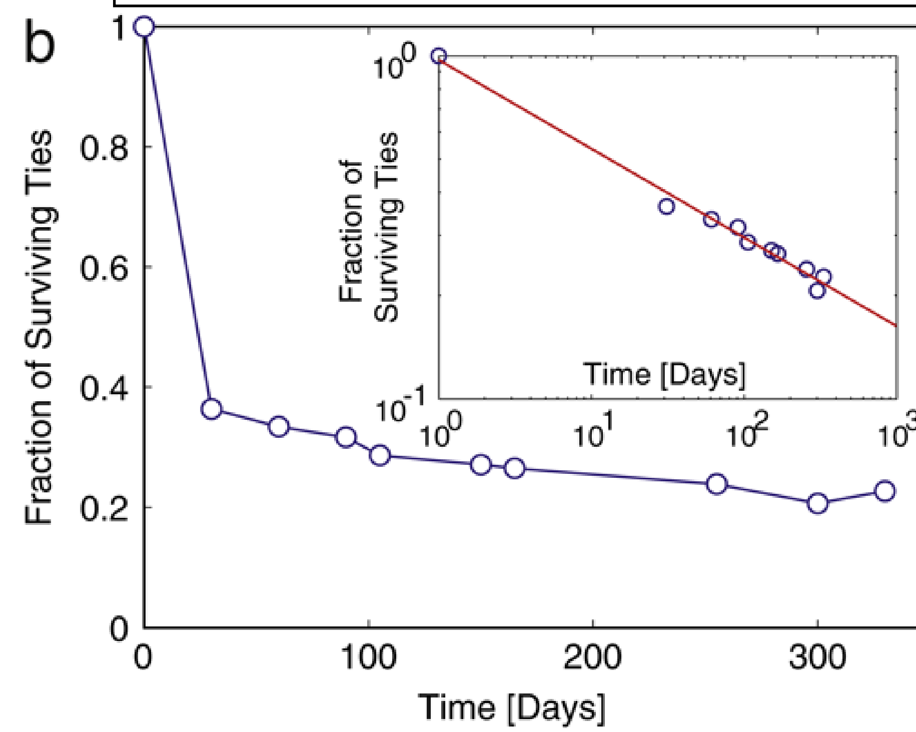


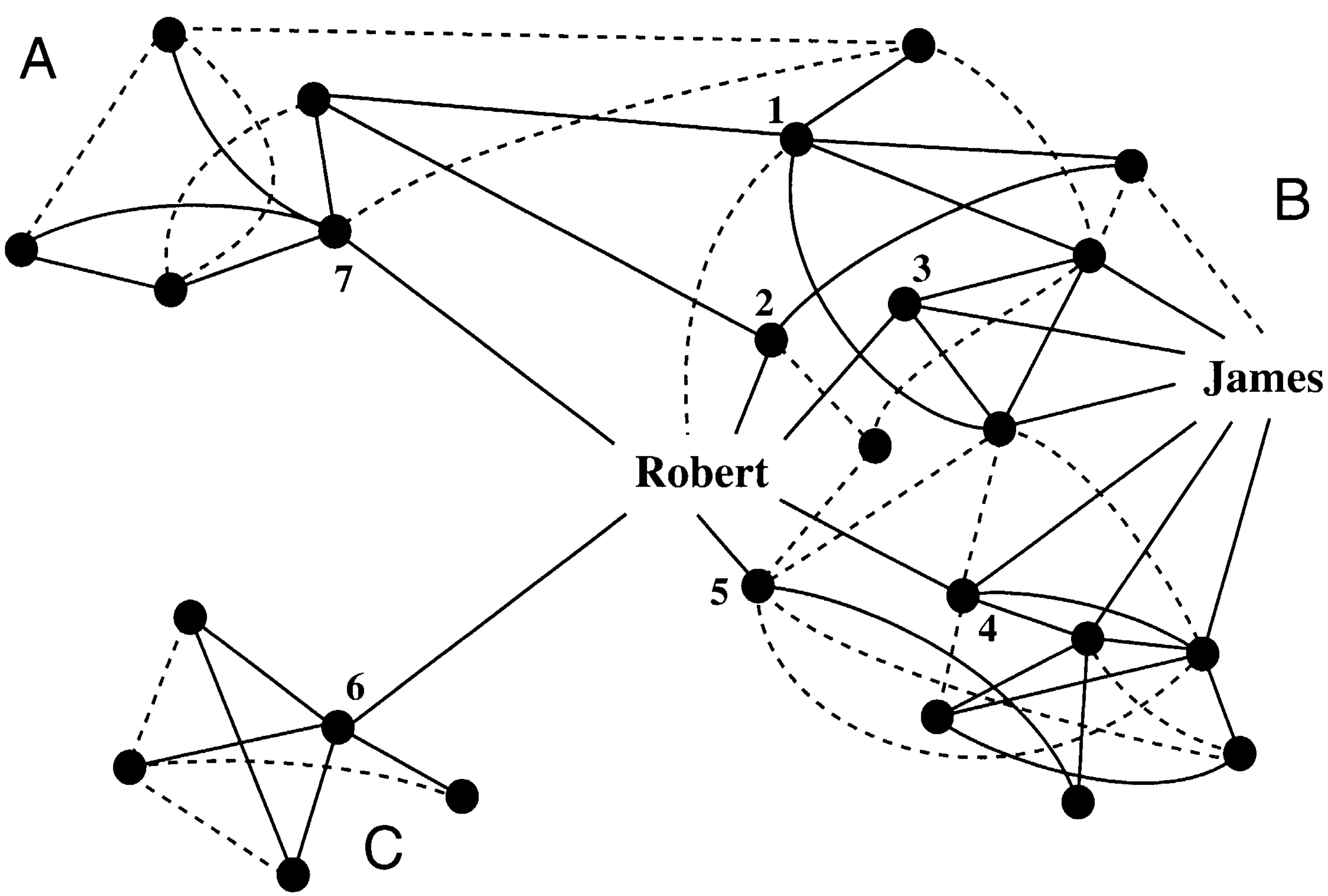
Table 1
Persistence of ties and link attributes

Pearson's correlation	ΔC	Δk	Δr	R	TO	Persistence
ΔC	1	0.023	0.15	0.11	0.23	0.15
Δk		1	0.02	−0.13	−0.19	−0.16
Δr			1	−0.68	−0.073	0.033
R				1	0.2964	0.5886
TO					1	0.3537
Regression coefficients	0.09	0.002	0.15	0.35	0.56	
Partial correlations	0.0027	0.0032	0.007	0.26	0.034	

- Links with large embeddedness and reciprocity are more likely to persist

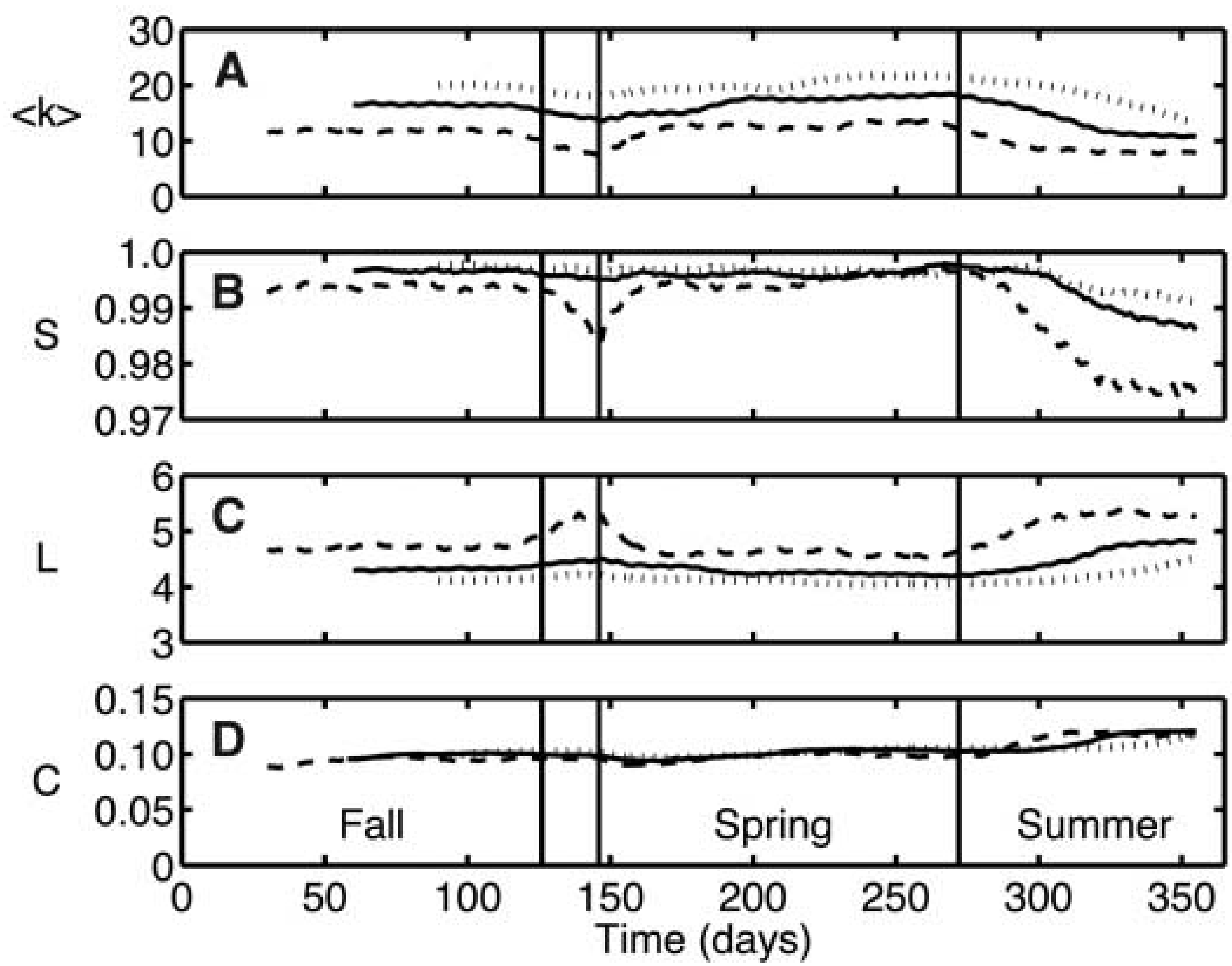
Tie dynamics

- Two paradoxes
  - Ties bridging distant parts of the network (the ones important for information diffusion, achievement) are not only the least likely to be created, but also the most likely to decay



Tie dynamics

- Two paradoxes
  - Tie formation tends to close triangles. But ties embedded in triangles are less likely to decay. Thus, network should become more clustered

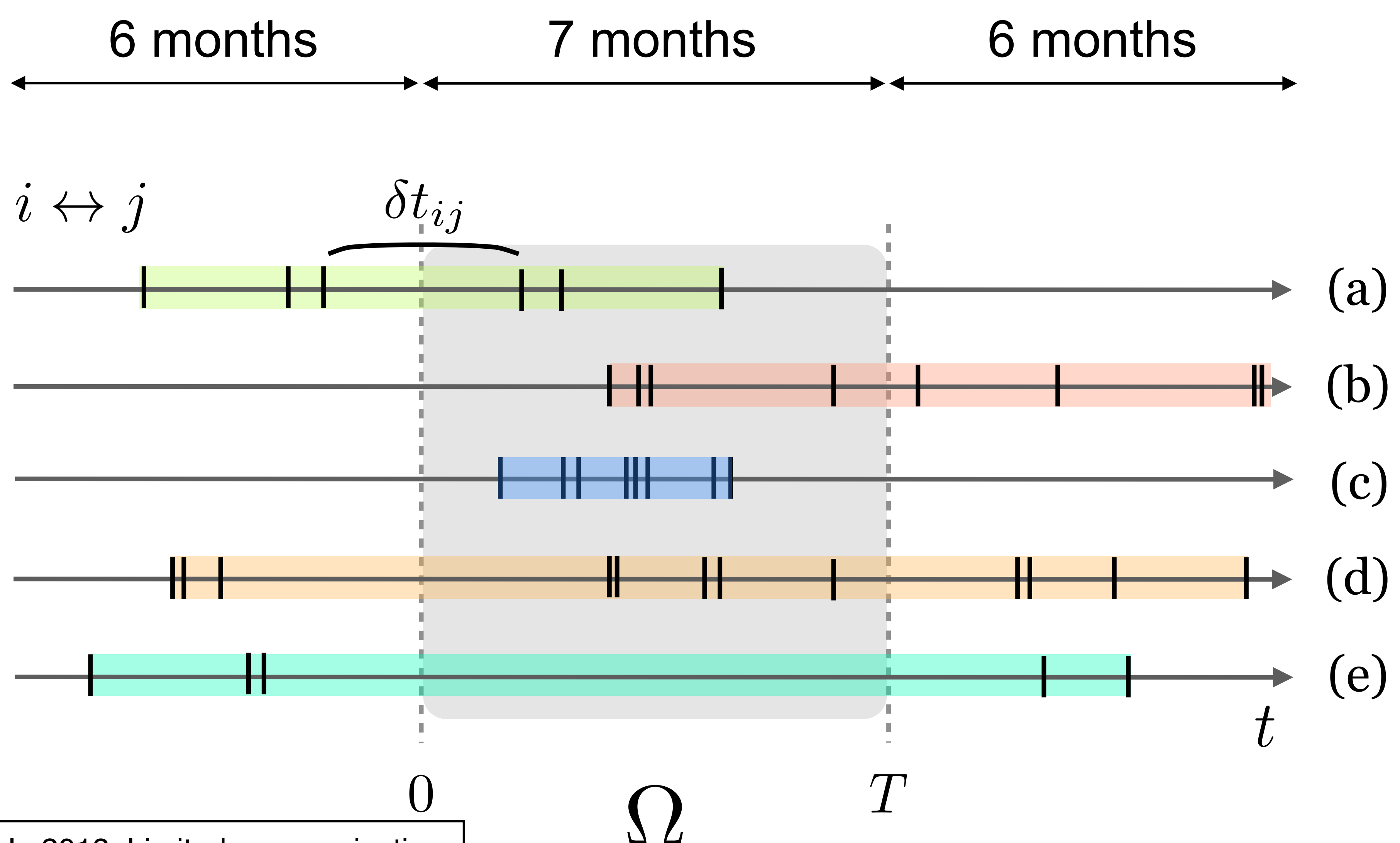


Tie dynamics

- How are ties formed and destroyed? Is there any strategy?
- Cognitive limitations, time limitations
    - Dunbar number: there is a limit to the number of people with whom one can maintain stable social relationships.
    - Time/attention is limited: how do we manage relationships if our time is limited?

Tie dynamics

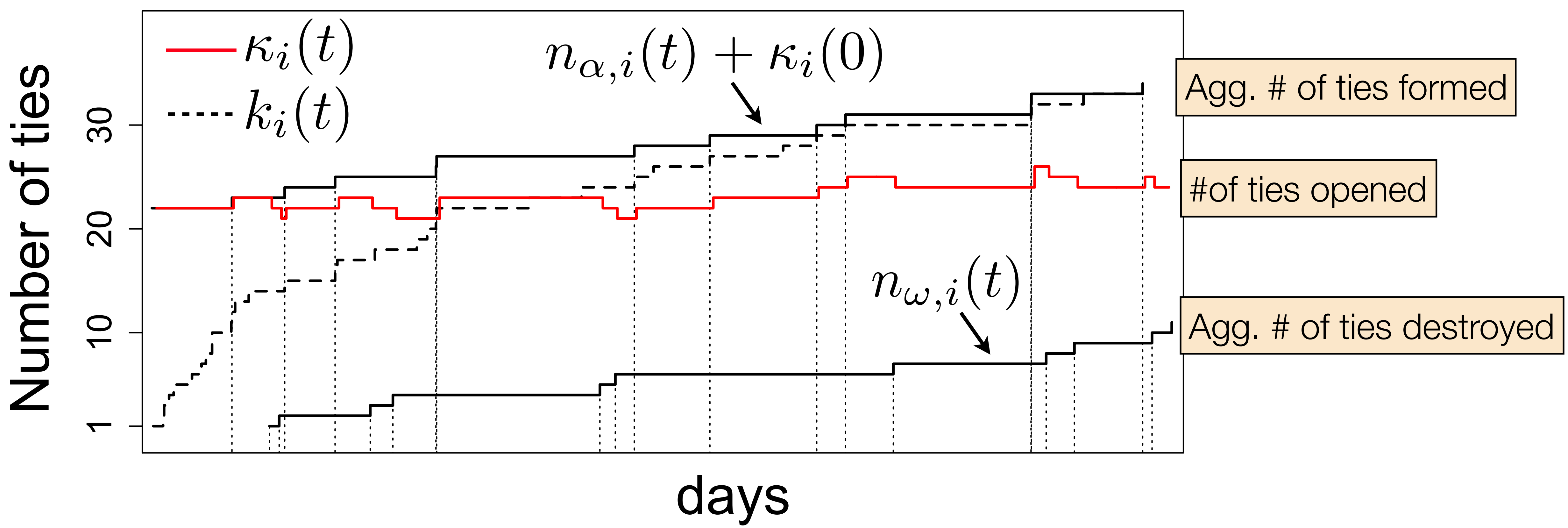
- How are ties formed and destroyed?
- Disentangling tie burstiness and formation/decay



Miritello, G. et al., 2013. Limited communication capacity unveils strategies for human interaction. *Scientific Reports*, 3.

Tie dynamics

- How are ties formed and destroyed?
- Very heterogeneous tie evolution

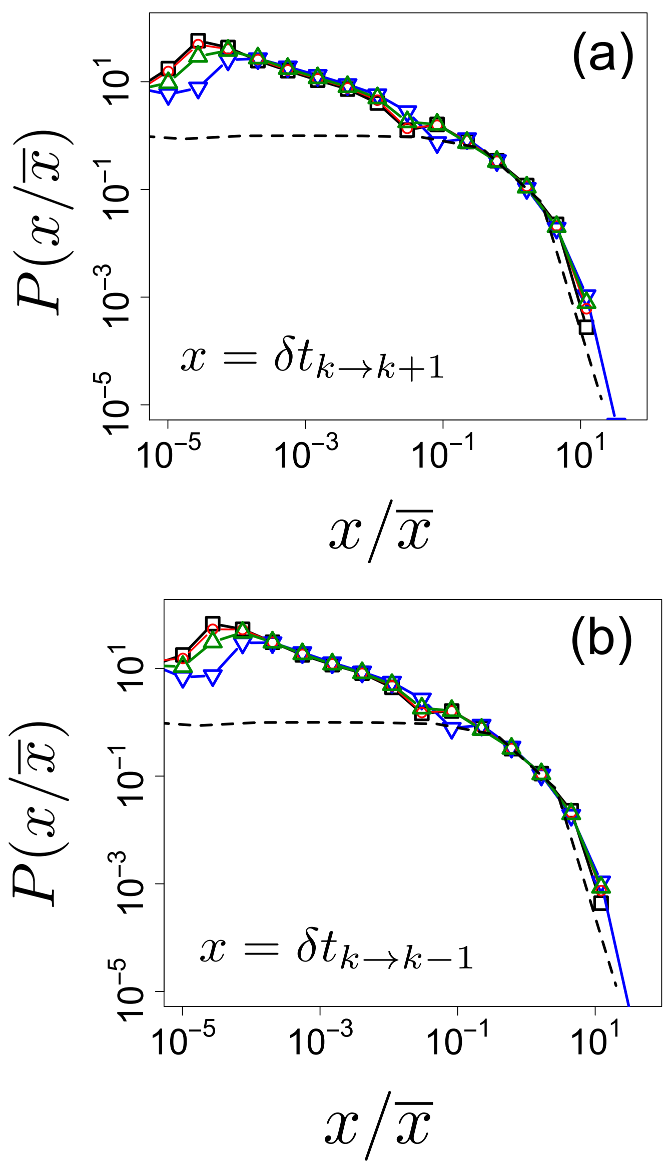


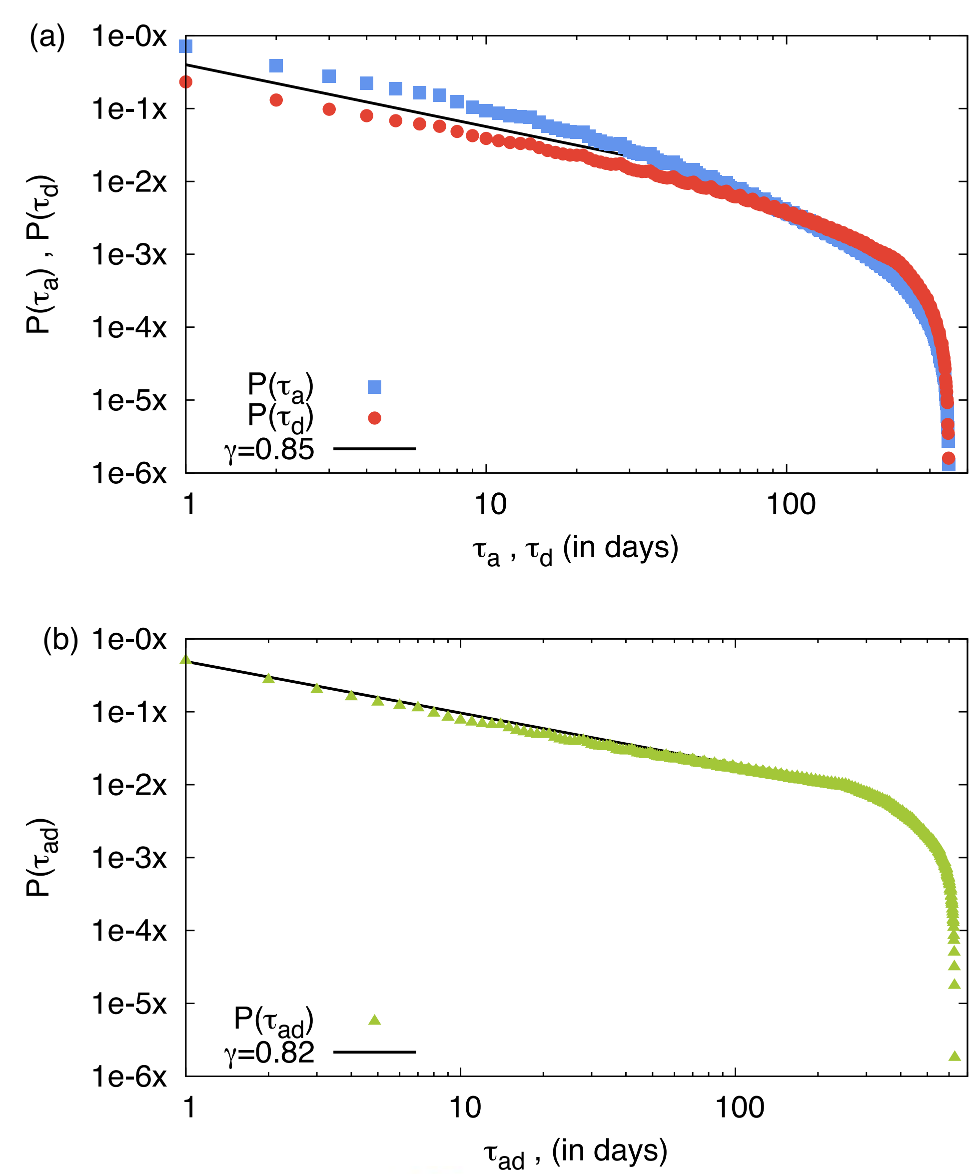
- Mean $\langle n_{\alpha,i} \rangle \simeq \langle n_{\omega,i} \rangle \simeq 8 \quad \langle k_i \rangle \simeq 16$
- But $n_{\alpha,i} > 15$ for 20% of nodes

Tie dynamics

- How are ties formed and destroyed?
- Tie formation/decay is bursty

Kikas, R., Dumas, M. & Karsai, M., 2012. Bursty egocentric network evolution in Skype. *arXiv.org*, physics.soc-ph.

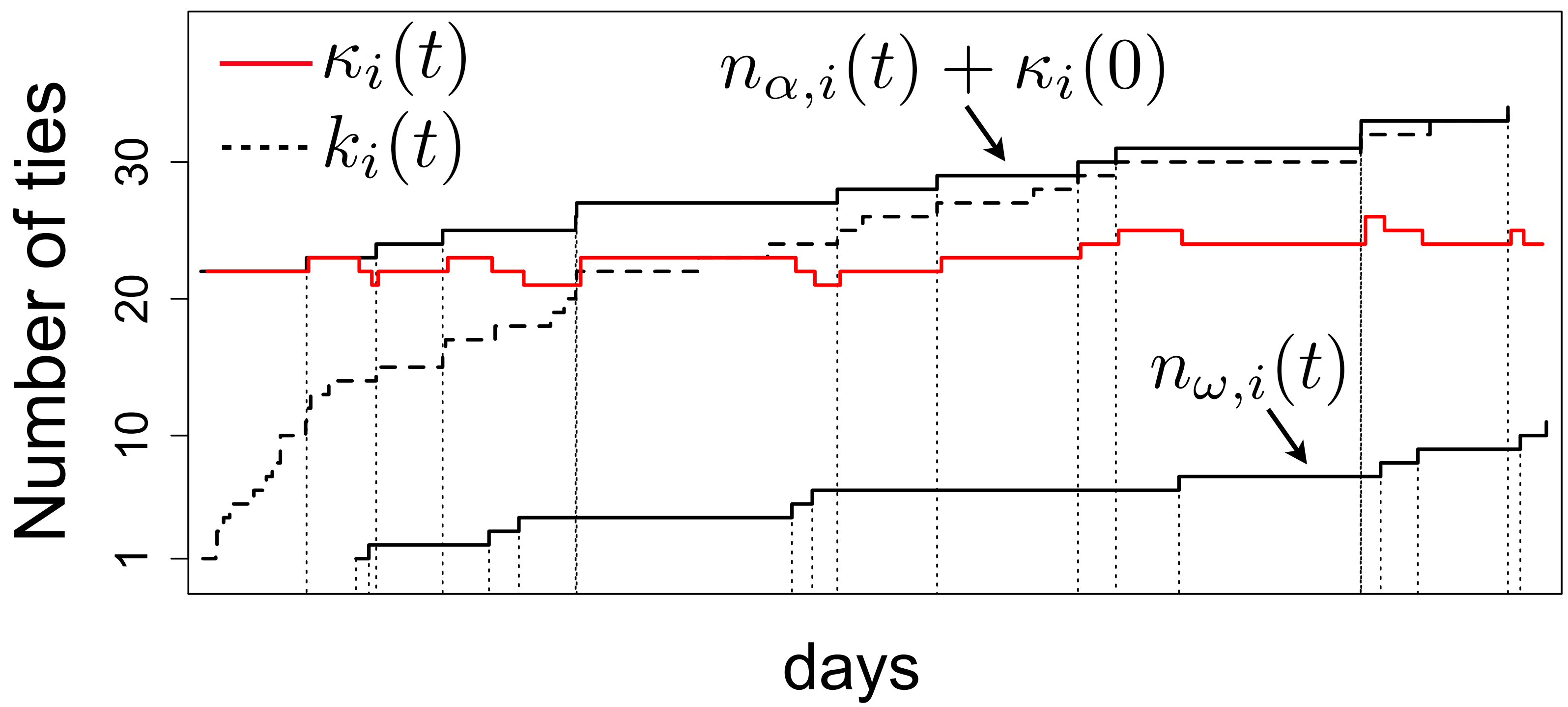


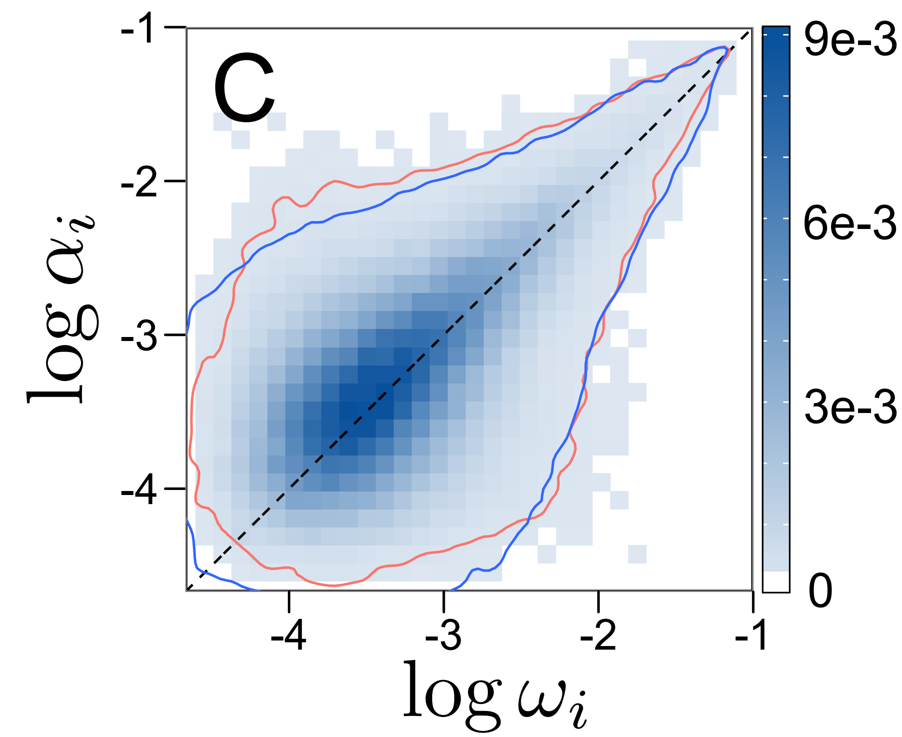


Miritello, G., Temporal Patterns of Communication in Social Networks, Springer 2013

Tie dynamics

- How are ties formed and destroyed?
- Linear tie formation/decay and conserved capacity





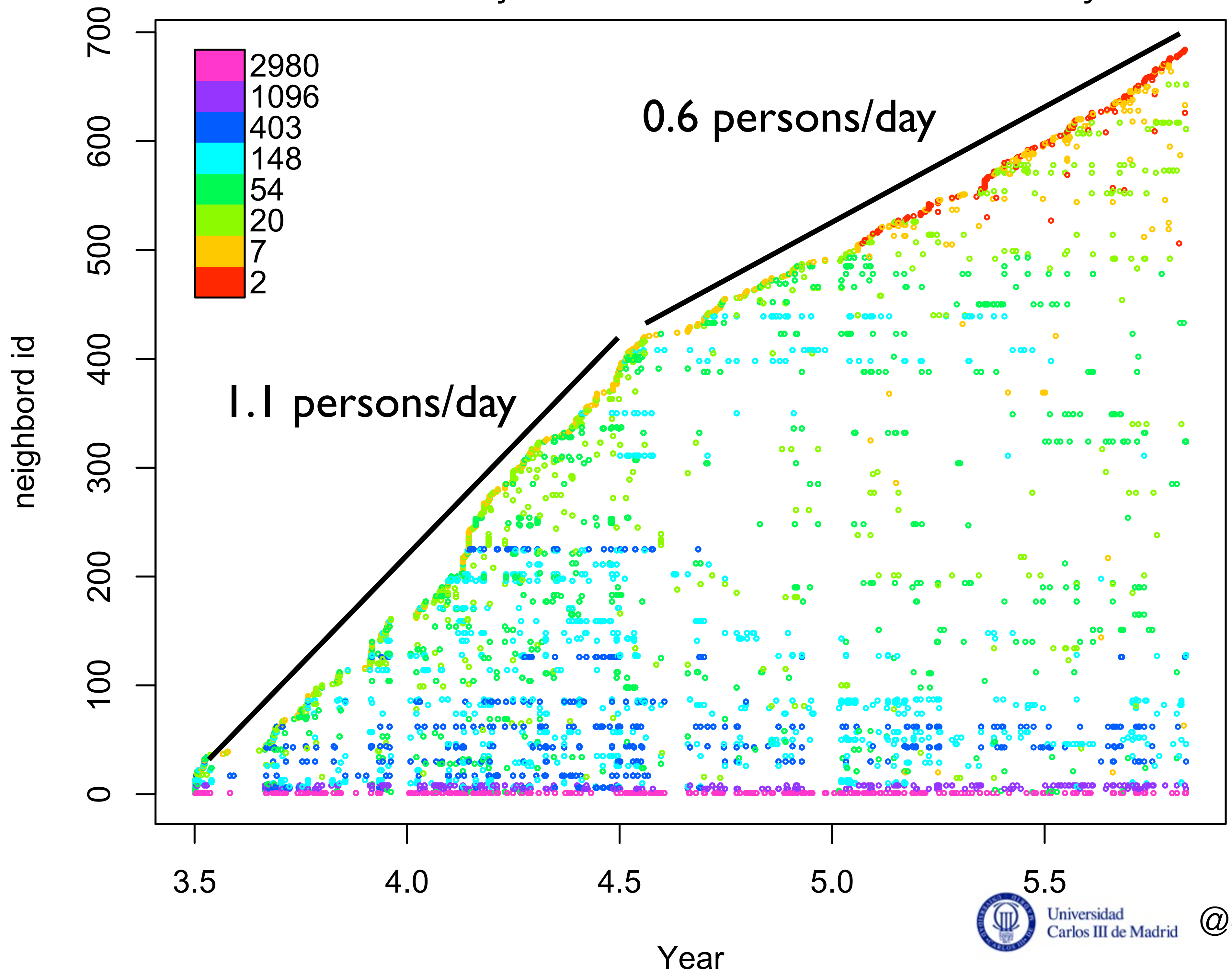
$$n_{\alpha,i}(t) \simeq \alpha_i t \qquad n_{\omega,i}(t) \simeq \omega_i t \qquad \alpha_i \simeq \omega_i$$

$$\boxed{\kappa_i(t) \simeq \kappa_i(0)}$$

Universidad Carlos III de Madrid @estebanmoro

Tie dynamics

- How are ties formed and destroyed? Linear tie formation/decay



Universidad Carlos III de Madrid

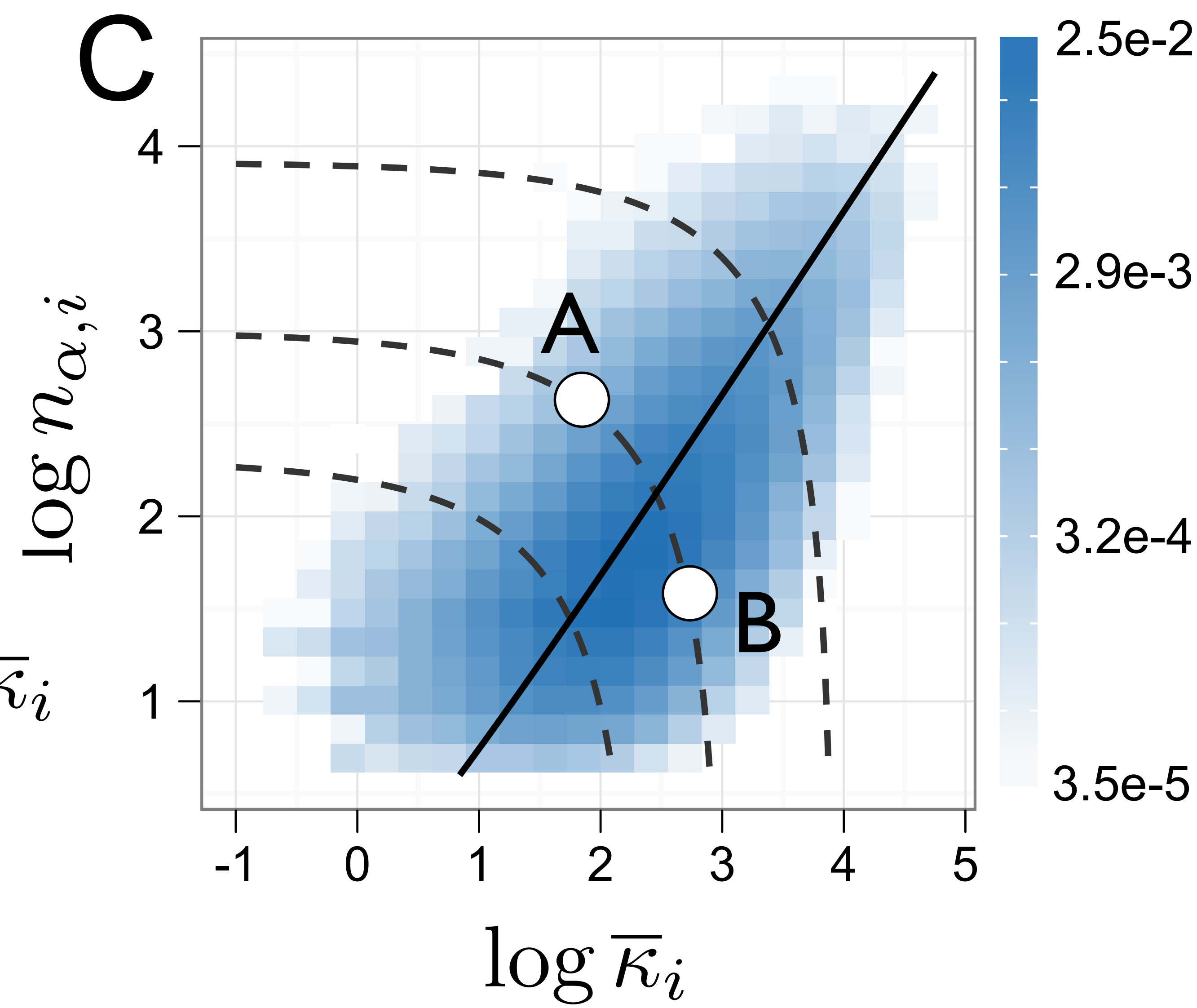
@estebanmoro

Tie dynamics

- How are ties formed and destroyed?
- Social capacity and activity are not independent

$$n_{\alpha,i} \propto \overline{\kappa}_i$$

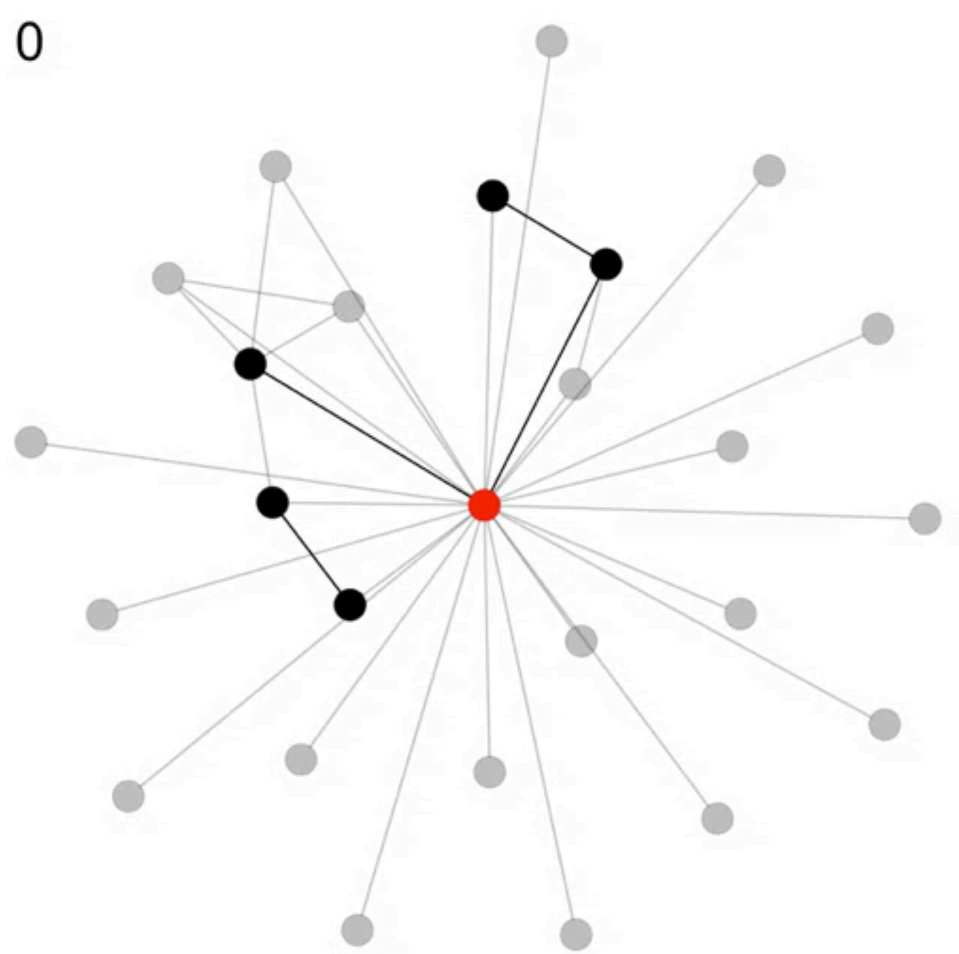
  - For a given k_i we have
    - Social explorers **(A)** $n_{\alpha,i} \gg \overline{\kappa}_i$
    - Balanced **(-)** $n_{\alpha,i} \simeq \overline{\kappa}_i$
    - Social keepers **(B)** $n_{\alpha,i} \ll \overline{\kappa}_i$



Miritello, Lara, Cebrián and EM
Scientific Reports 3, 1950 (2013)

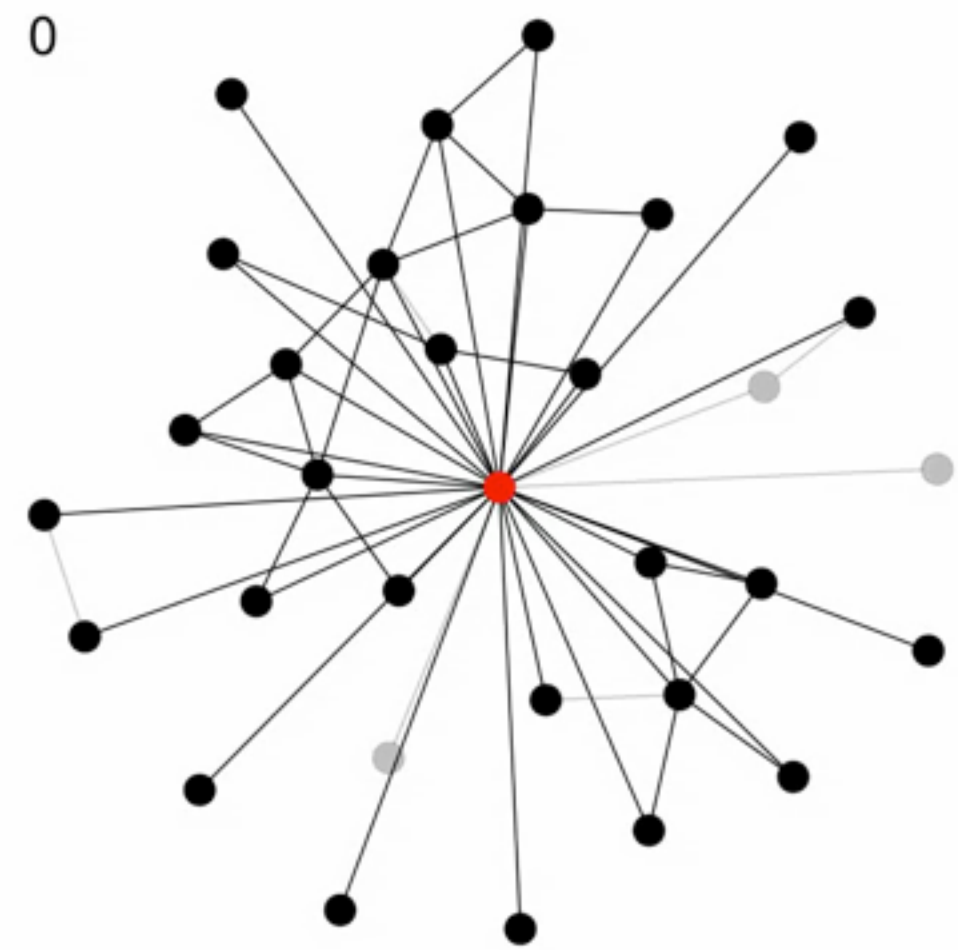
Tie dynamics

- H



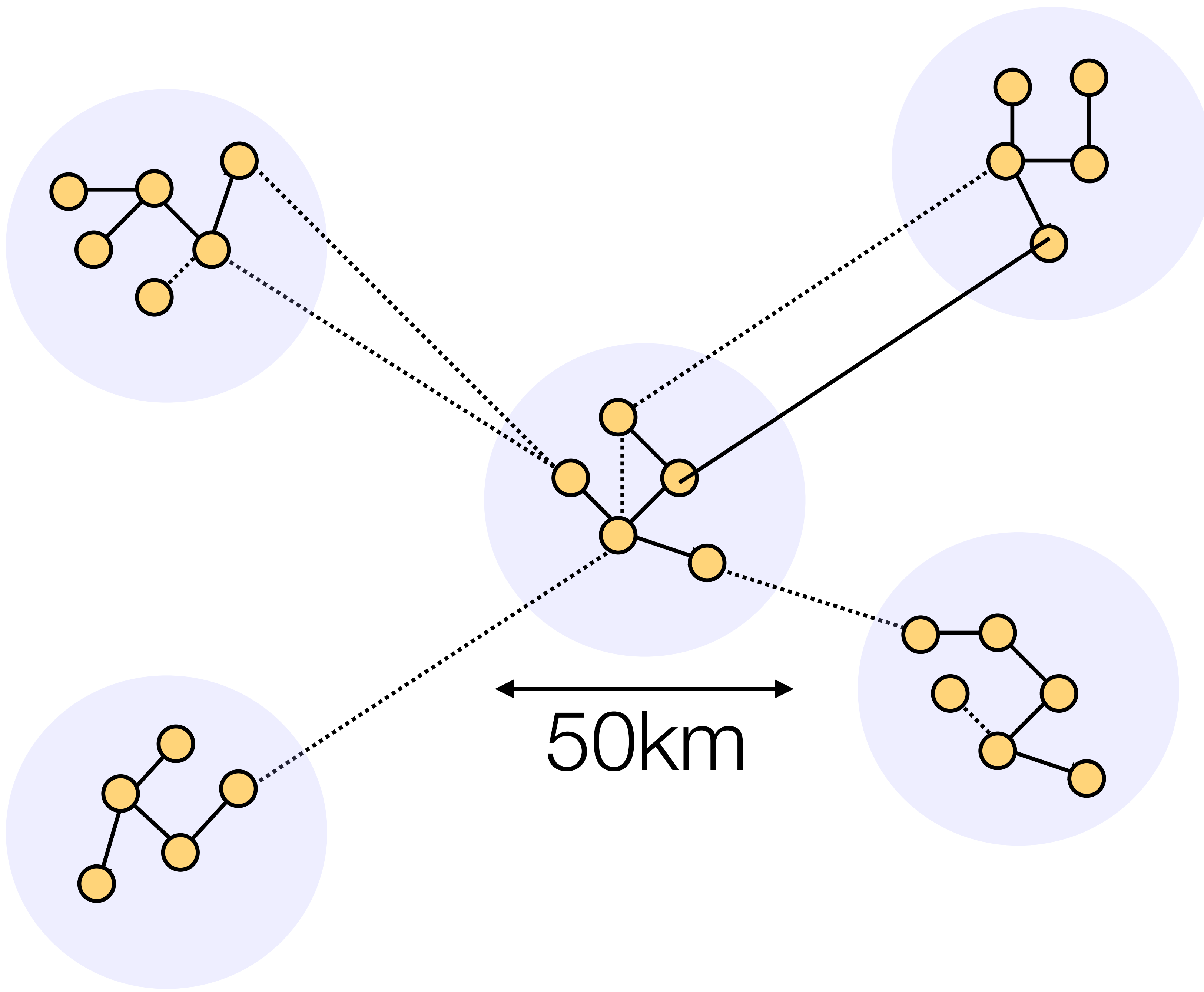
$$n_{\alpha,i} = 23, \overline{\kappa}_i = 4$$

Social explorer



$$n_{\alpha,i} = 3, \overline{\kappa}_i = 24$$

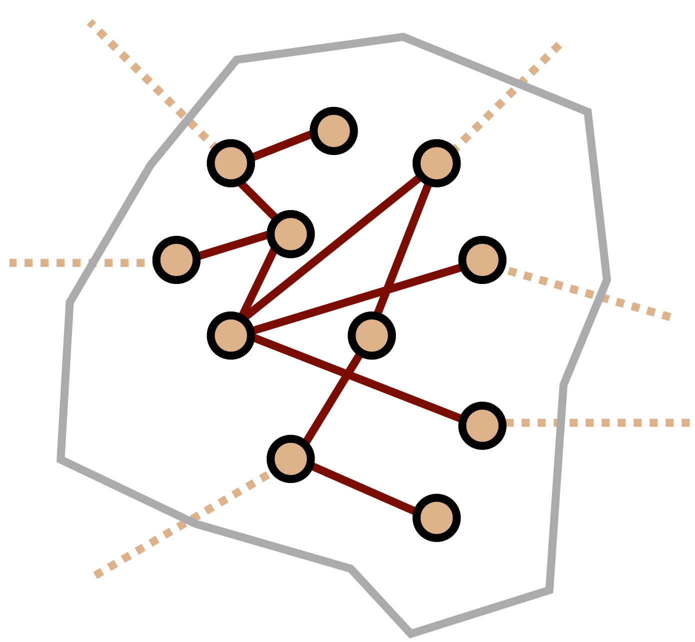
Social keeper

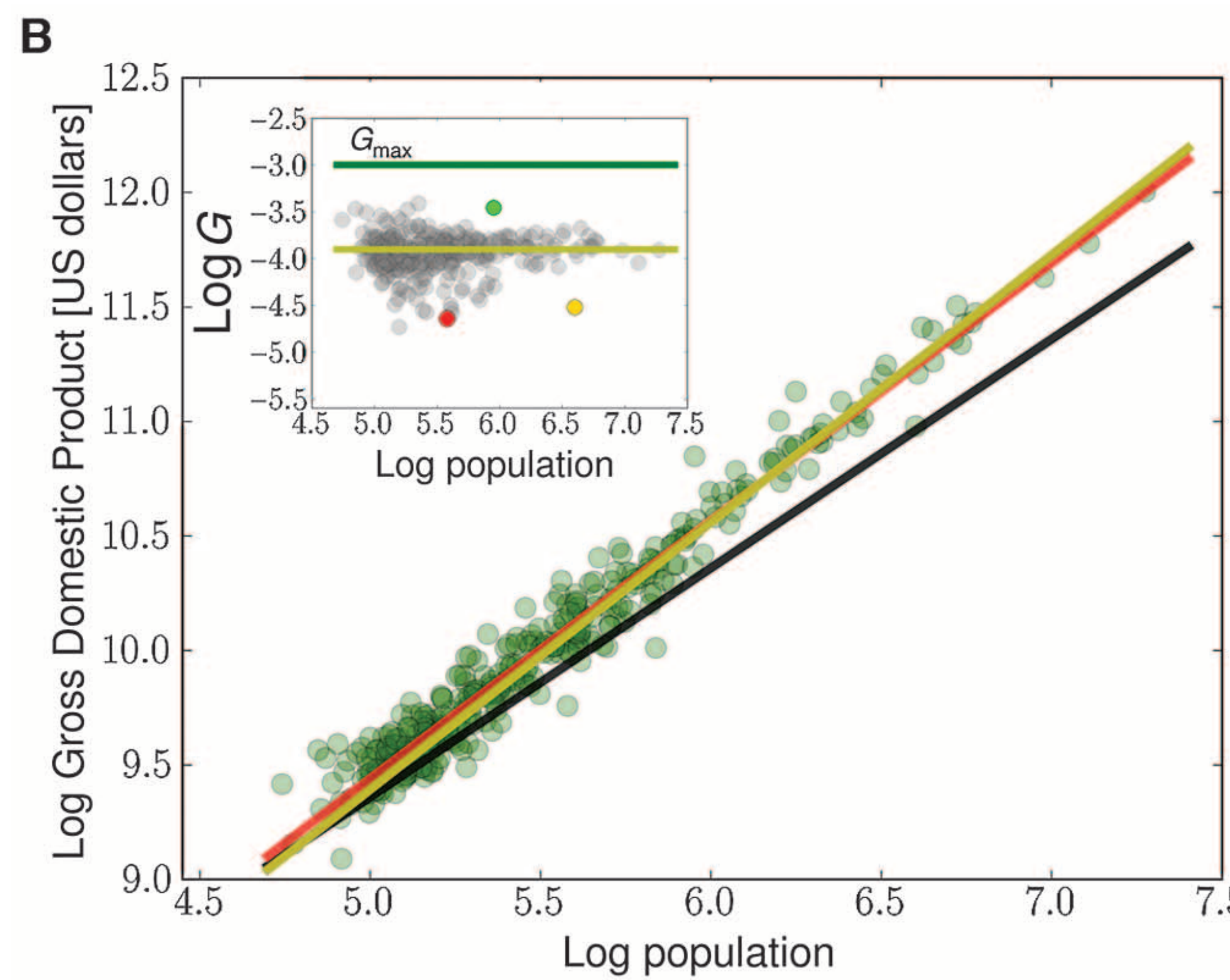


Geographical dynamics 2.3

Geography and network dynamics

- Does geography play a role in the dynamics of human communication?
- Known results:
  - City properties scale super-linearly with population
  - Including #links within the city!!



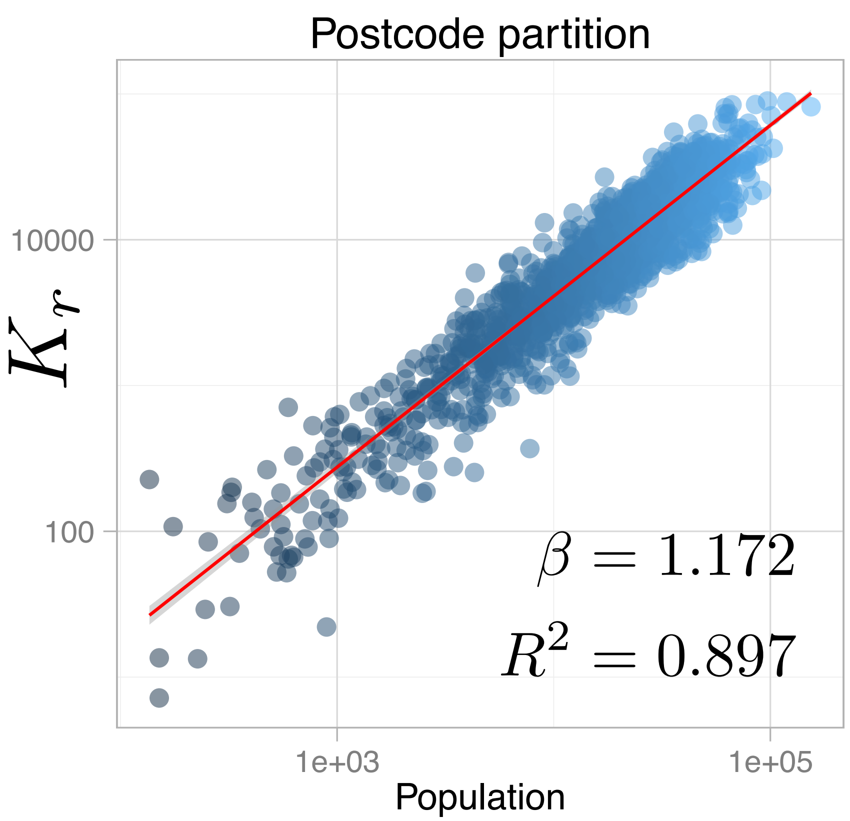


Bettencourt, L.M.A., 2013. The Origins of Scaling in Cities. Science, 340(6139), pp.1438–1441.

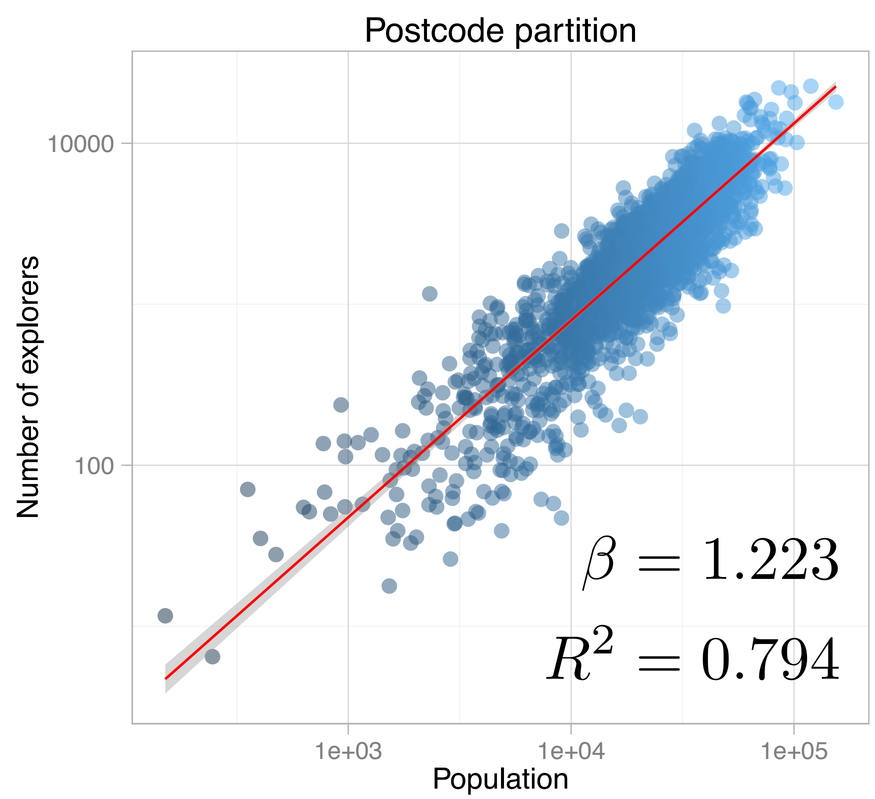
Geography and network dynamics

- Dynamics also scale super-linearly with the size of the city
    - Bigger cities have more dynamical networks

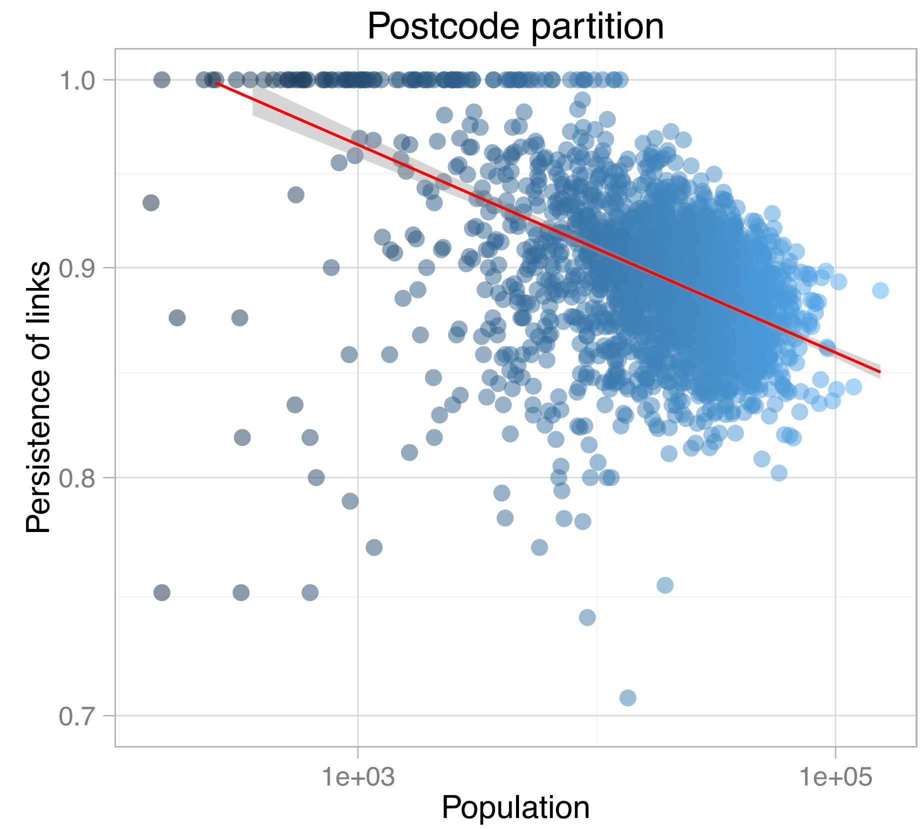
Number of links



Number of explorers



Tie persistence

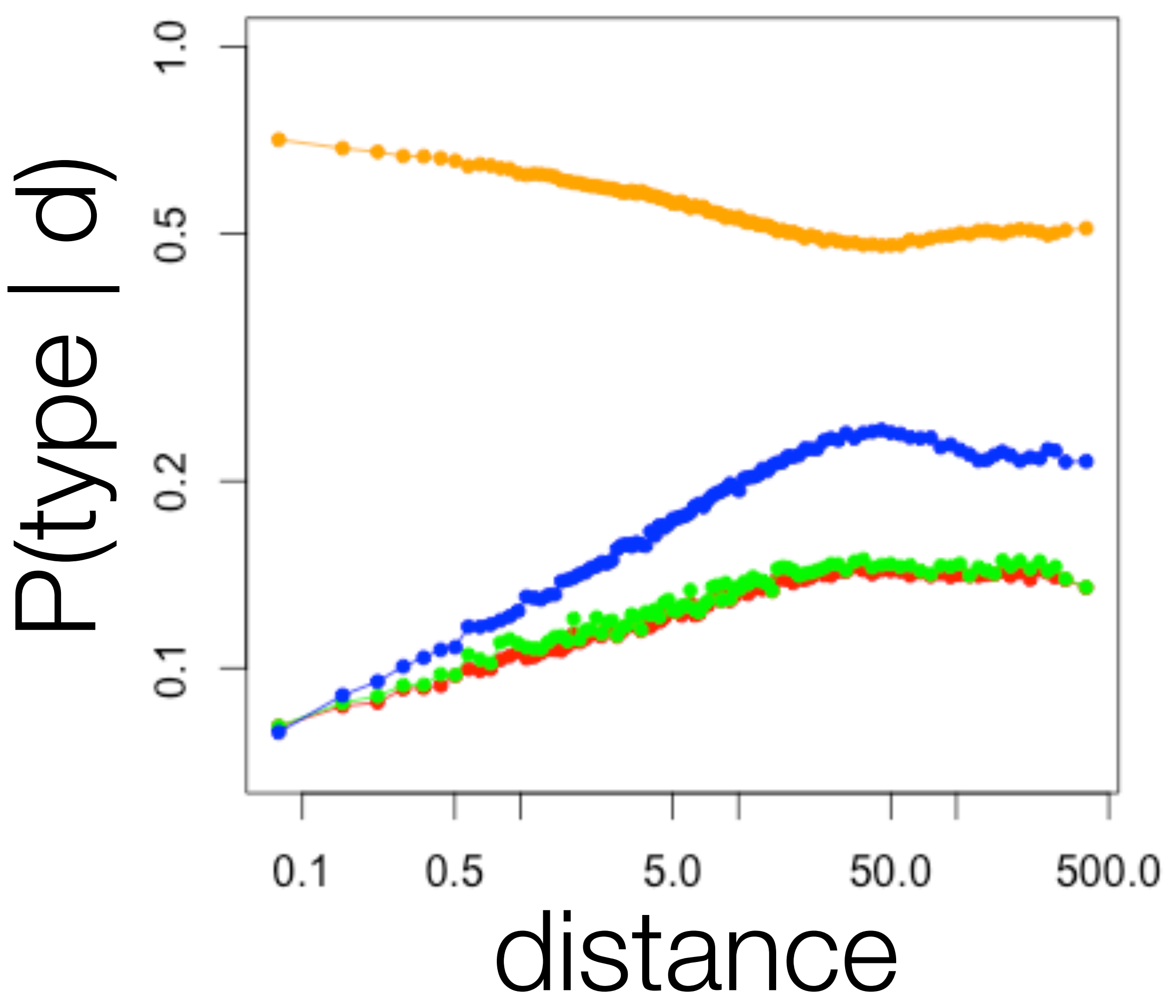


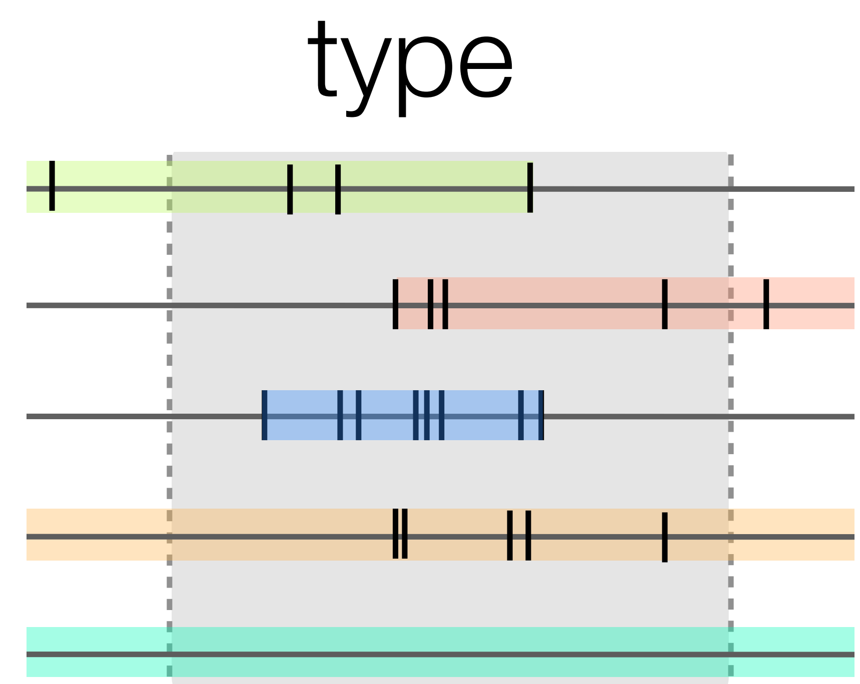
$$Persistence = \frac{|E(t+T) \cap E(t)|}{|E(t)|}$$

@estebanmoro

Geography and network dynamics

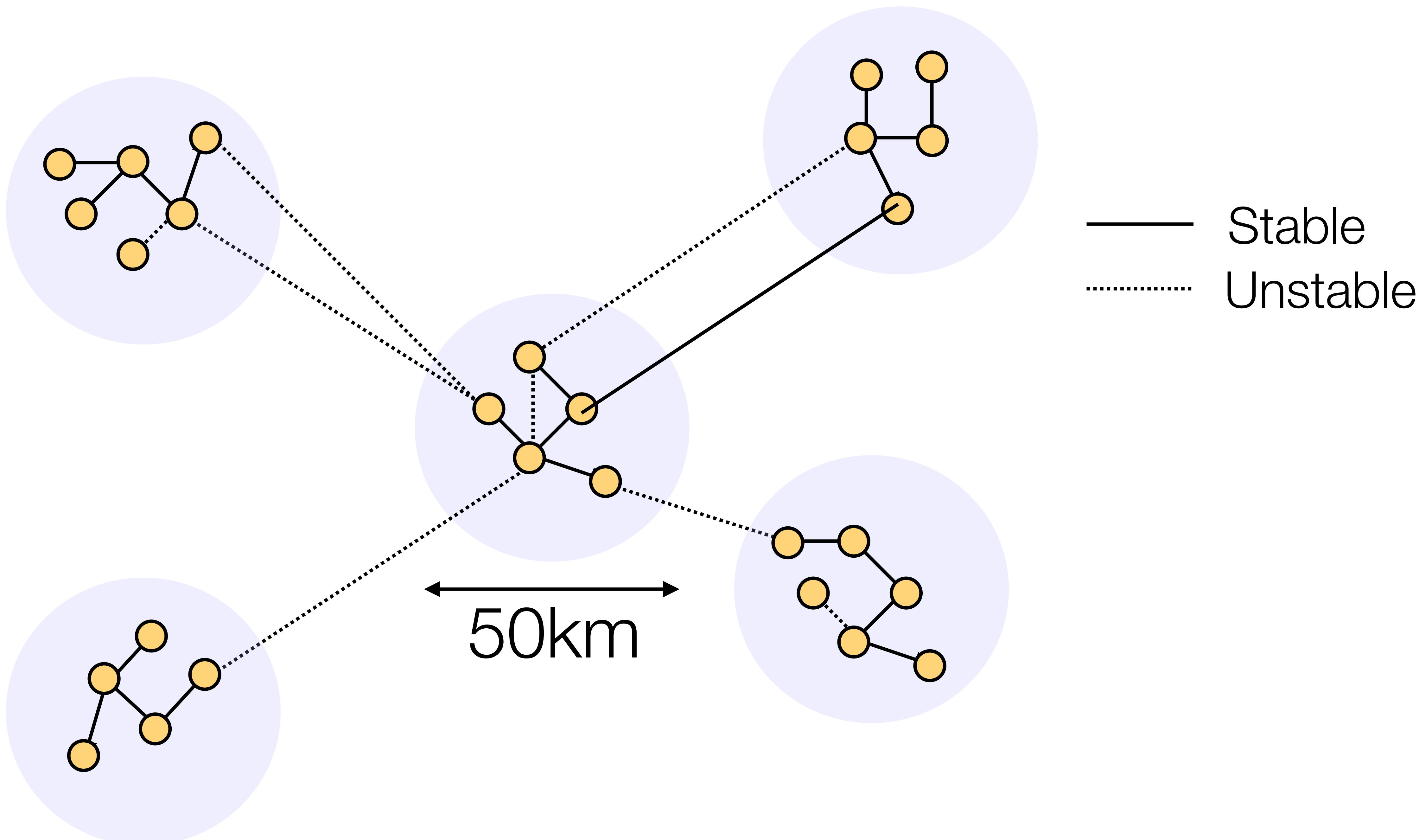
- Tie Dynamics also depend on the distance between people:
  - At small distances ties are very stable
  - At large distances ties are very unstable





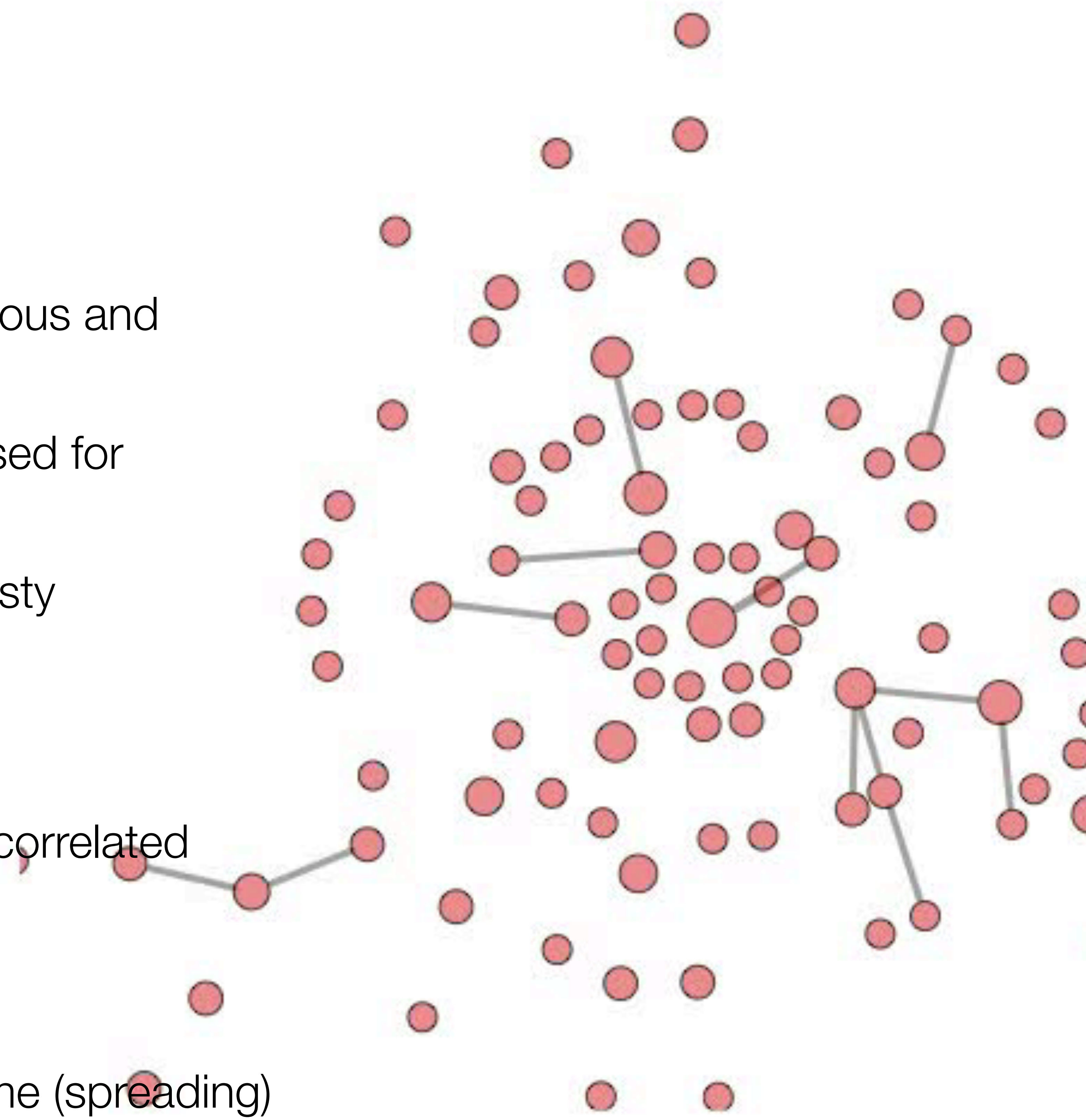
There is a geographical scale for dynamics
d ~ 50km

The geographical picture of temporal networks



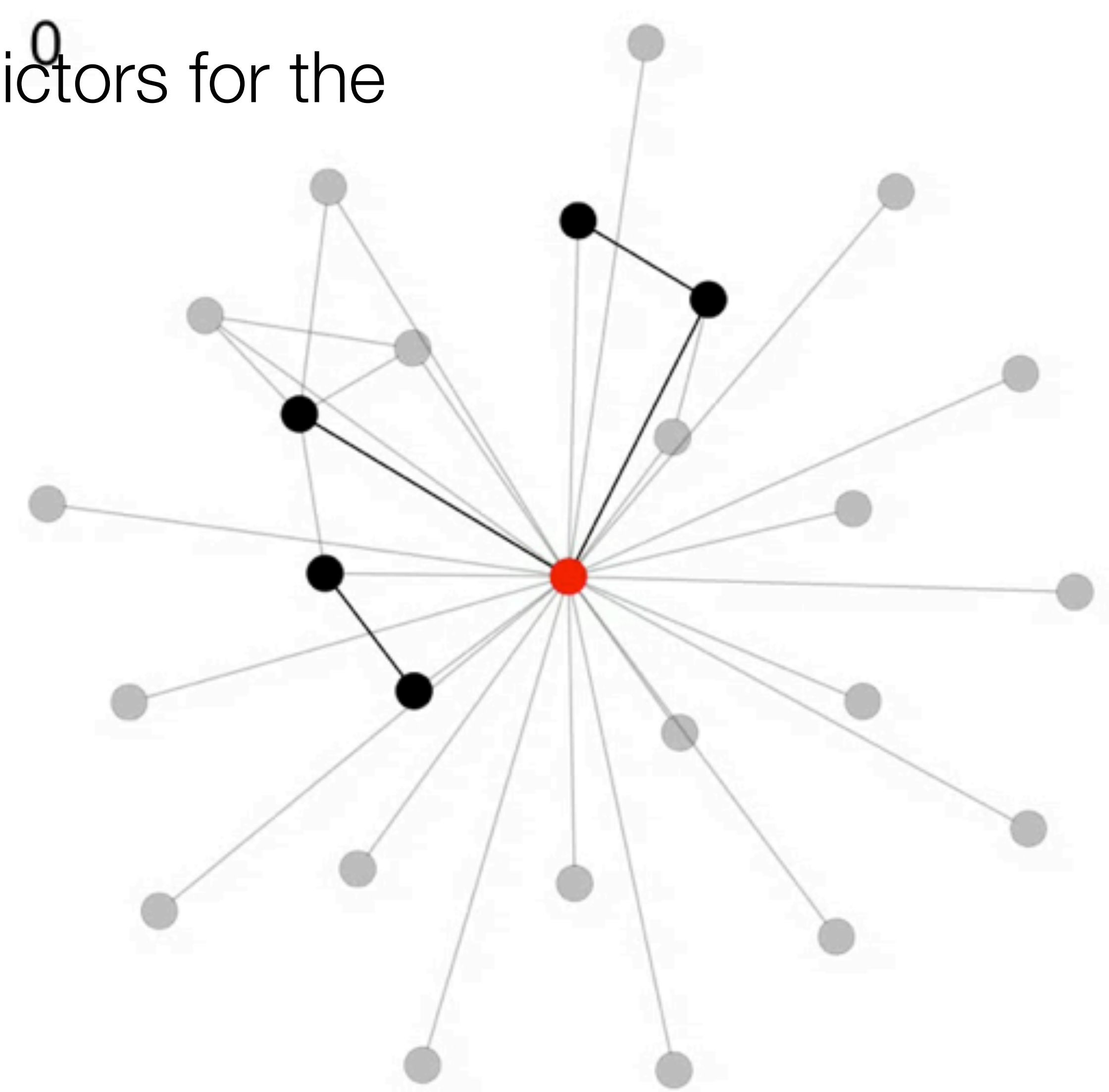
Tie dynamics

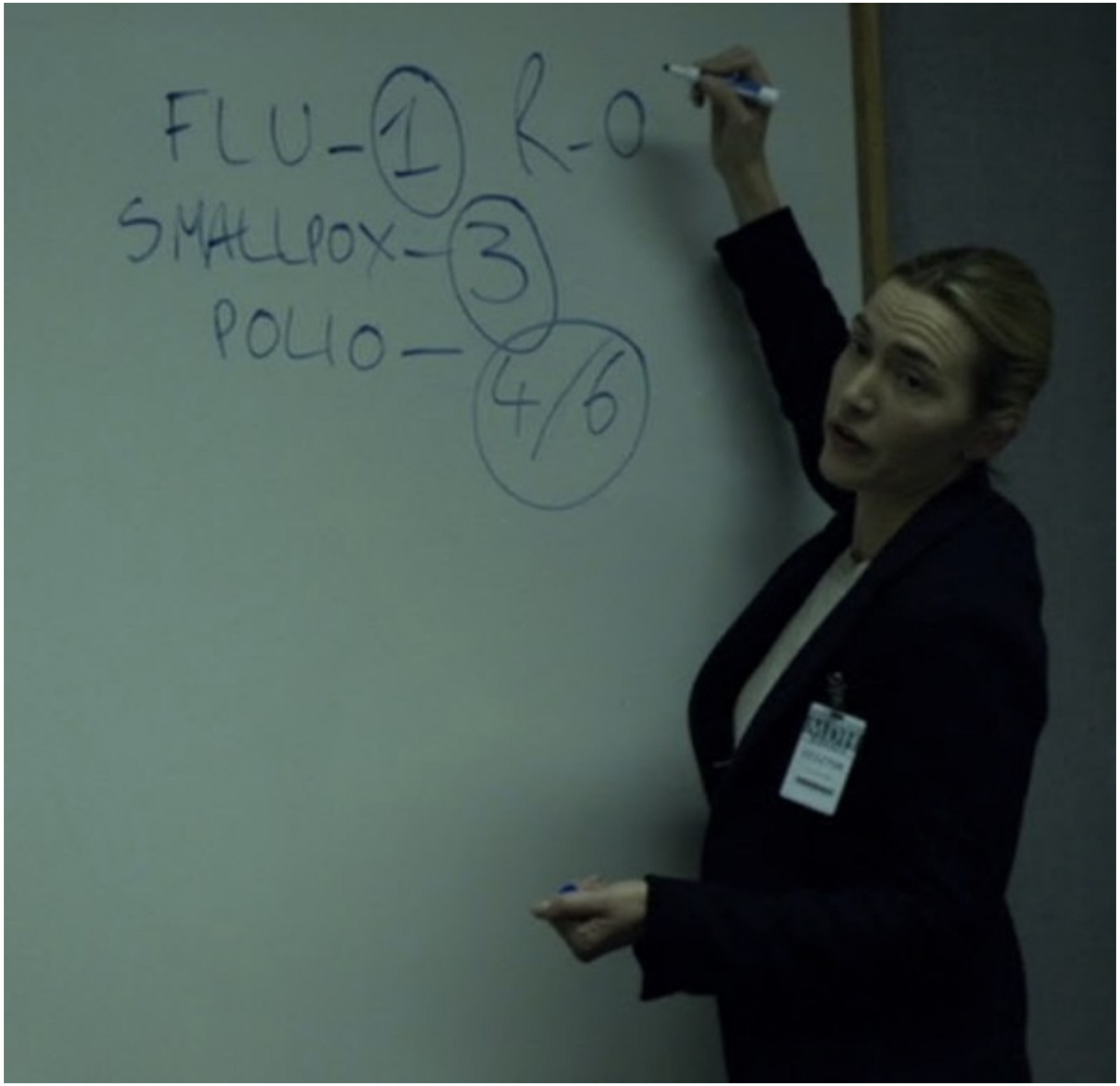
- Wrap-up
  - Individual activity is heterogeneous and persistent
  - Different parts of the day are used for different social tasks
  - Activity within a single tie is bursty
    - P(dt) is a heavy tailed
    - Bursts are correlated
  - Activity across adjacent ties is correlated
    - Two adjacent ties
      - Group conversations
      - Impact on the waiting time (spreading)



Tie dynamics

- *Triadic closure*, *reciprocity* are *predictors* for the formation of a link
    - *Embeddedness*, *reciprocity* are predictors for the persistence of a link
- Tie formation/decay is *bursty*
- Tie formation/decay *strategy*:
    - Heterogeneous
    - Linear in time
    - Social explorers / social keepers
- *Geography*:
    - Larger cities are more dynamical
    - At larger distances links are more unstable

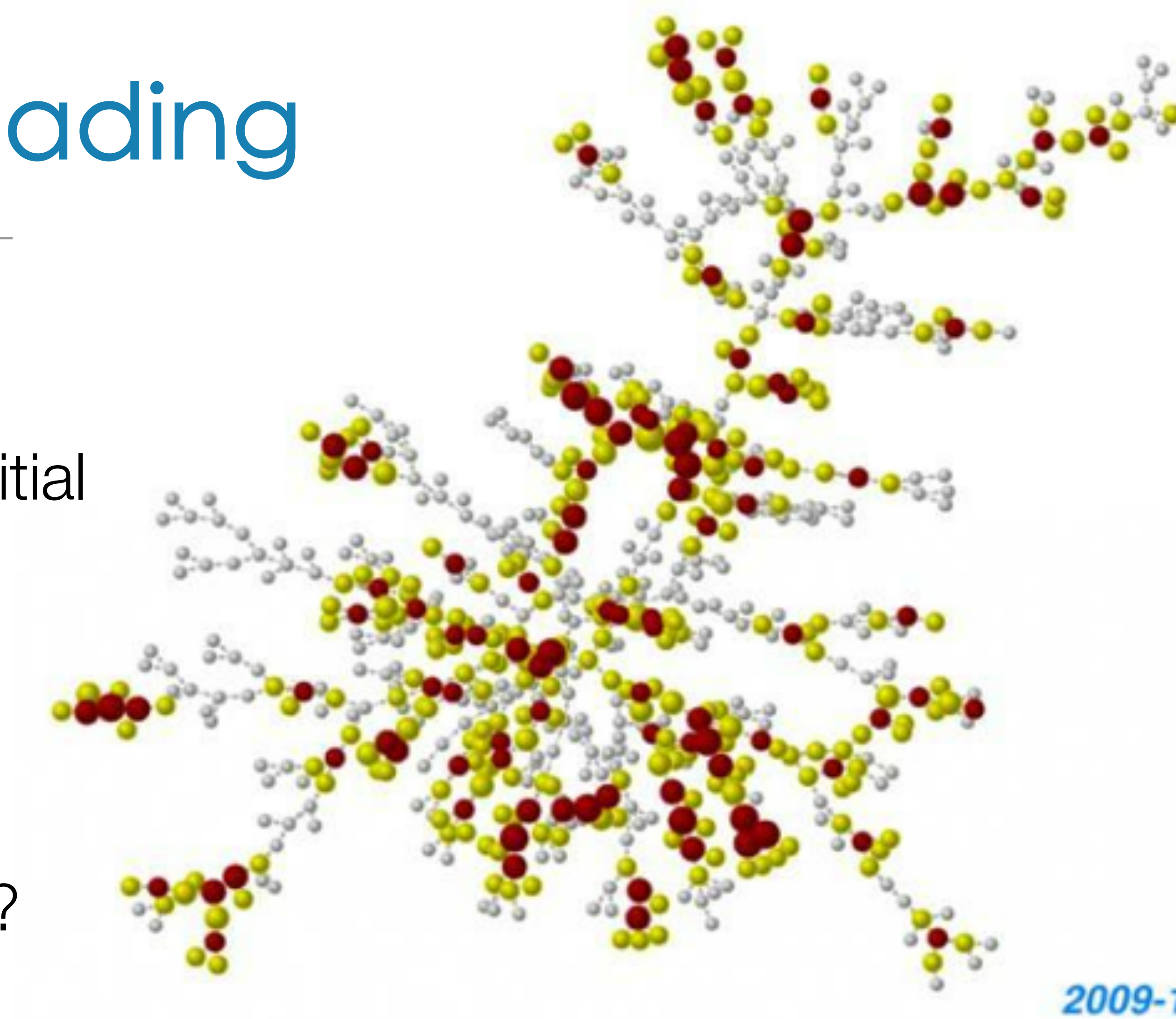




Impact on diffusion processes

3

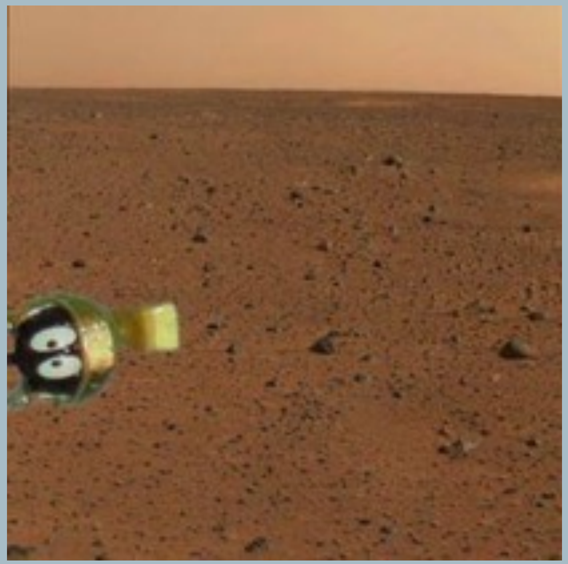
Relevant question in spreading



Christakis & Fowler ’10

- **Reach**
  - How many people are infected from a initial spreader?
- **Time**
  - How long does it take to infect them?
  - Early detection of an outbreak, possible?
- **Optimization**
  - How do we choose a given a number N of initial spreaders, so that reach is maximize in a given time? What is the optimal N for a given cost?
  - How do we choose a given number of immune people so that reach of the disease is minimized? (resiliance of networks)
  - How do we choose sensors to detect propagation?

Information spreading as cascades



http://www.facebookstories.com/stories/2200/data-visualization-photo-sharing-explosions

2012.08.05 23:05

Simple model for spreading

- SI / SIR / SIS models (Kermack & McKendrick ’27)
  - S: suceptible (non infected)
  - I: Infected
  - R: resiliant
  - S + I + R = N

$$S \xrightarrow{\lambda} I \xrightarrow{\gamma} R$$

$$\frac{dS}{dt} = -\lambda IS \qquad R_0 = N\frac{\lambda}{\gamma}$$

$$\frac{dI}{dt} = \lambda IS - \gamma I$$

$$\frac{dR}{dt} = \gamma I \qquad \frac{dI}{dt} = \gamma(R_0 S/N - 1)I$$

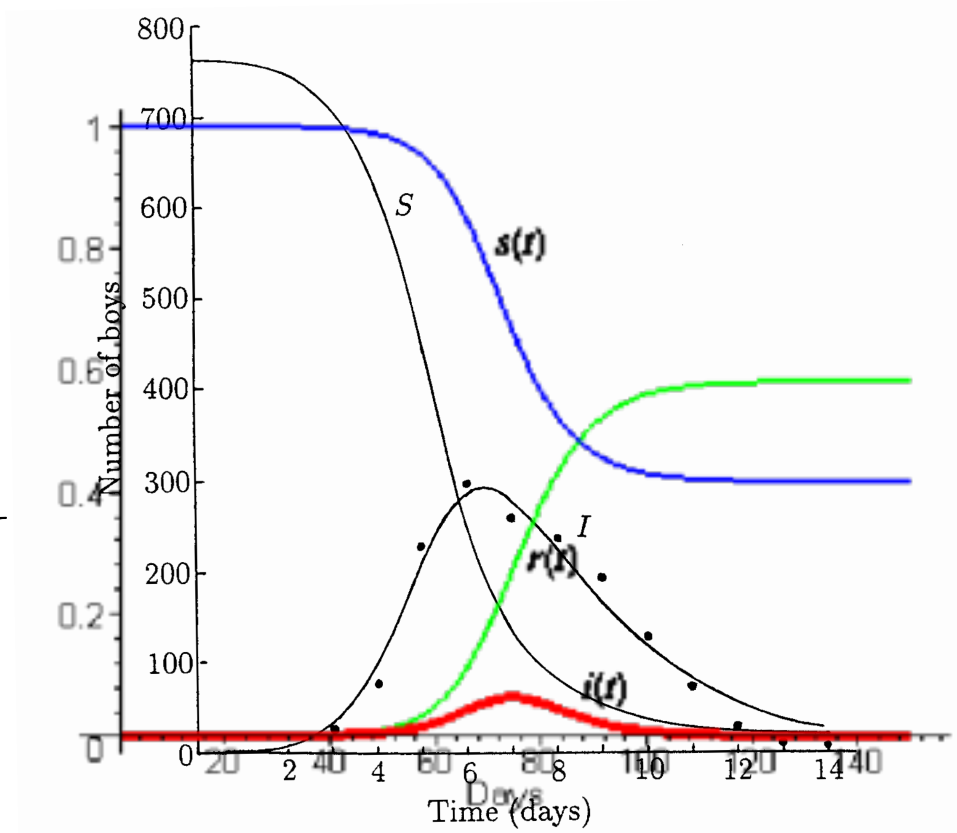
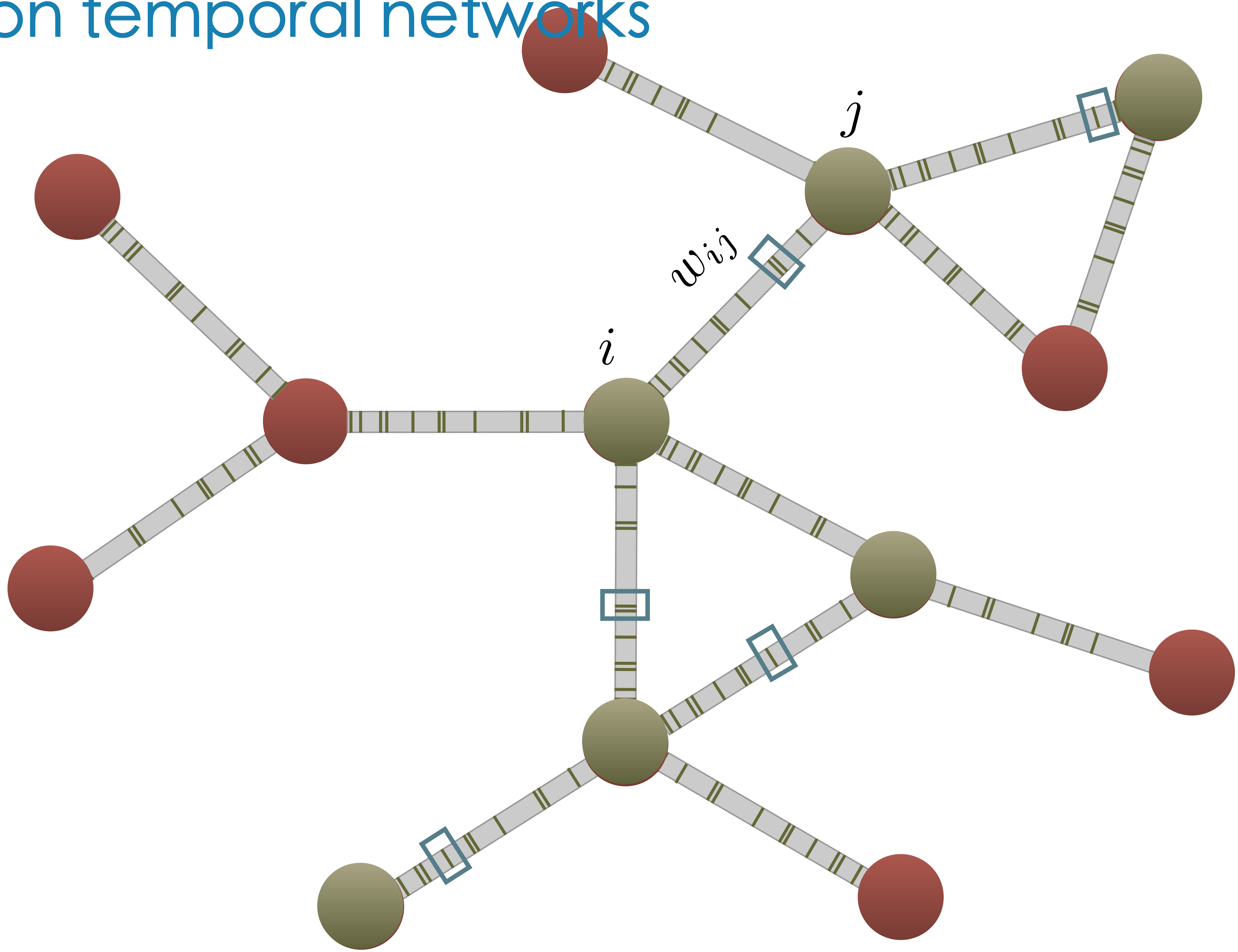


Fig. 19.3. Influenza epidemic data (•) for a boys boarding school as reported in British Me
Journal, 4th March 1978. The continuous curves for the infectives (I) and susceptibles (S)
obtained from a best fit numerical solution of the *SIR* system (19.1)–(19.3): parameter v
$N = 763$, $S_0 = 762$, $I_0 = 1$, $\rho = 202$, $r = 2.18 \times 10^{-3}$/day. The conditions for an epiden
occur, namely $S_0 > \rho$ is clearly stisfied and the epidemic is severe since R/ρ is not small

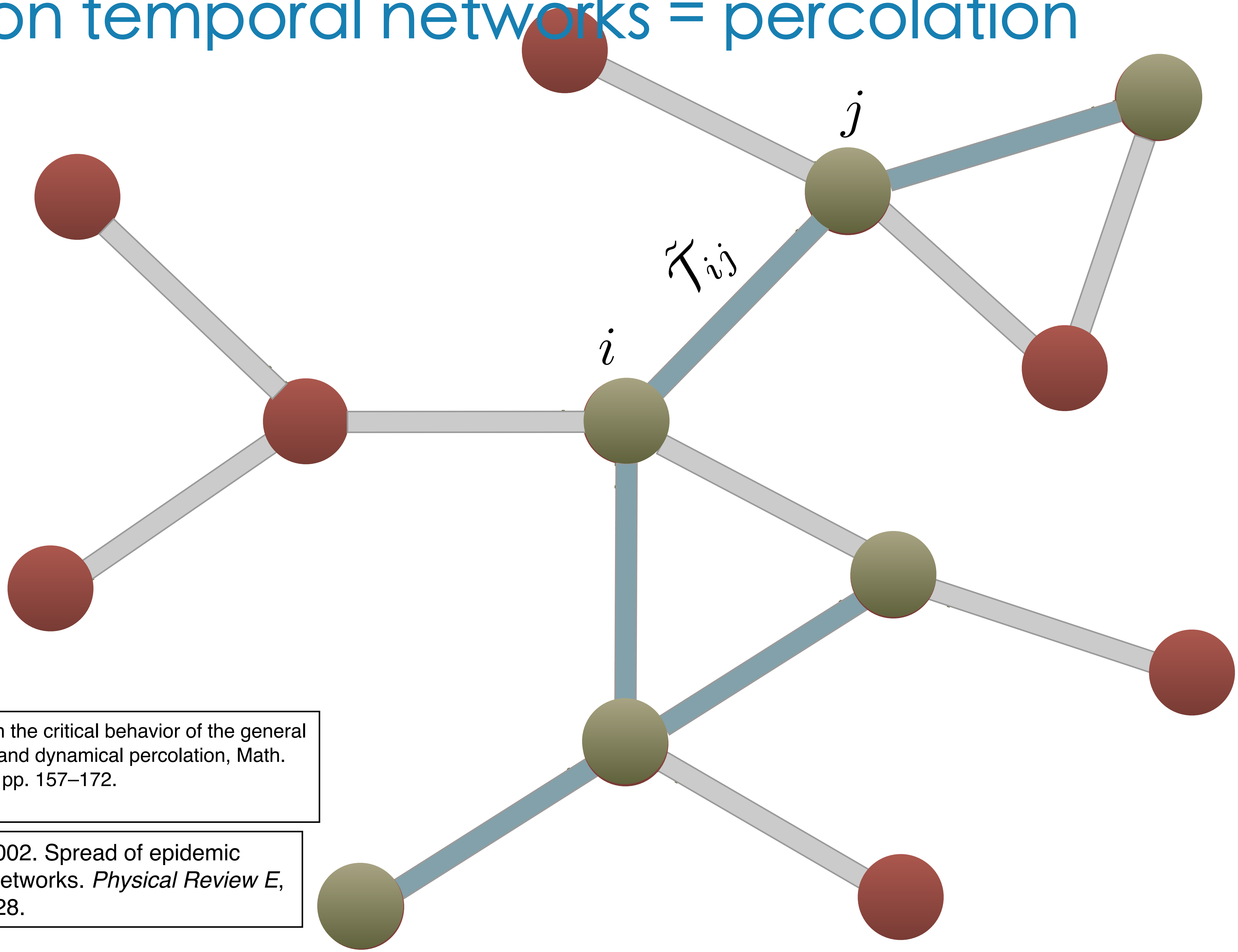
- R_0: basic reproductive number

$$R_0 > N/S(0) \Rightarrow dI/dt > 0$$
$$R_0 < N/S(0) \Rightarrow dI/dt < 0$$

SIR on temporal networks



SIR on temporal networks = percolation

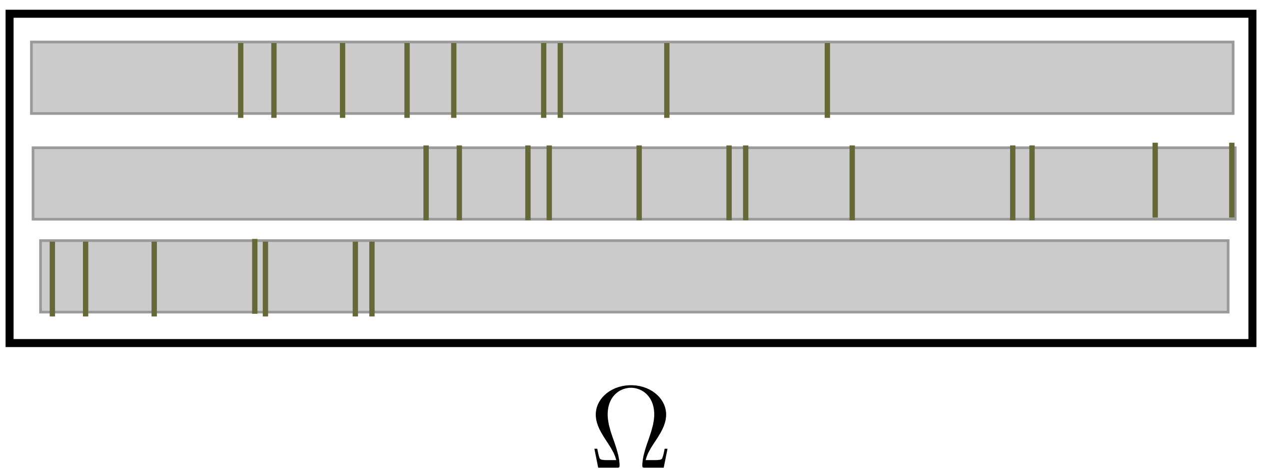


P. Grassberger, On the critical behavior of the general epidemic process and dynamical percolation, Math. Biosci., 63 (1983), pp. 157–172.

Newman, M., 2002. Spread of epidemic disease on networks. *Physical Review E*, 66(1), p.16128.

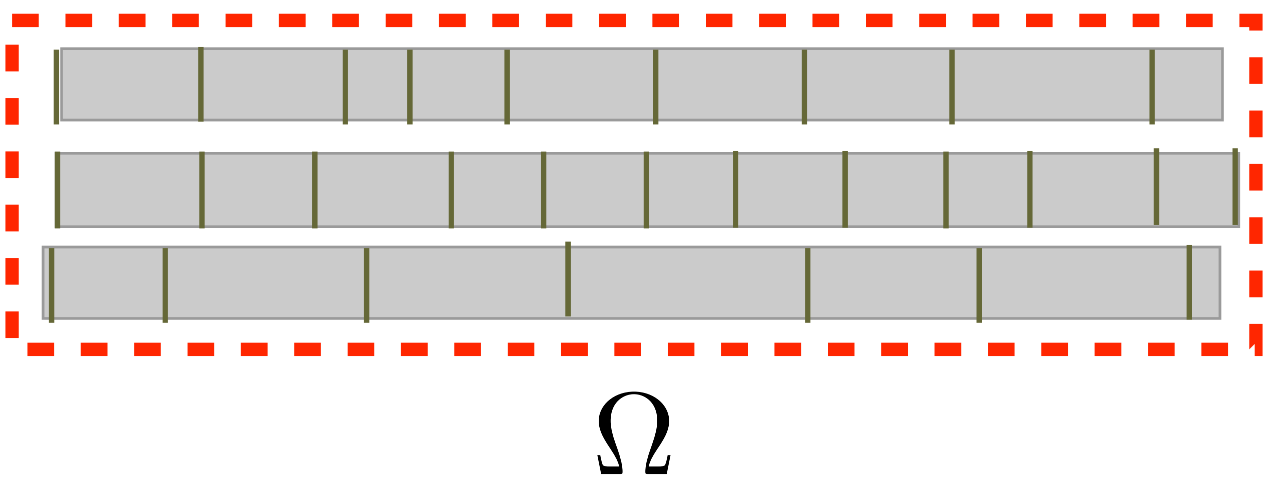
Comparison with null models

- Real data



P(dt) heavy tailed
Correlated bursts
Correlated tie activity
Temporal motifs
Tie dynamics

- Time Shuffled data



P(dt) exponential
Uncorrelated bursts
Uncorrelated tie activity
No temporal motifs
No tie dynamics

Data-driven simulations

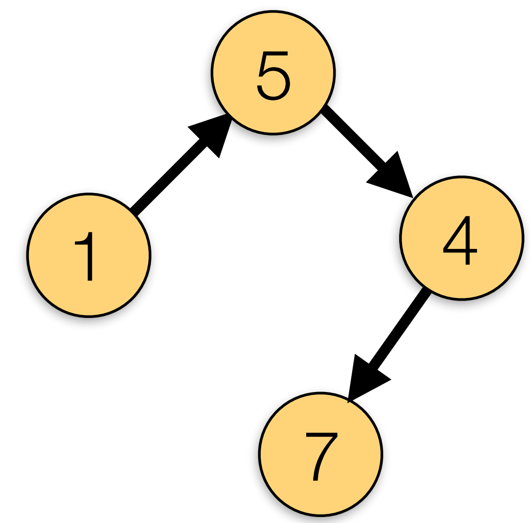
- SIR model on real contact data

v_i, v_j, t
1,5,412
2,3,523
5,4,631
3,7,782
1,2,921
2,7,999

Select seed
+
infect in each contact with probability

v_i, v_j, t
1,5,412
2,3,523
5,4,631
3,7,782
1,2,921
4,7,999

=



Data-driven simulations

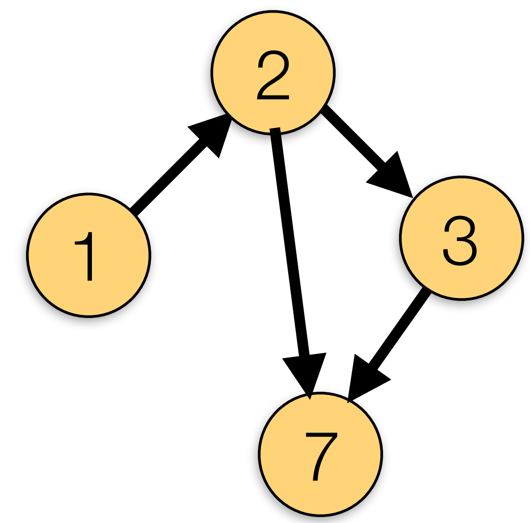
- SIR model on shuffled contact data

v_i, v_j, t		v_i, v_j, t
1,5,412		1,2,412
2,3,523		2,3,523
5,4,631	→	1,5,631
3,7,782		2,7,782
1,2,921		3,7,921
2,7,999		5,4,999
Real data		Shuffled data

Select seed
+
infect in each contact with probability

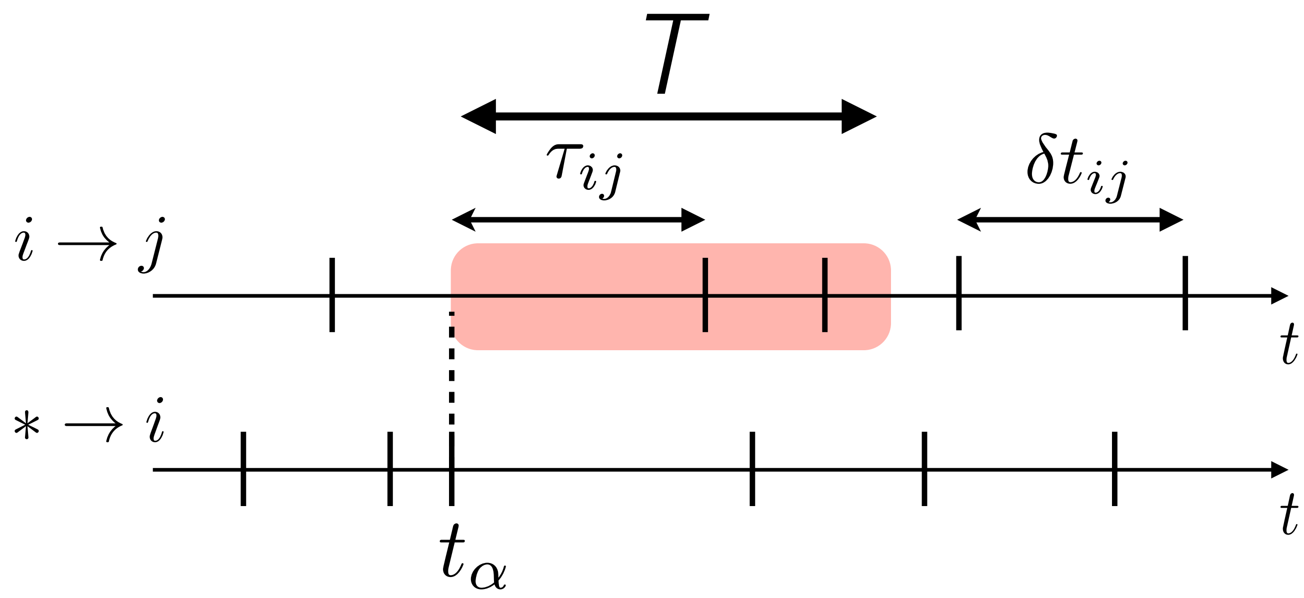
v_i, v_j, t
1,2,412
2,3,523
1,5,631
2,7,782
3,7,921
5,4,999

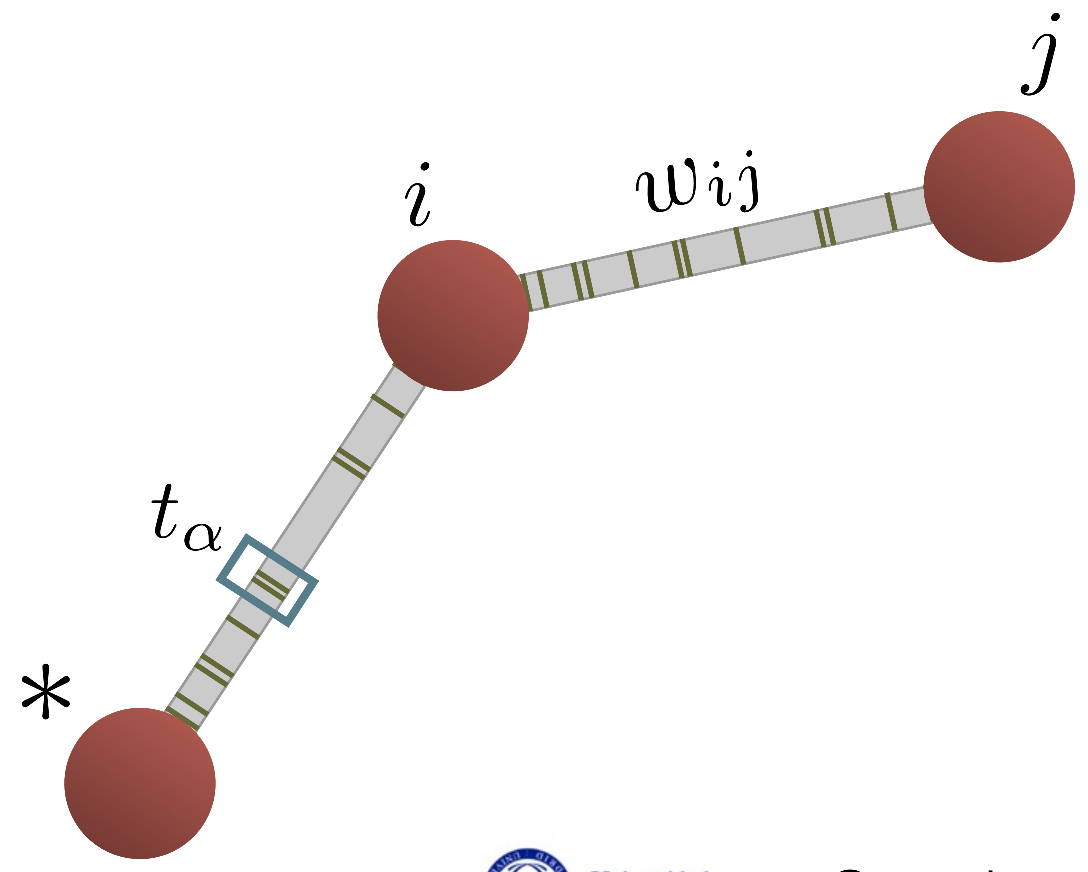
=



Spreading including tie activity

- Spreading (SIR) on contact networks
- Hypothesis:
  - In every contact there is a probability λ to infect
  - Nodes only remain infected for a time “$T \simeq 1/\gamma$“
- **Transmissibility**: probability that *i* infects j after being infected at t_α



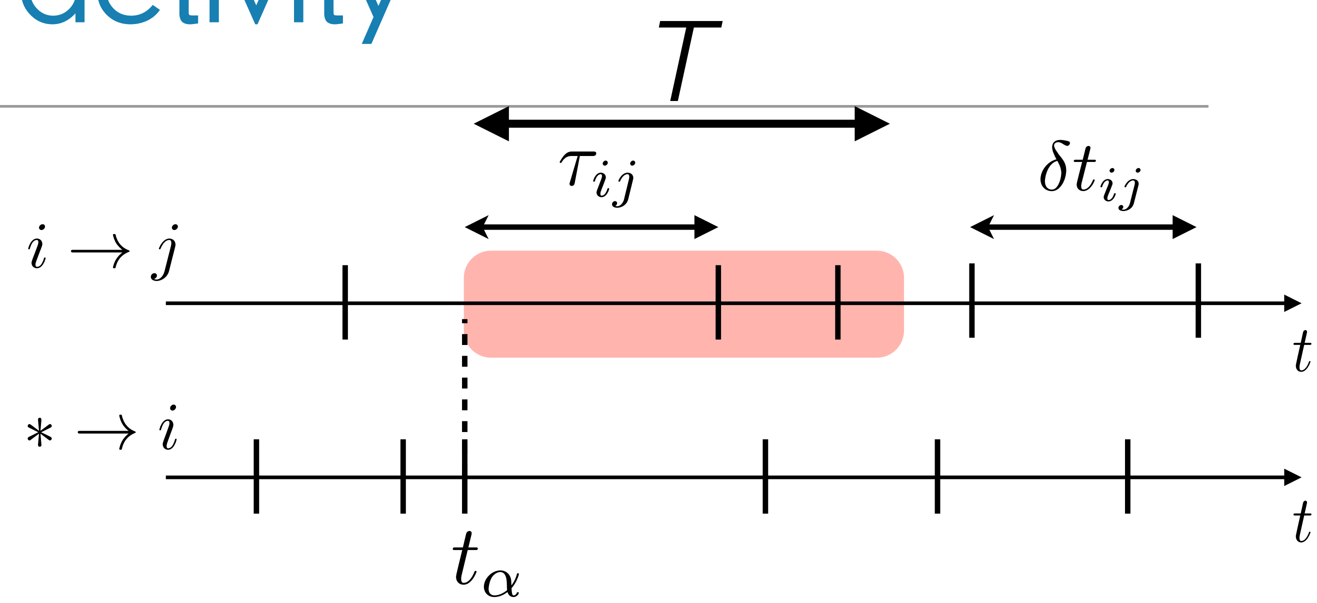


Miritello, G., Moro, E. & Lara, R., 2011. Dynamical strength of social ties in information spreading. *Physical Review E*, 83(4), p.045102.

Spreading including tie activity

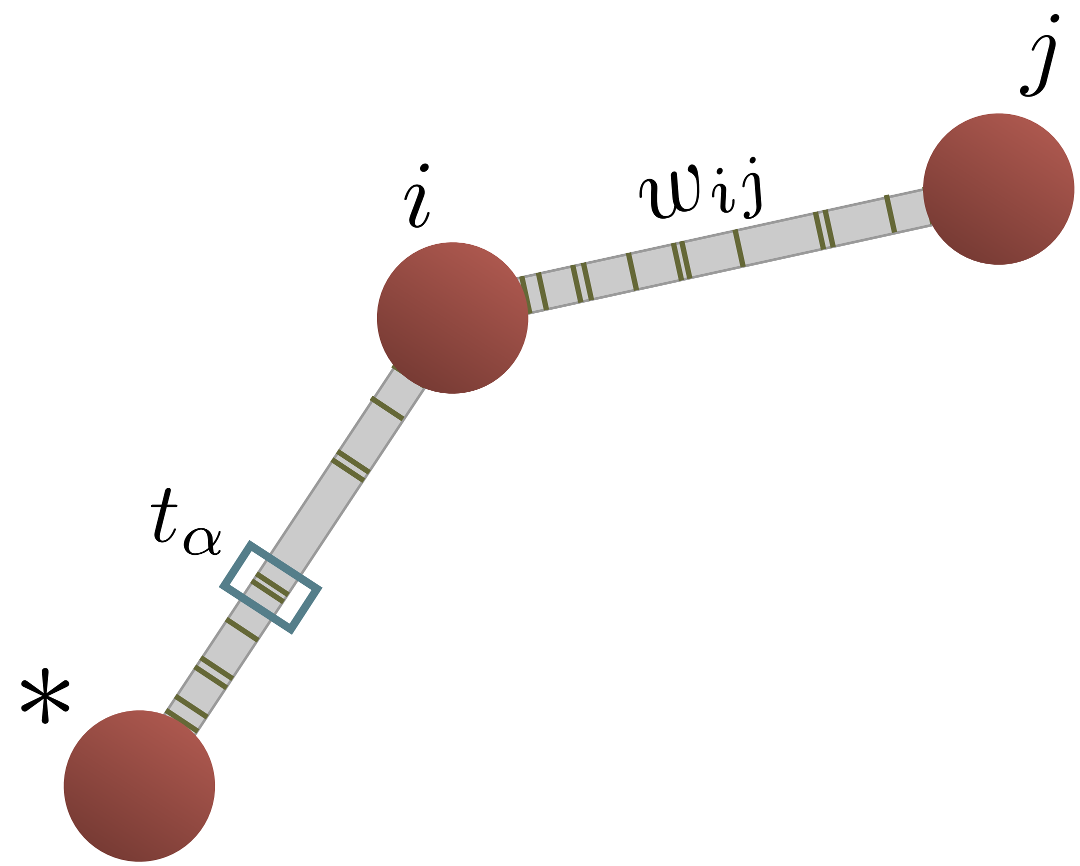
- **Transmissibility**:

$$\mathcal{T}_{ij} = 1 - (1 - \lambda)^{n_{ij}(t_\alpha)}$$



- where

$n_{ij}(t_\alpha)$ = number of $i \to j$ events in the time interval $[t_\alpha, t_\alpha + T]$



Spreading including tie activity

- Assuming $* \to i$ contacts are independent and equally probable in the observation period

$$\mathcal{T}_{ij}[\lambda, T] = \langle 1 - (1-\lambda)^{n_{ij}(t_\alpha)} \rangle_\alpha .$$

$$\mathcal{T}_{ij}[\lambda, T] = \sum_{n=0}^{\infty} P(n_{ij} = n; T)[1 - (1-\lambda)^n]$$

Probability of having *n* interactions between *i* and *j* in a time interval of length T

Spreading including tie activity

$$\mathcal{T}_{ij}[\lambda, T] = \sum_{n=0}^{\infty} P(n_{ij} = n; T)[1 - (1 - \lambda)^n]$$

- General process. Approximations
  - If $\lambda \ll 1 \Rightarrow 1 - (1 - \lambda)^n \simeq \lambda n$

$$\mathcal{T}_{ij} \simeq \lambda \boxed{\langle n_{ij} \rangle_{t_o}}$$

  - If $\lambda \simeq 1 \Rightarrow 1 - (1 - \lambda)^n \simeq 1 \text{ for } n > 0$

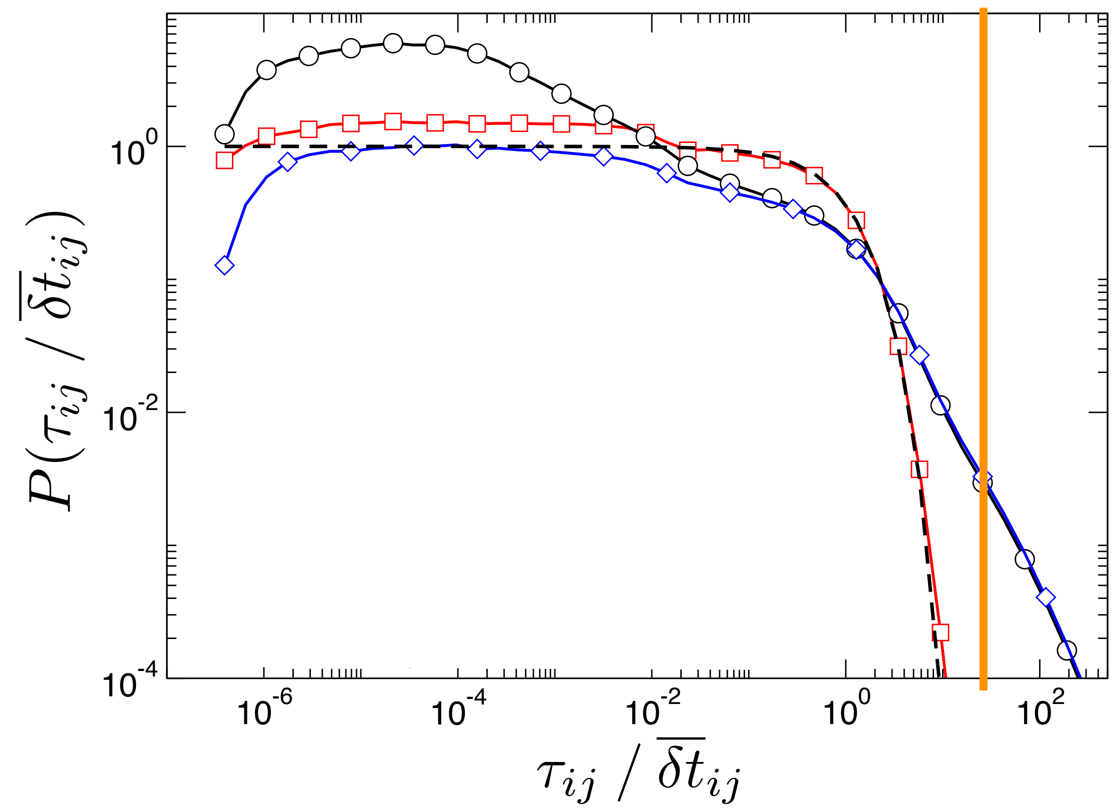
$$\mathcal{T}_{ij} \simeq 1 - \boxed{P_{ij}^0}$$

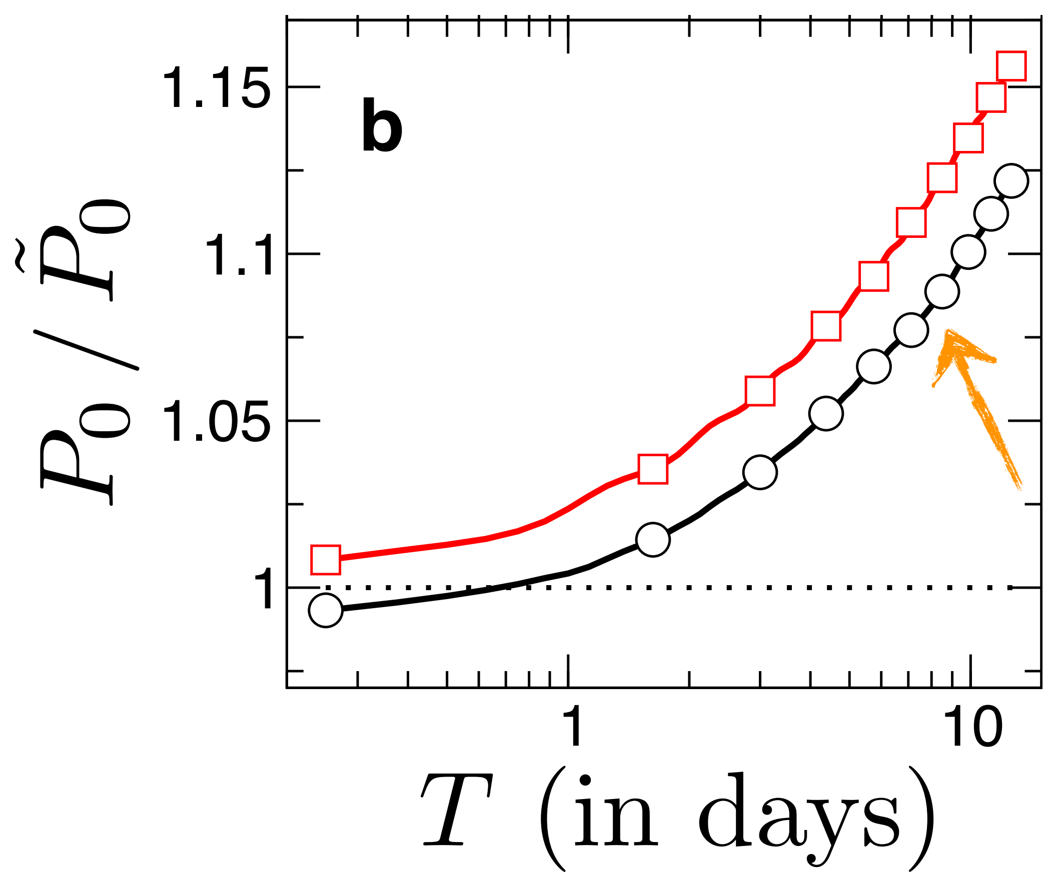
where $P_{ij}^0 = P(n_{ij} = 0; T) = \int_T^{\infty} P(\tau_{ij}) d\tau_{ij}$

Spreading including tie activity

- $\lambda \simeq 1$
- Probability of no event P^0_{ij}

$$P^0_{ij} = P(n_{ij} = 0; T) = \int_T^\infty P(\tau_{ij}) d\tau_{ij}$$





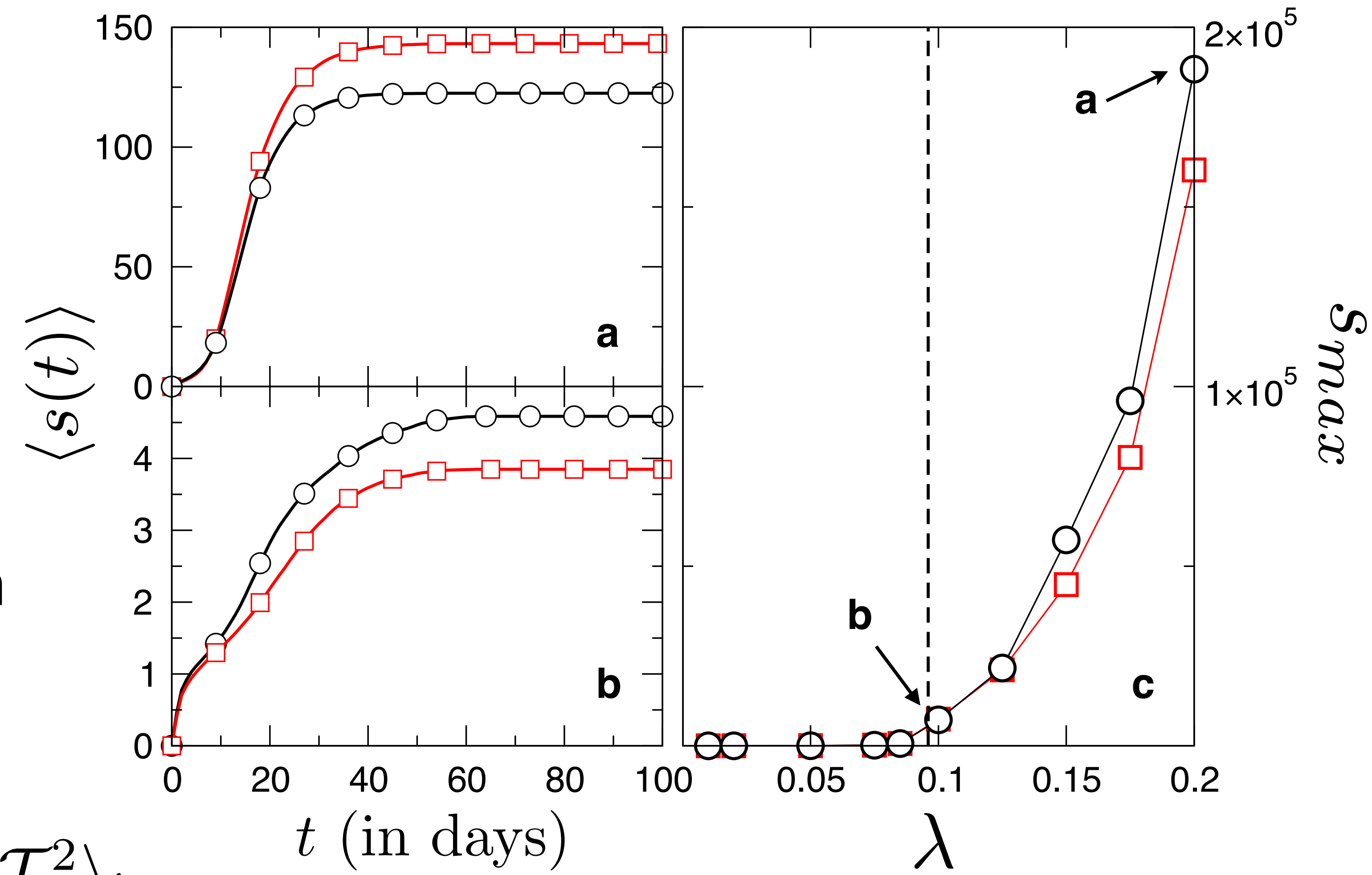
Long waiting times (bursts) make transmissibility smaller

$$\mathcal{T}_{ij} \simeq 1 - P^0_{ij}$$
$$\mathcal{T}_{ij} \leq \tilde{\mathcal{T}}_{ij}$$

Spreading including tie activity

- Smaller transmissibility =
  - Slower propagation
  - Smaller propagation

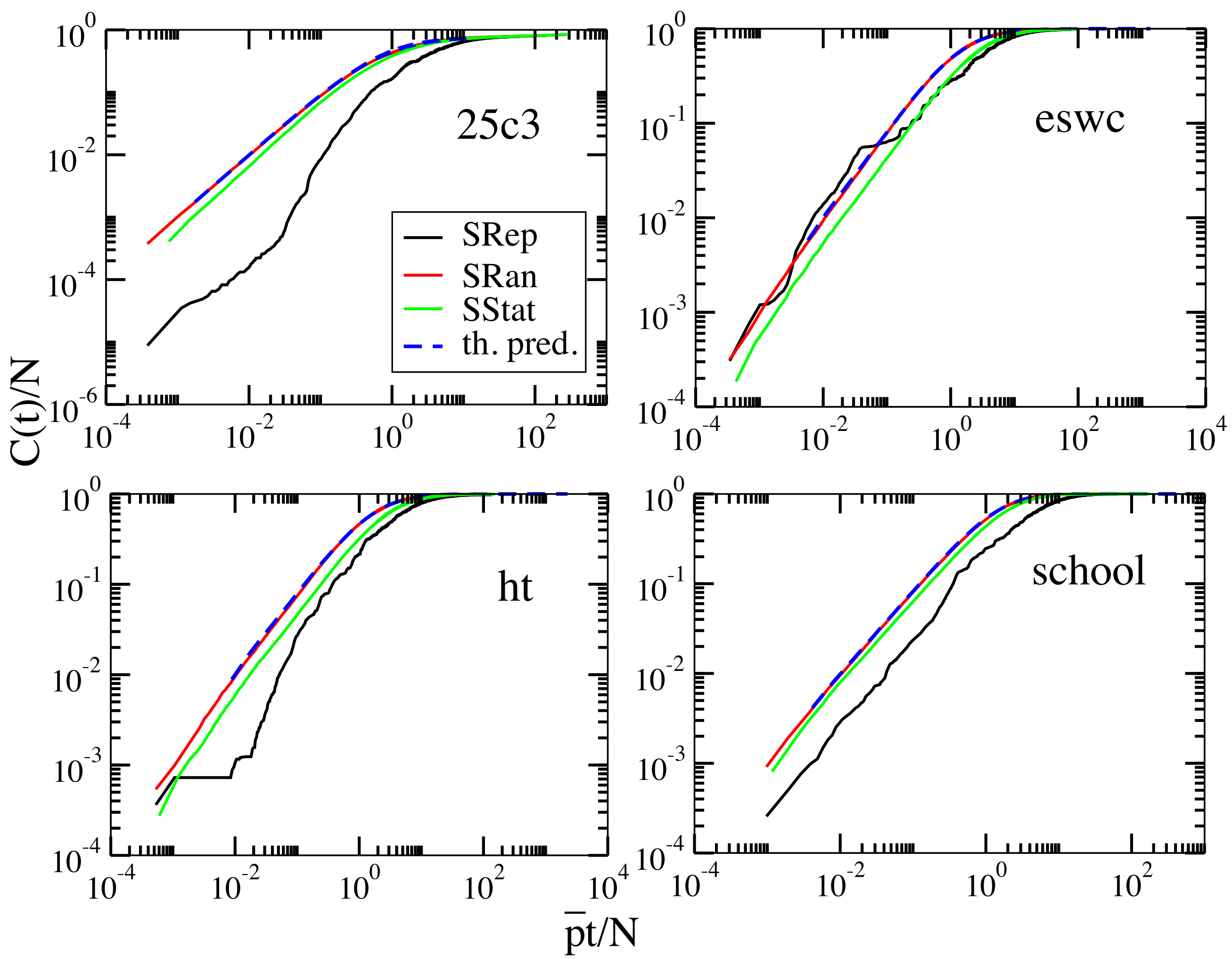
- Transmissibility can be used to predict the dynamical percolation transition



$$R_1[\lambda, T] = \frac{\langle (\sum_j \mathcal{T}_{ij})^2 \rangle_i - \langle \sum_j \mathcal{T}_{ij}^2 \rangle_i}{\langle \sum_j \mathcal{T}_{ij} \rangle_i}.$$

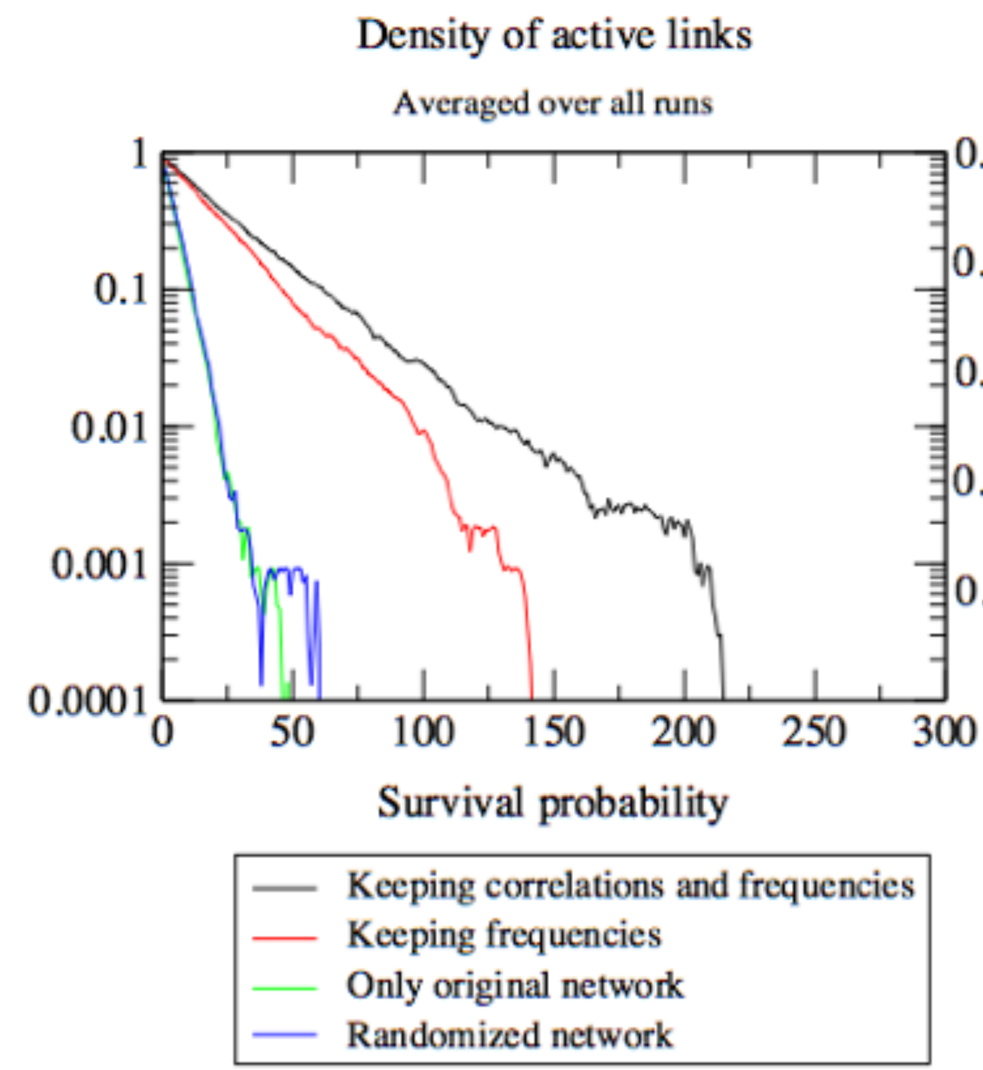
Other models

Random Walks



Starnini, M. et al., 2012. Random walks on temporal networks. *Physical Review E*, 85, pp.056115–056115.

Voter Model

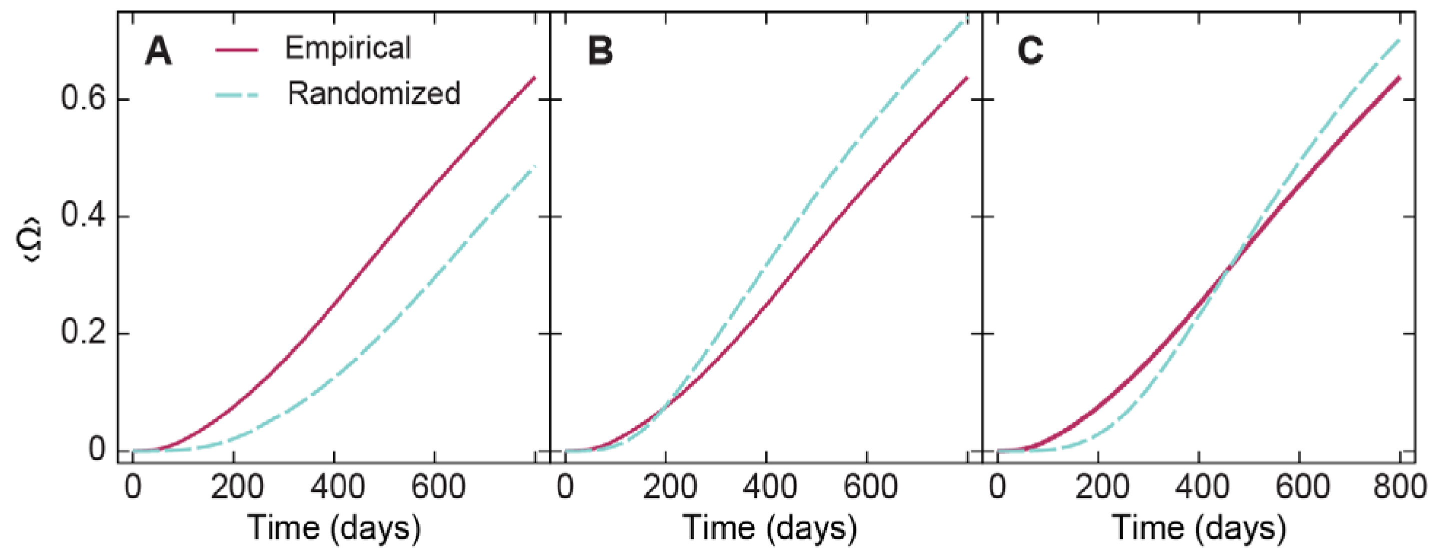


V. Eguiluz & EM, unpublished, 2011

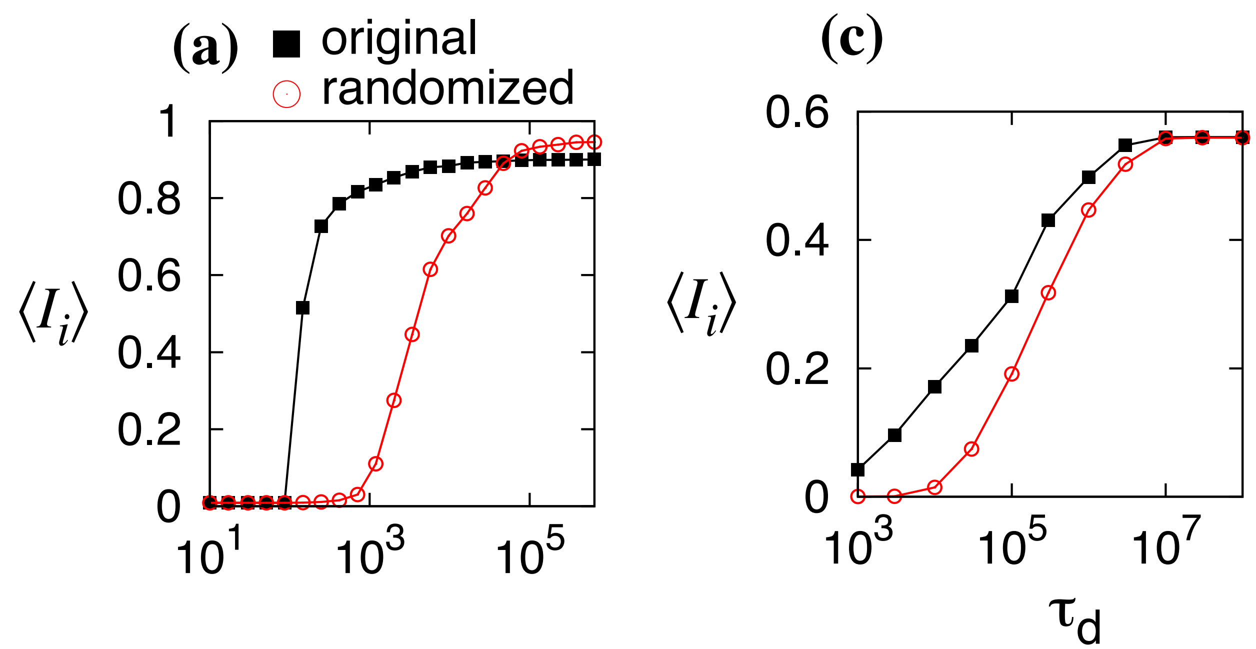
Spreading including tie activity

- Some data/models shows that burstiness accelerates contagion

Rocha, L., Liljeros, F. & Holme, P., 2011. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Computational Biology*, 7(3), p.e1001109.

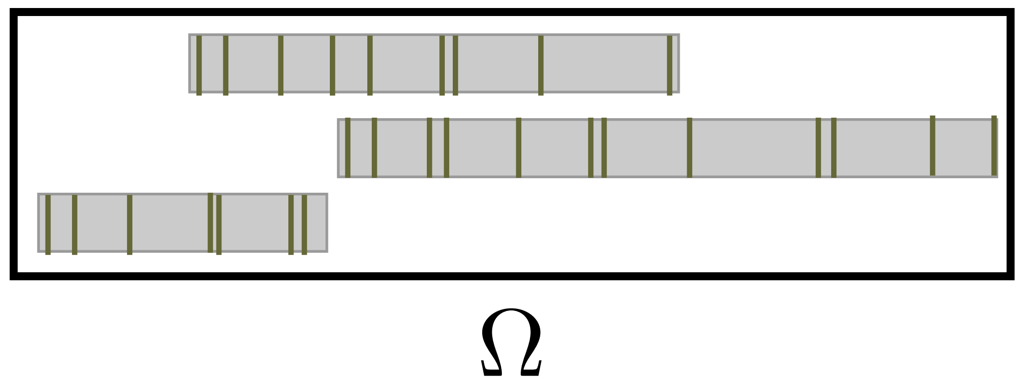


Takaguchi, T., Masuda, N. & Holme, P., 2012. Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics. *PLoS ONE*, 8(7), pp.e68629–e68629.



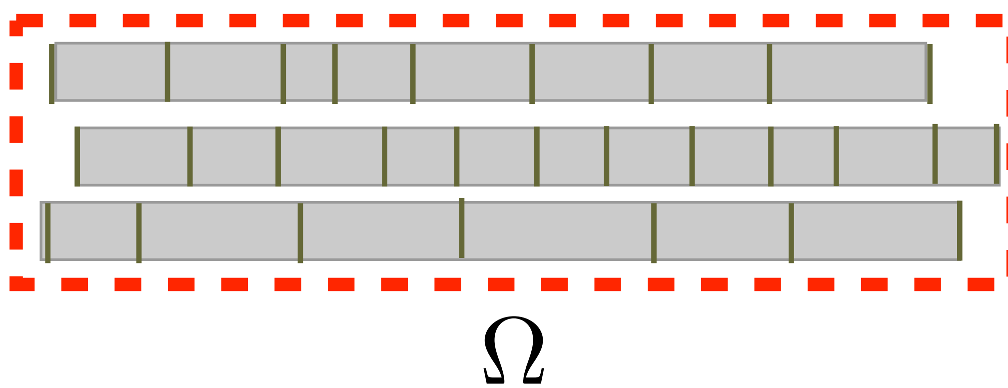
Spreading including tie dynamics

- Real data



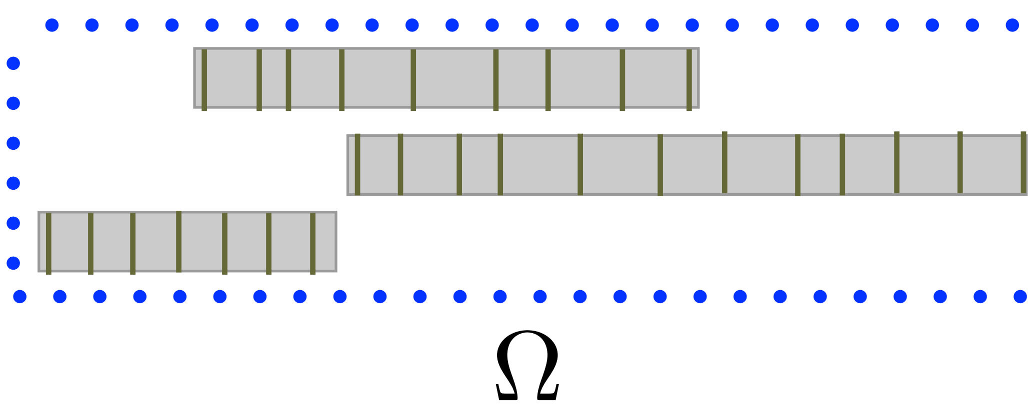
P(dt) heavy tailed
Correlated bursts
Correlated tie activity
Temporal motifs
Tie dynamics

- Shuffled data (1)



P(dt) exponential
Uncorrelated bursts
Uncorrelated tie activity
No temporal motifs
No tie dynamics

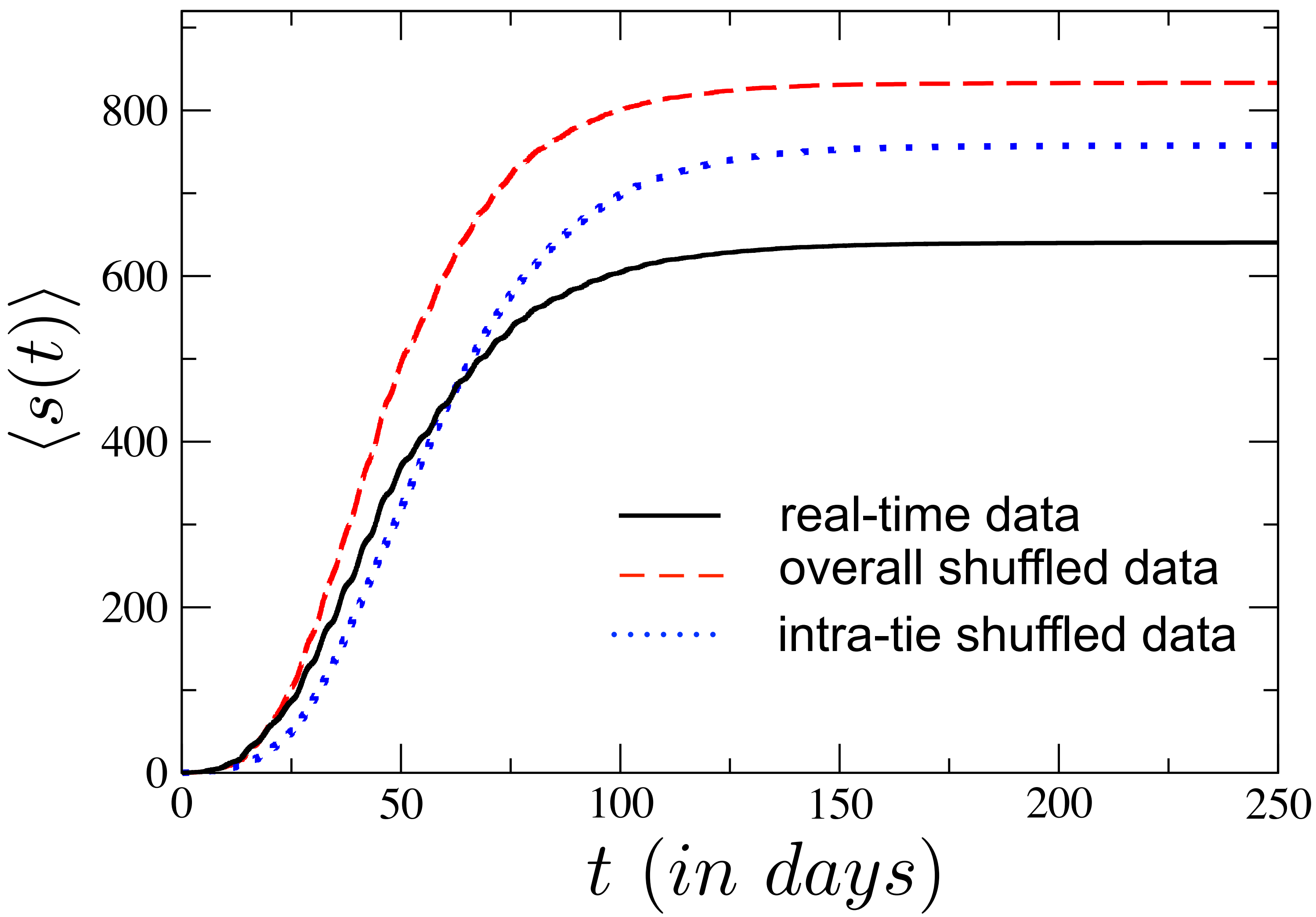
- Shuffled data (2)

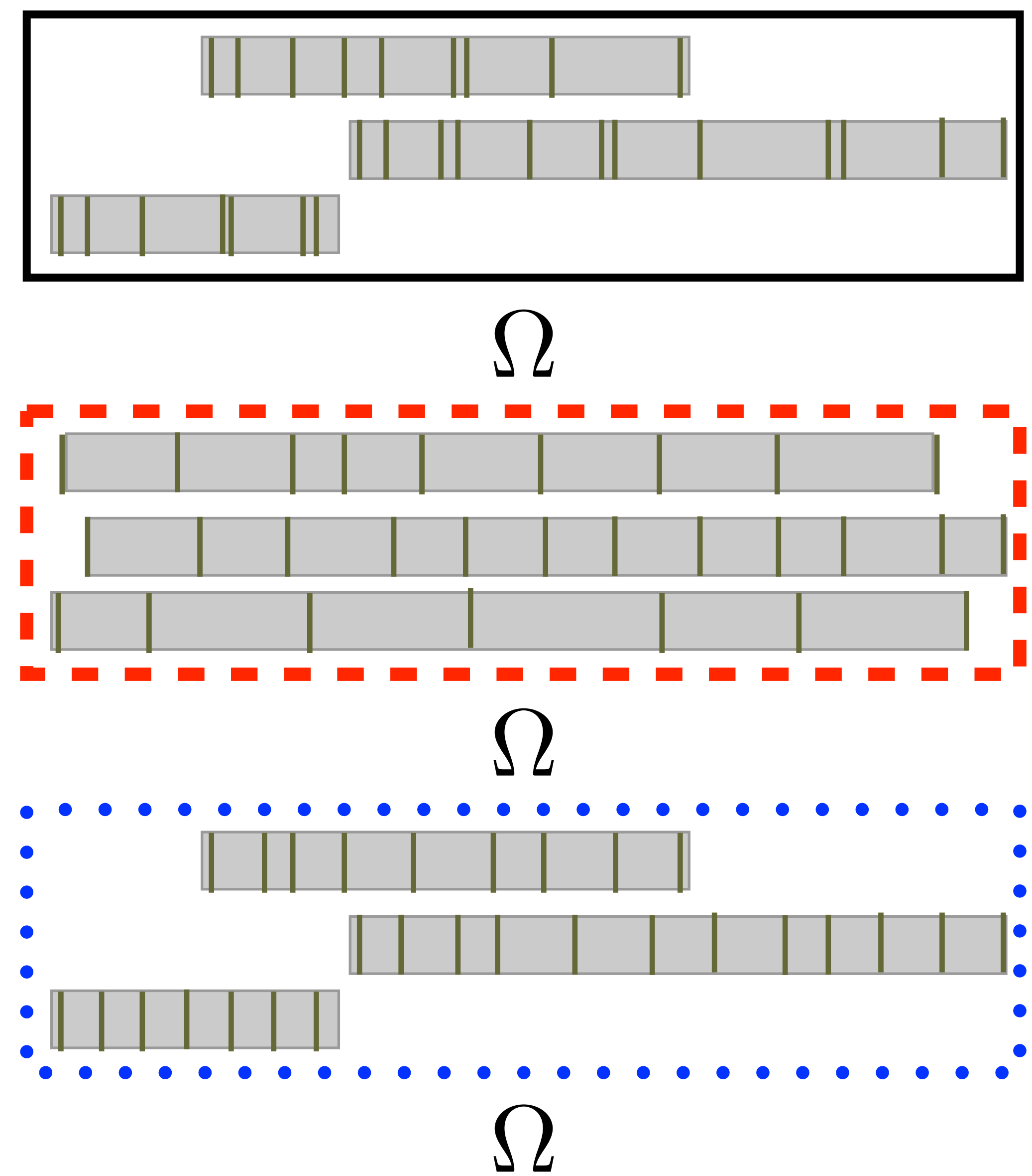


P(dt) exponential
Uncorrelated bursts
Uncorrelated tie activity
Temporal motifs ?
Tie dynamics

Spreading including time dynamics

- Burstiness + conservation of ties



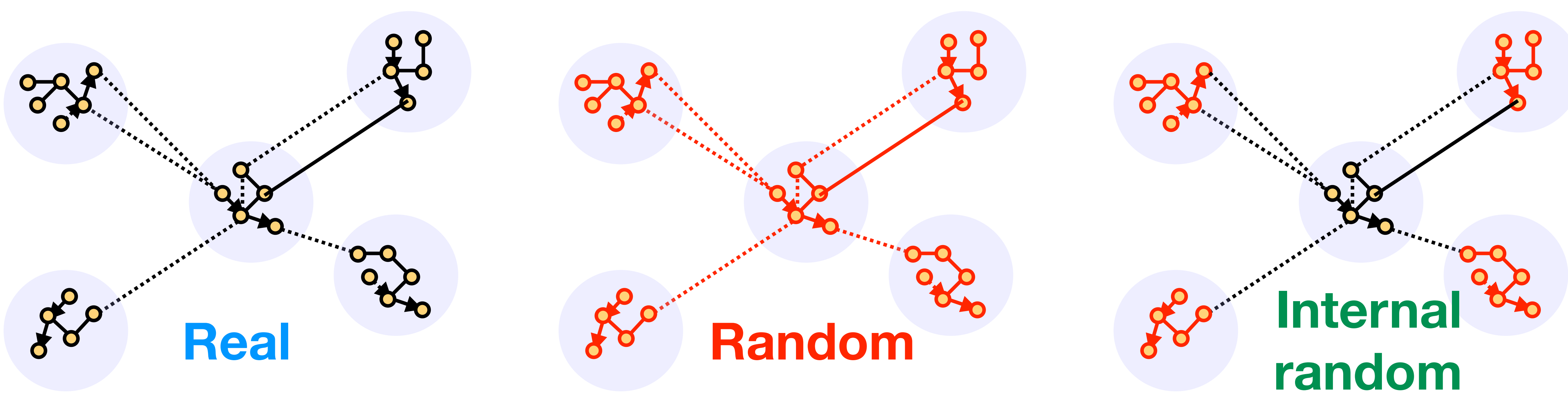


- Half of the slowing effect comes from destroying tie dynamics in the **shuffling**

Miritello, G. (2013). *Temporal Patterns of Communication in Social Networks*. Springer.

Spreading including geography

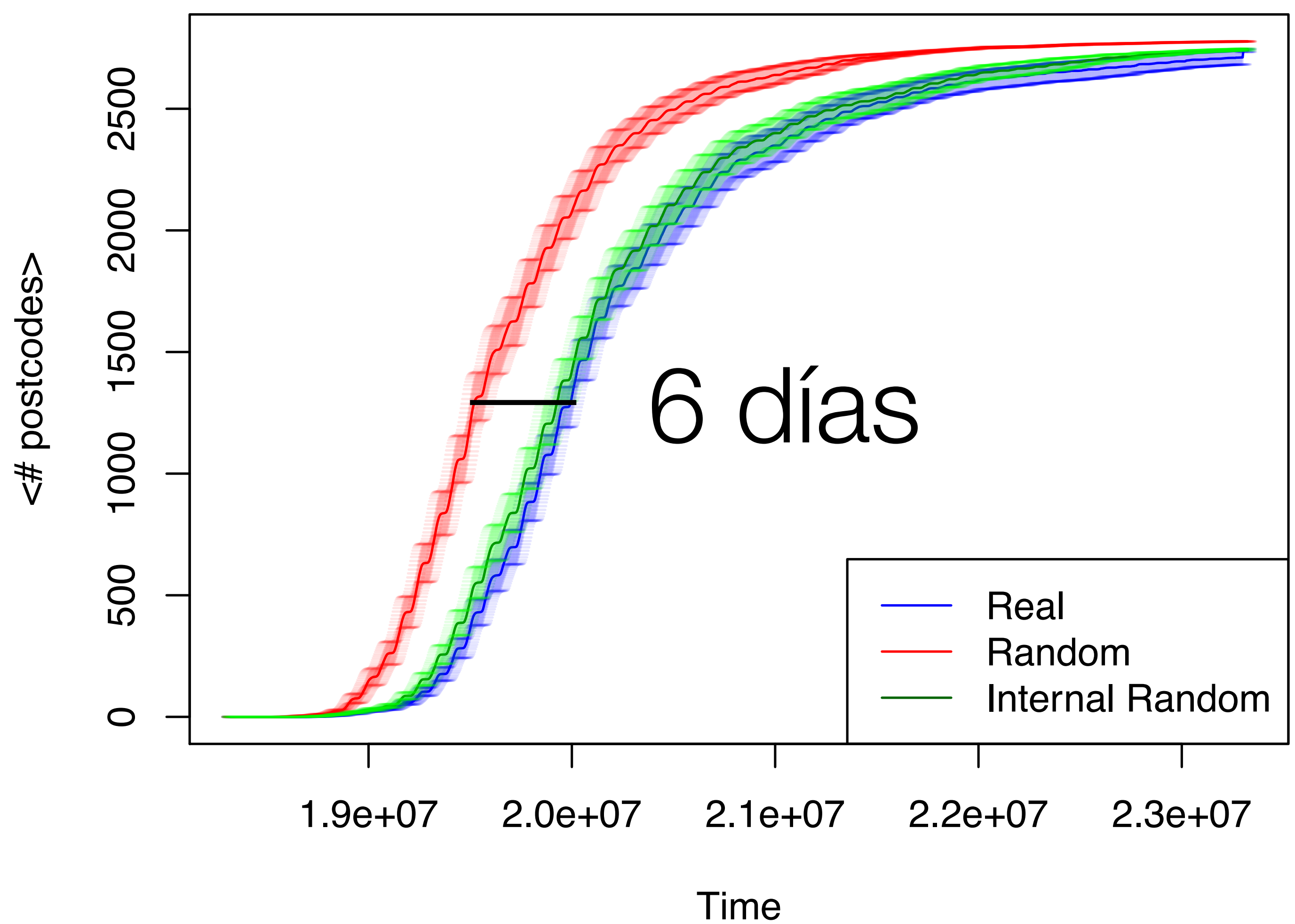
- Reminder: the fraction of unstable links is high at larger distance
- We consider another shuffling: we only shuffle ties within geographical areas

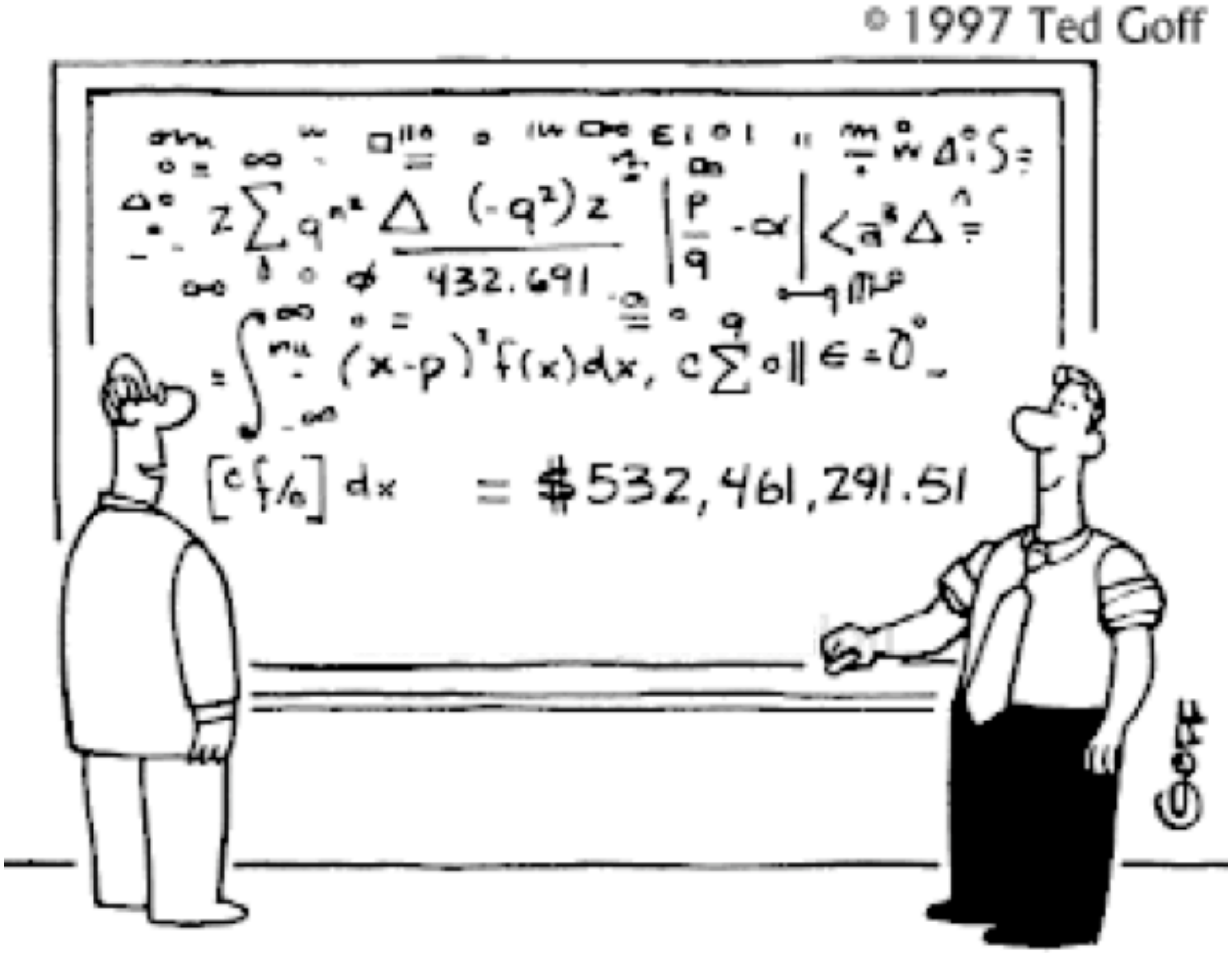


- We study propagation of information across geographical areas with the SI model
  - A geographical area is “infected” if at least a fraction of the nodes in the area is infected.

Spreading including geography

- Information spreads geographical much slower than in the shuffled case
- Most of the slowing down of information diffusion comes from inter-city links. Those links are the most unstable



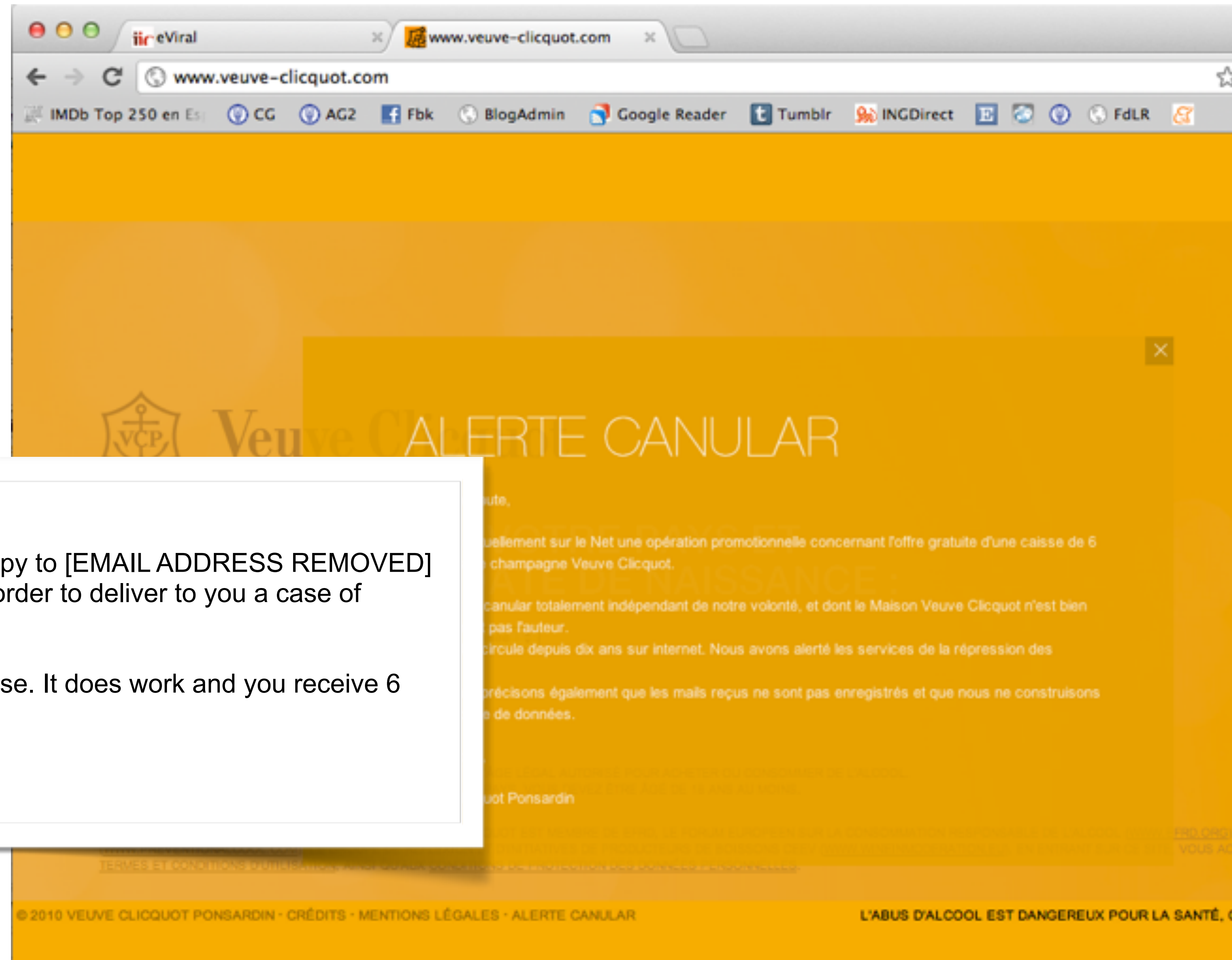


"Now that's what I call a breakthrough!"

Applications to real world 4

Information

- Hoaxes prevail for years



Hello all Champagne lovers.

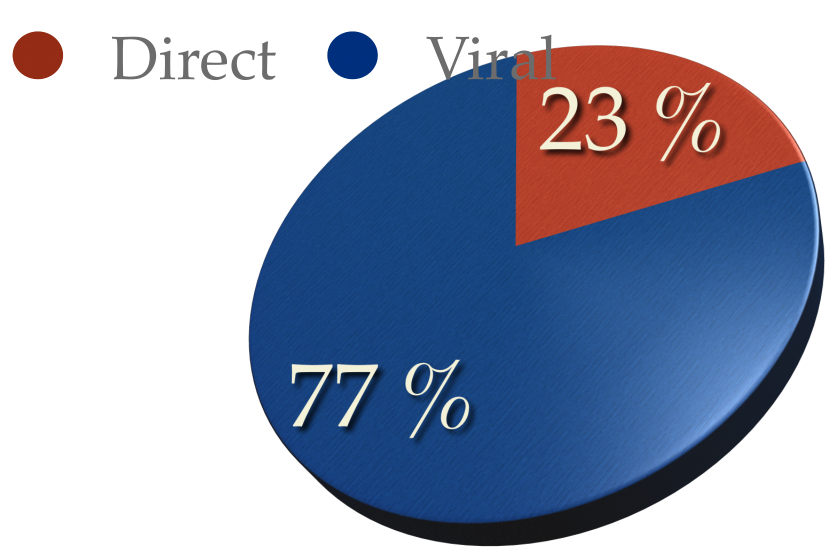
Send this message to 10 people, with a copy to [EMAIL ADDRESS REMOVED] Veuve Clicquot France will contact you in order to deliver to you a case of champagne in three weeks.

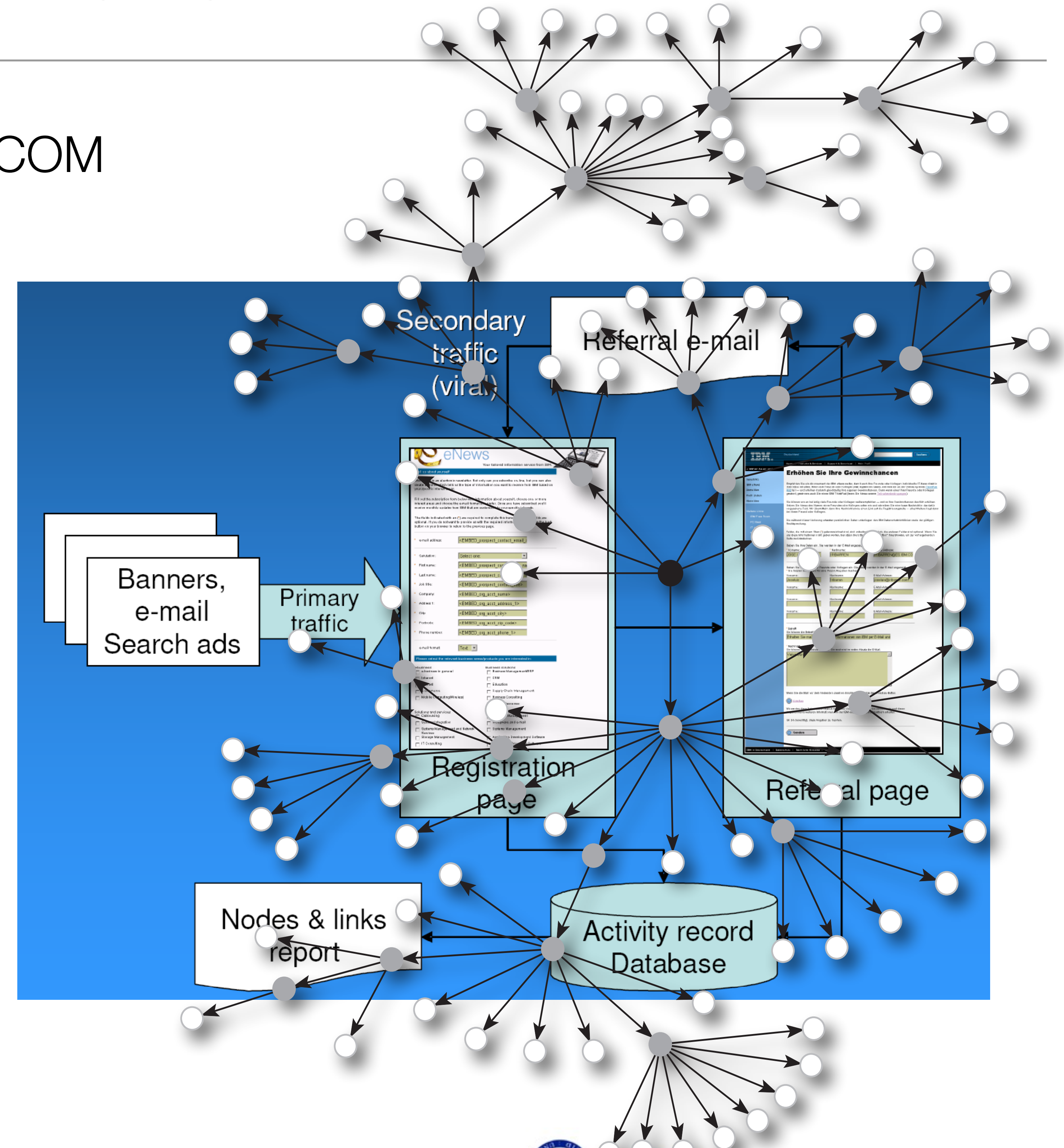
They are doing this to enlarge their database. It does work and you receive 6 bottles in 15 days.

Salut a tous les amoureux du champagne.

Viral marketing experiments

- Viral marketing campaigns IBM.COM
  - 2003-2005 IBM.COM
  - 30000 B2B clients
  - 11 european countries
  - 2 months of campaign



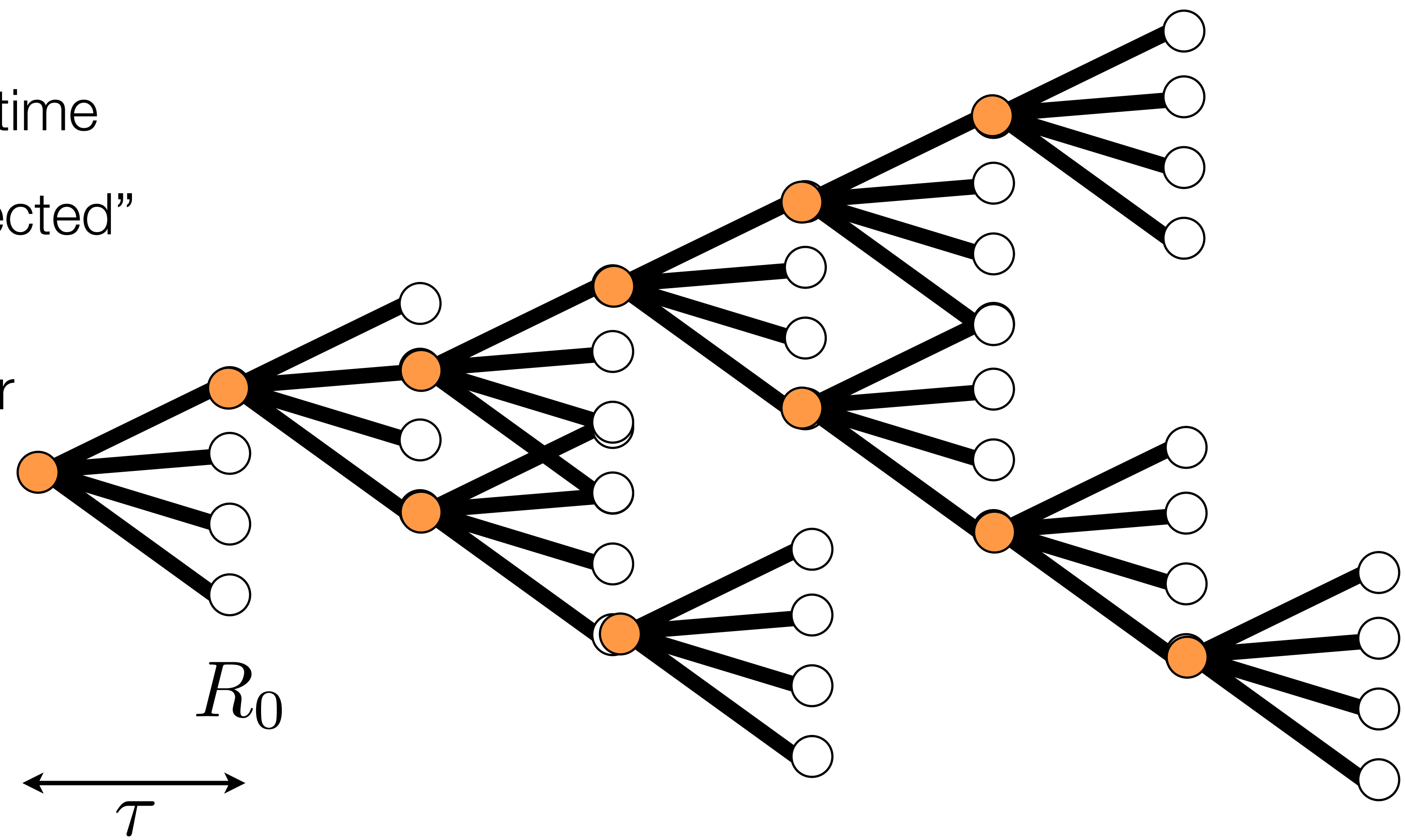


Universidad Carlos III de Madrid

@estebanmoro

Back-of-the-envelope calculation

- Assuming
  - constant response time
  - and number of “infected” friends
- What is *i(t)*, the number of infected people at time t?

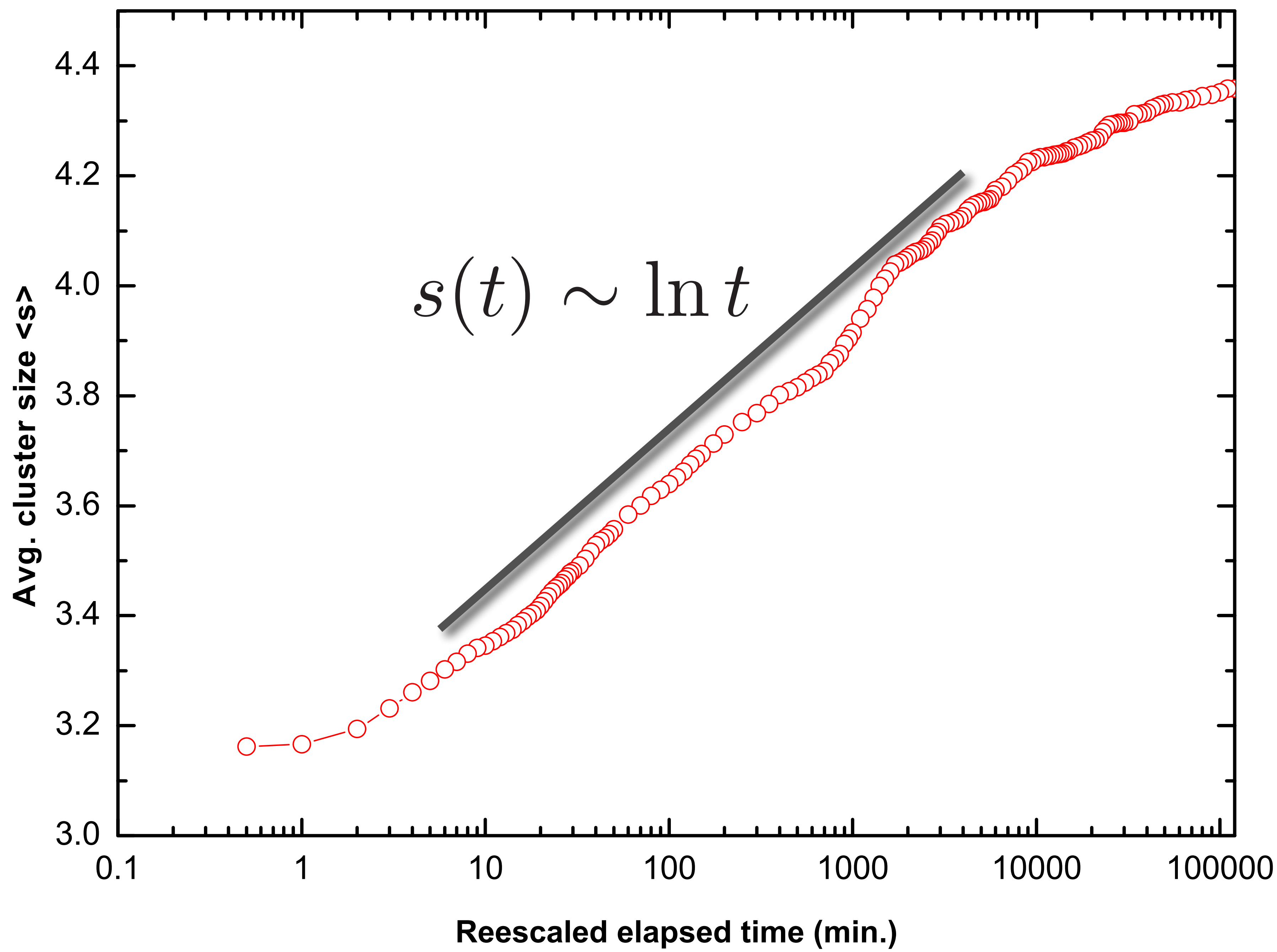


$$1 \quad + \quad R_0 \quad + \quad R_0^2 \quad + \quad R_0^3 \quad + \quad R_0^4 \quad + \quad R_0^5 + \cdots$$

$$\frac{di}{dt} = \frac{R_0 - 1}{\tau} i \qquad i(t) = i(0)e^{\alpha t}$$

Viral marketing experiments

- Viral marketing campaigns prevailed for weeks/months



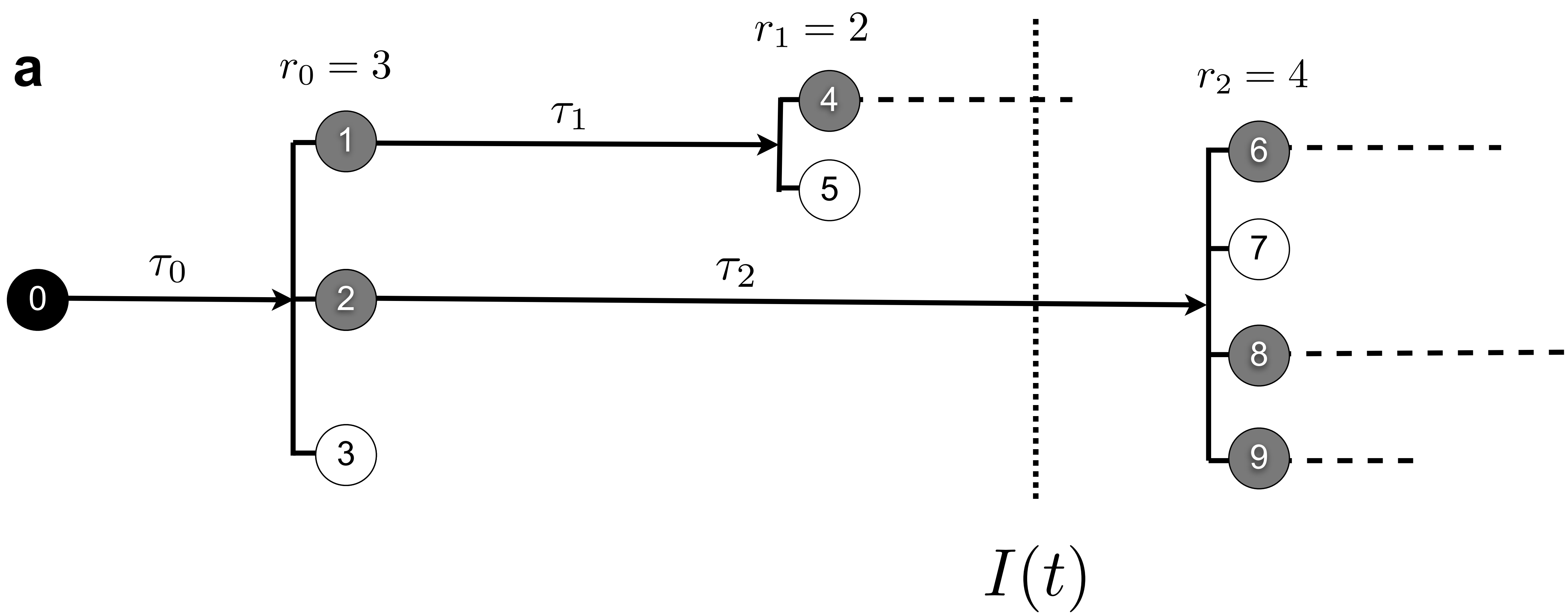
$$R_0 \simeq 0.25$$

$$\overline{\tau} \simeq 1\ day$$

$$s(t) = \int_0^t i(s)ds$$

Bellman-Harris process

- The process is characterised by the distribution of number of recommendations and response time $P(k)\ \ [P(r)]\ \ G(\tau)$



@estebanmoro

Bellman-Harris Process

- It is the well-known Bellman-Harris Process

$$i(t) = 1 - G(t) + R_0 \int_0^t dG(t)\ i(t-\tau)$$

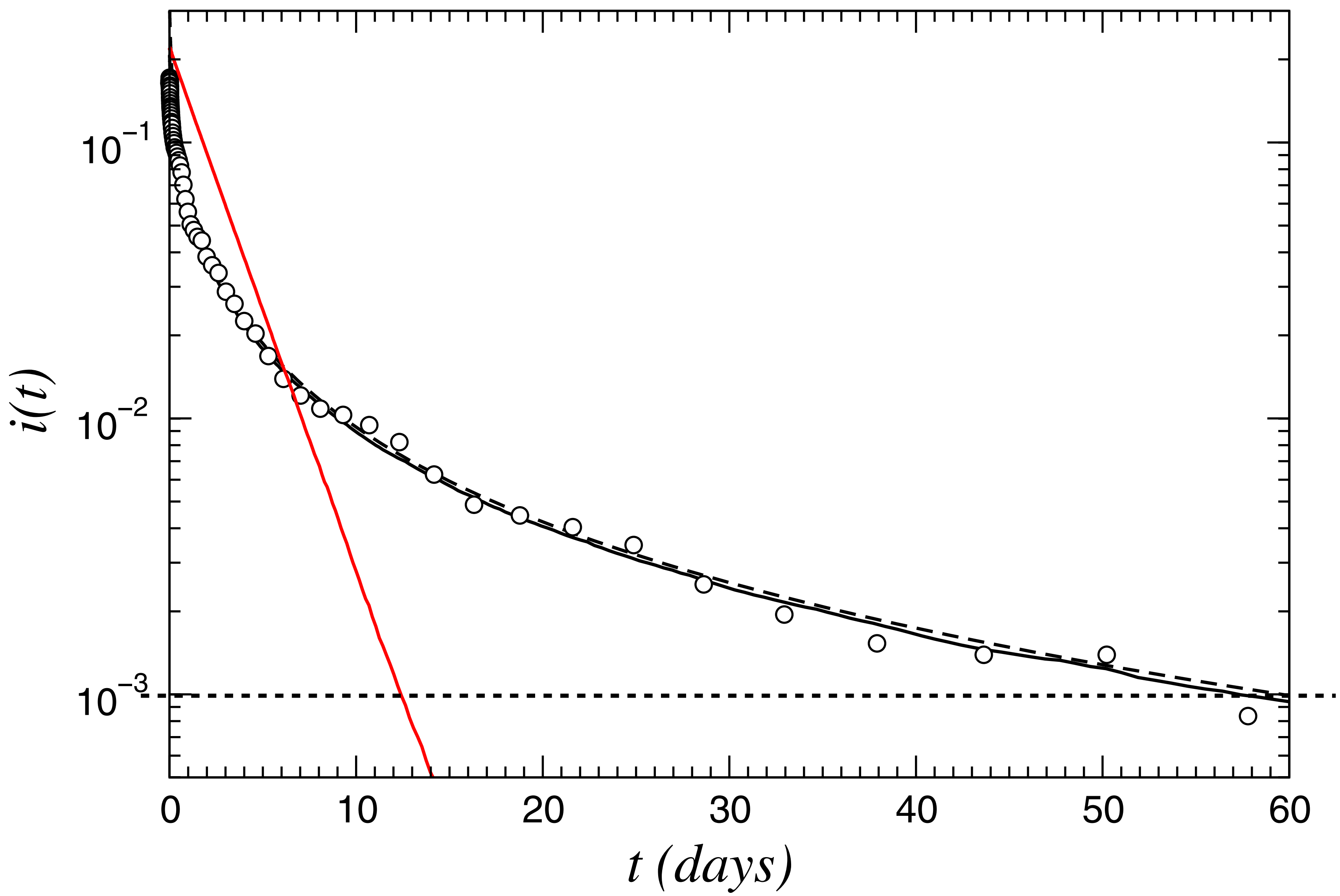
- where R_0 is the secondary reproductive number and $i(t) = \langle I(t) \rangle$
- The dynamics is determined by the tail of the distribution

Theorem. (Athereya & Ney '70s) If $R_0 < 1$ and G is in the sub-exponential class $\mathscr{S}$, then

$$i(t) \sim \frac{1 - G(t)}{1 - R_0}$$

Information travels in logarithmic time

- Prevalence



$$i(t_f) \simeq 10^{-3}$$

$$t_f \simeq 12\ days$$

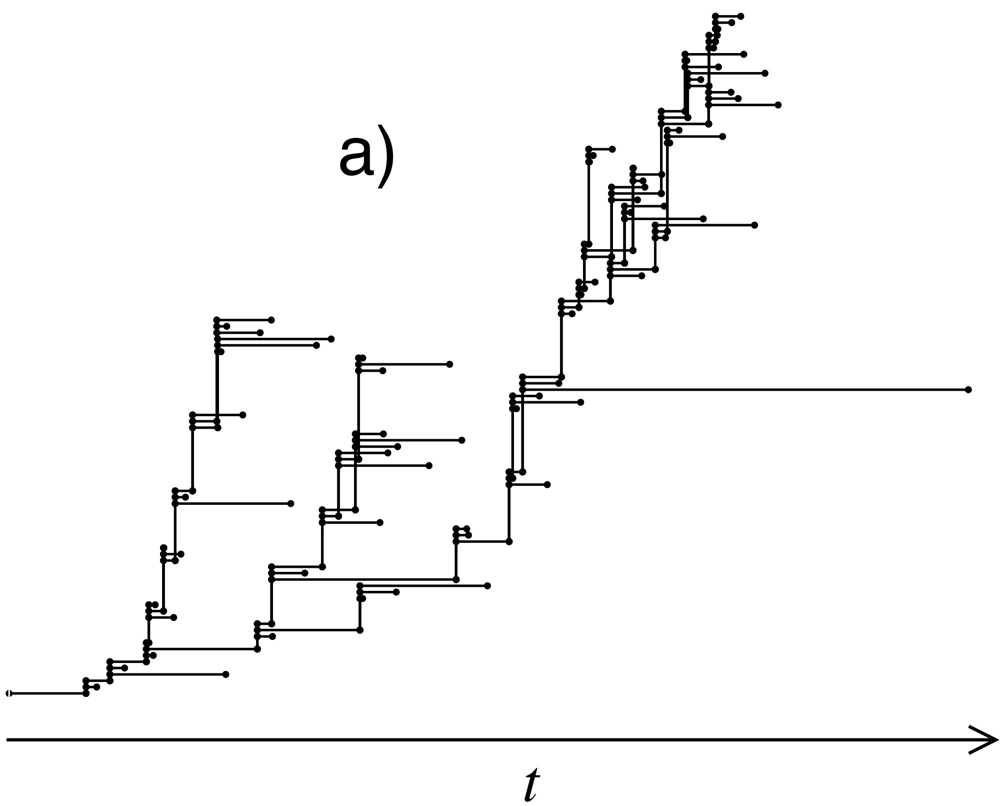
$$t_f \simeq 2\ months$$

$$i(t_f) = 10^{-4}$$

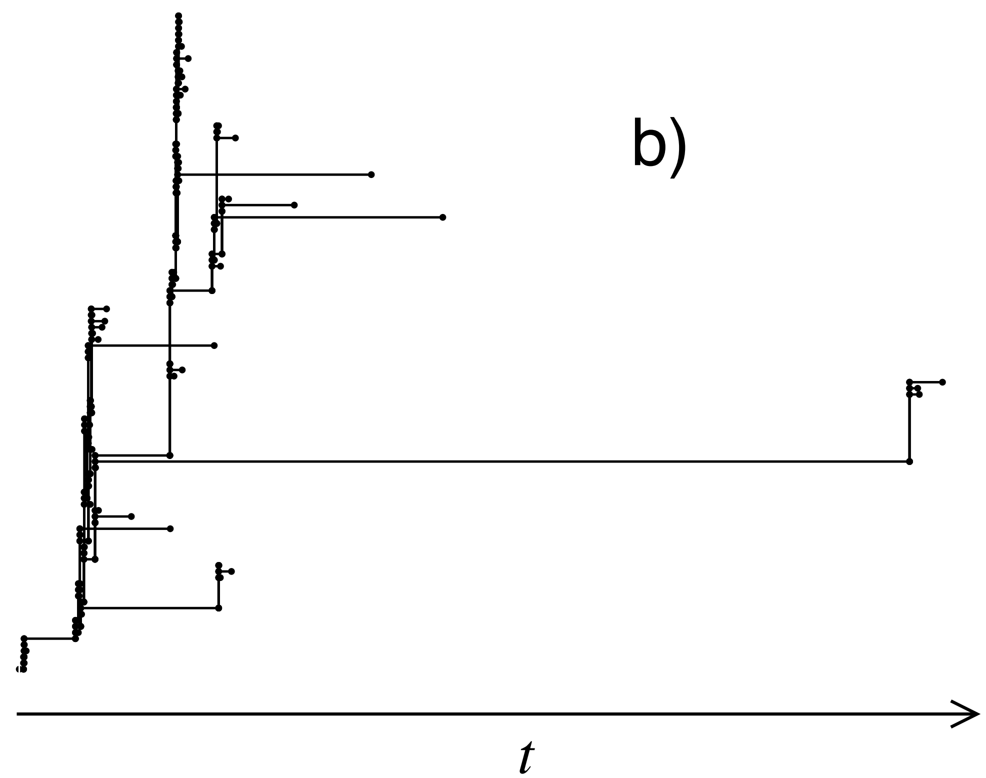
$$t_f \simeq 17\ days$$

$$t_f \simeq 1\ year$$

Information spreading is dominated by the tail of the distribution



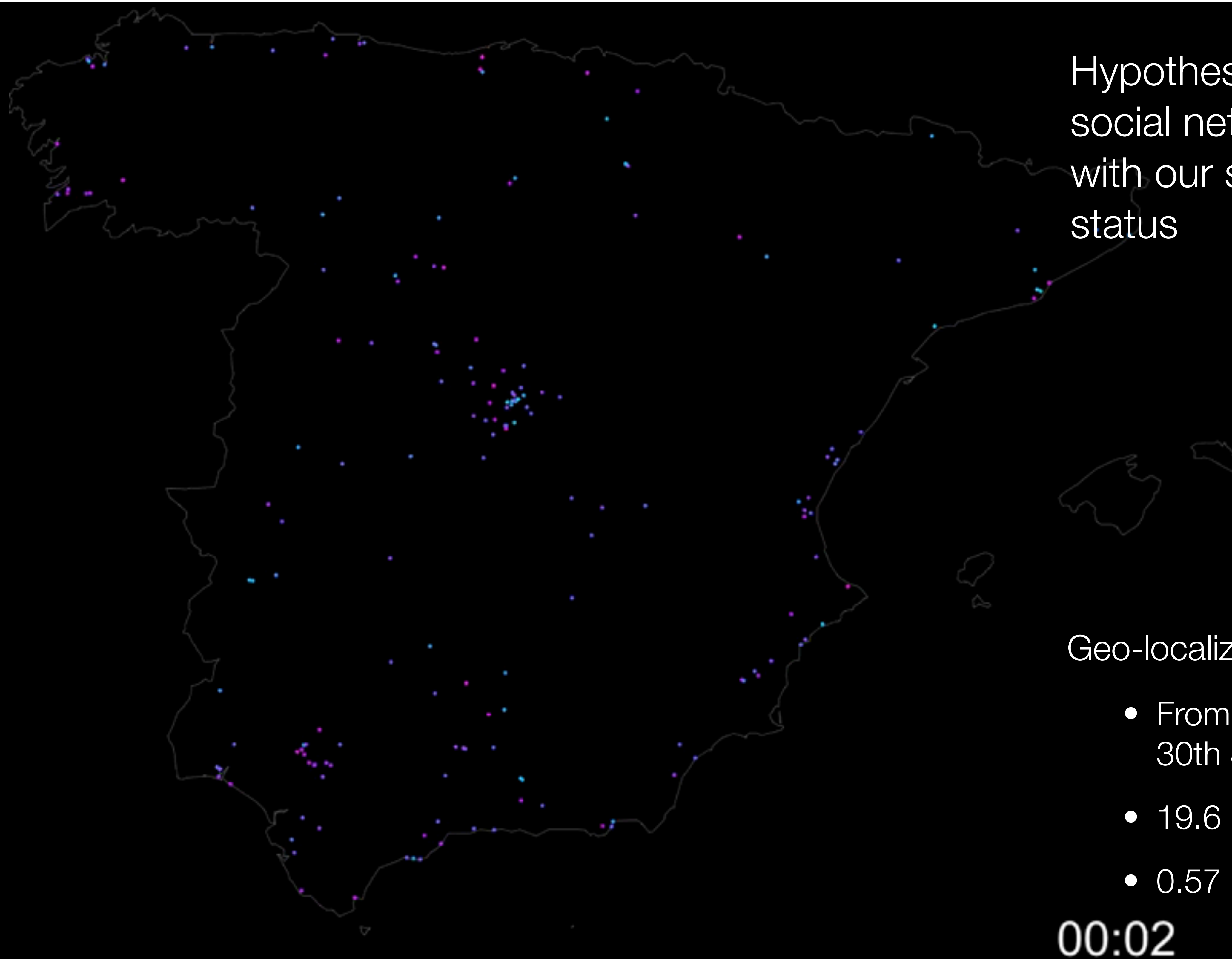
Homogeneous response time
(exponential)



Heterogeneous response time
(subexponential)

1. Iribarren, J. E. L., & Moro, E. (2009). Impact of human activity patterns on the dynamics of information diffusion, 103(3), 038702–038702. http://doi.org/10.1103/PhysRevLett.103.038702
2. Iribarren, J. L., & Moro, E. (2011). Branching dynamics of viral information spreading, 84(4), 46116. http://doi.org/10.1103/PhysRevE.84.046116

Individual activity and socio-economical situation



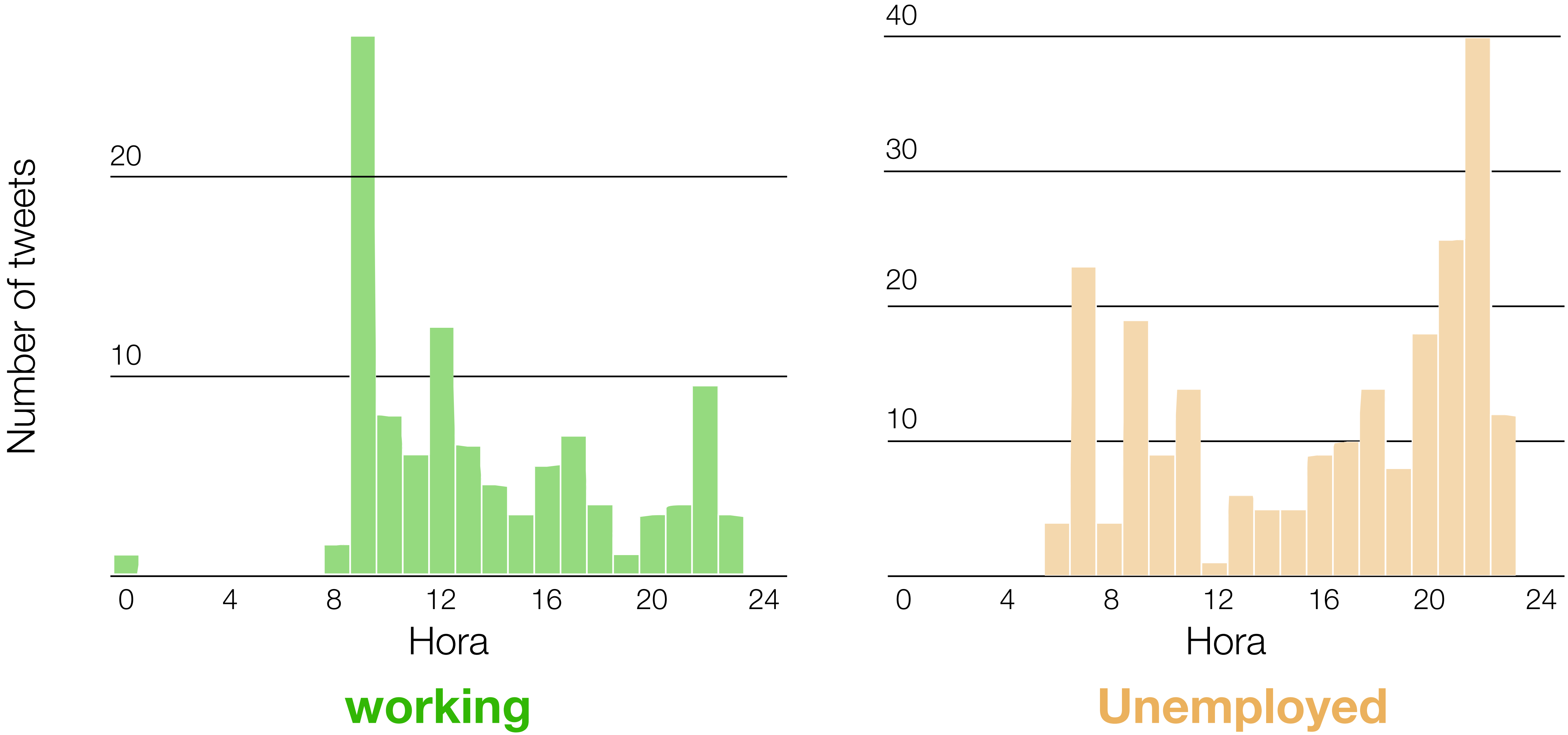
Hypothesis: our activity in social networks is correlated with our socio-economical status

Geo-localized tweets in Spain

- From 29th Nov 2012 to 30th June 2013
- 19.6 million tweets
- 0.57 million unique users

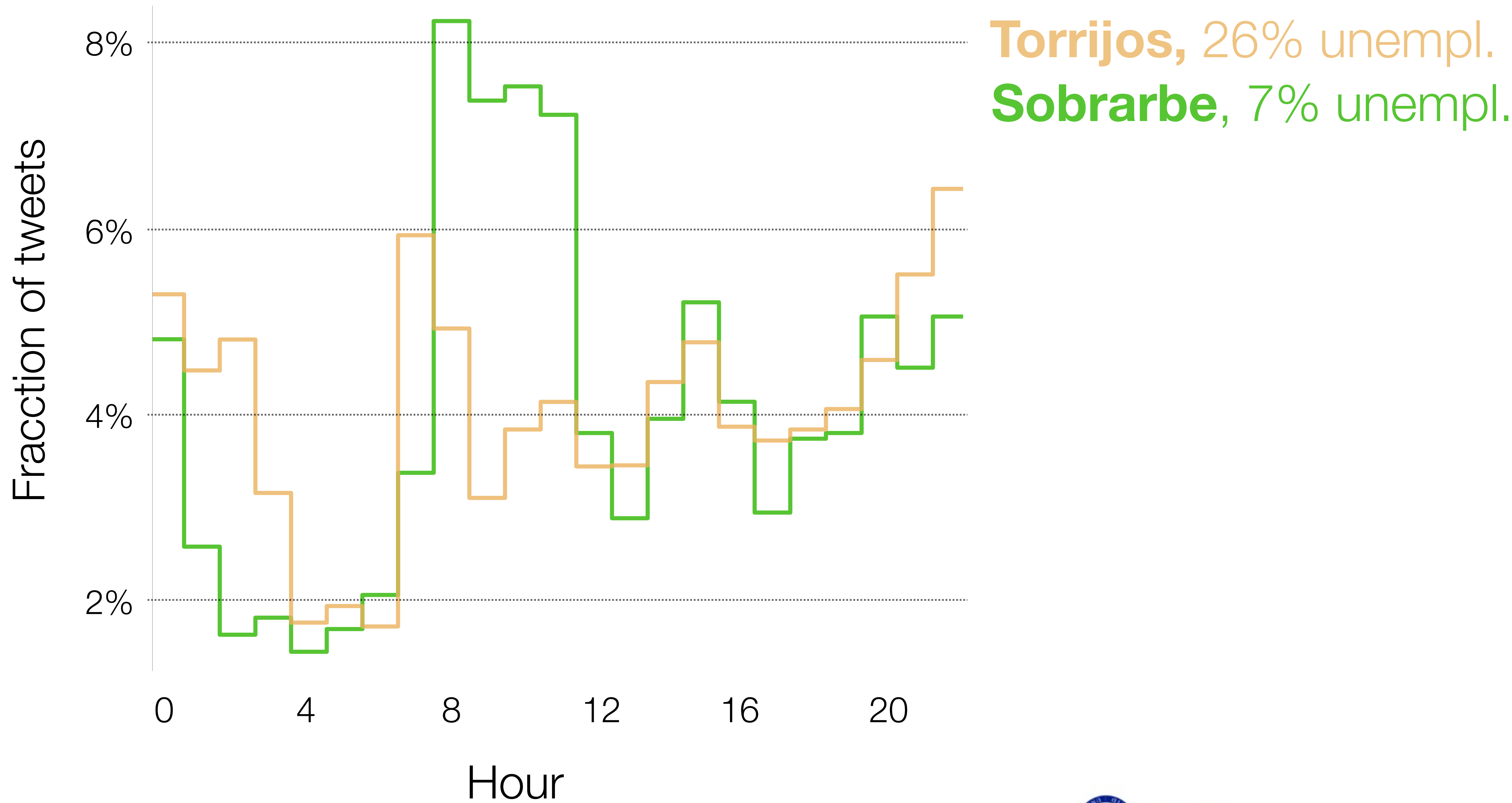
Individual activity and socio-economical situation

- Our daily activity is impacted by our socio-economical situation
    - At the individual level



Individual activity and socio-economical situation

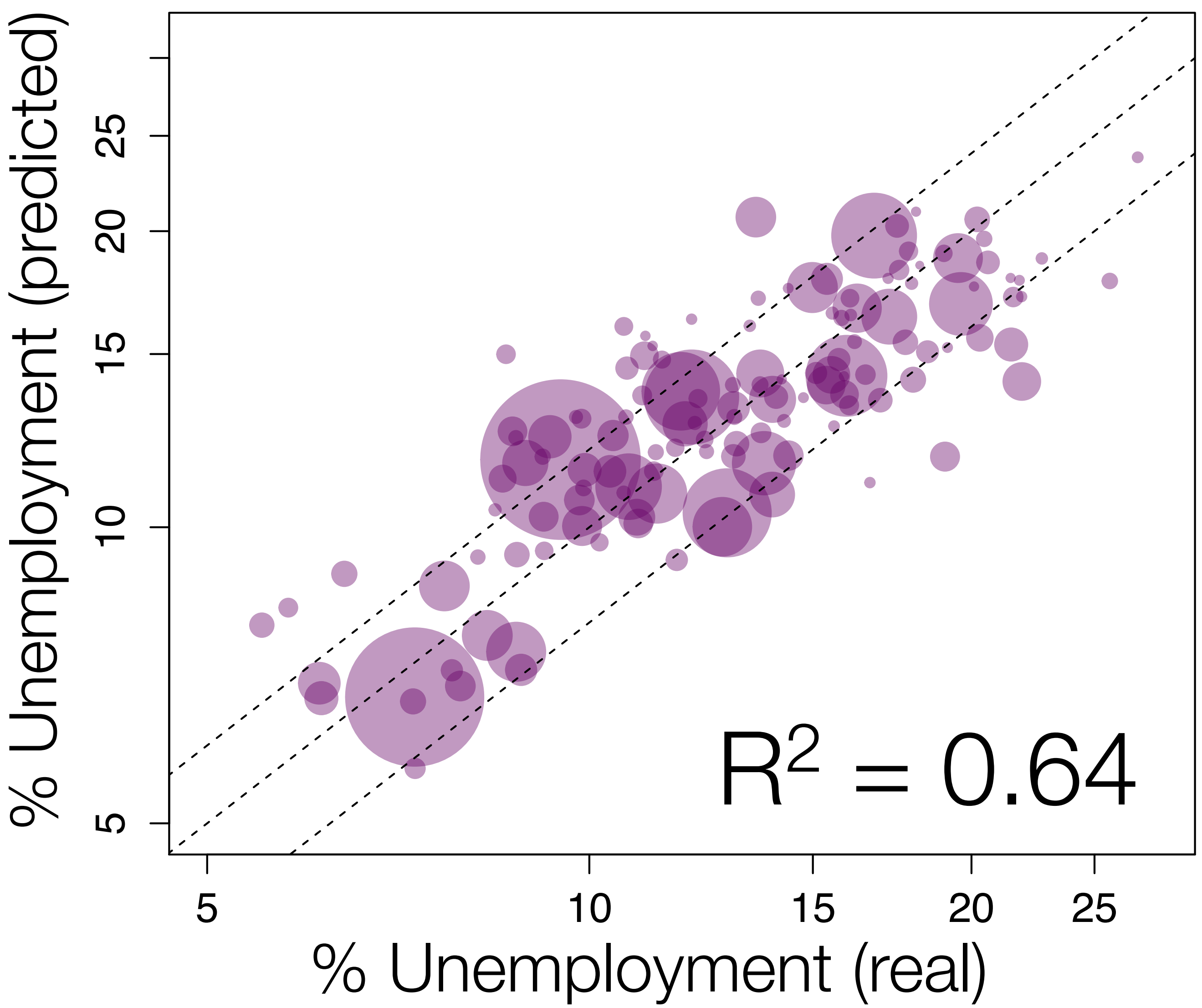
- Our daily activity is impacted by our socio-economical situation
  - At group/city level

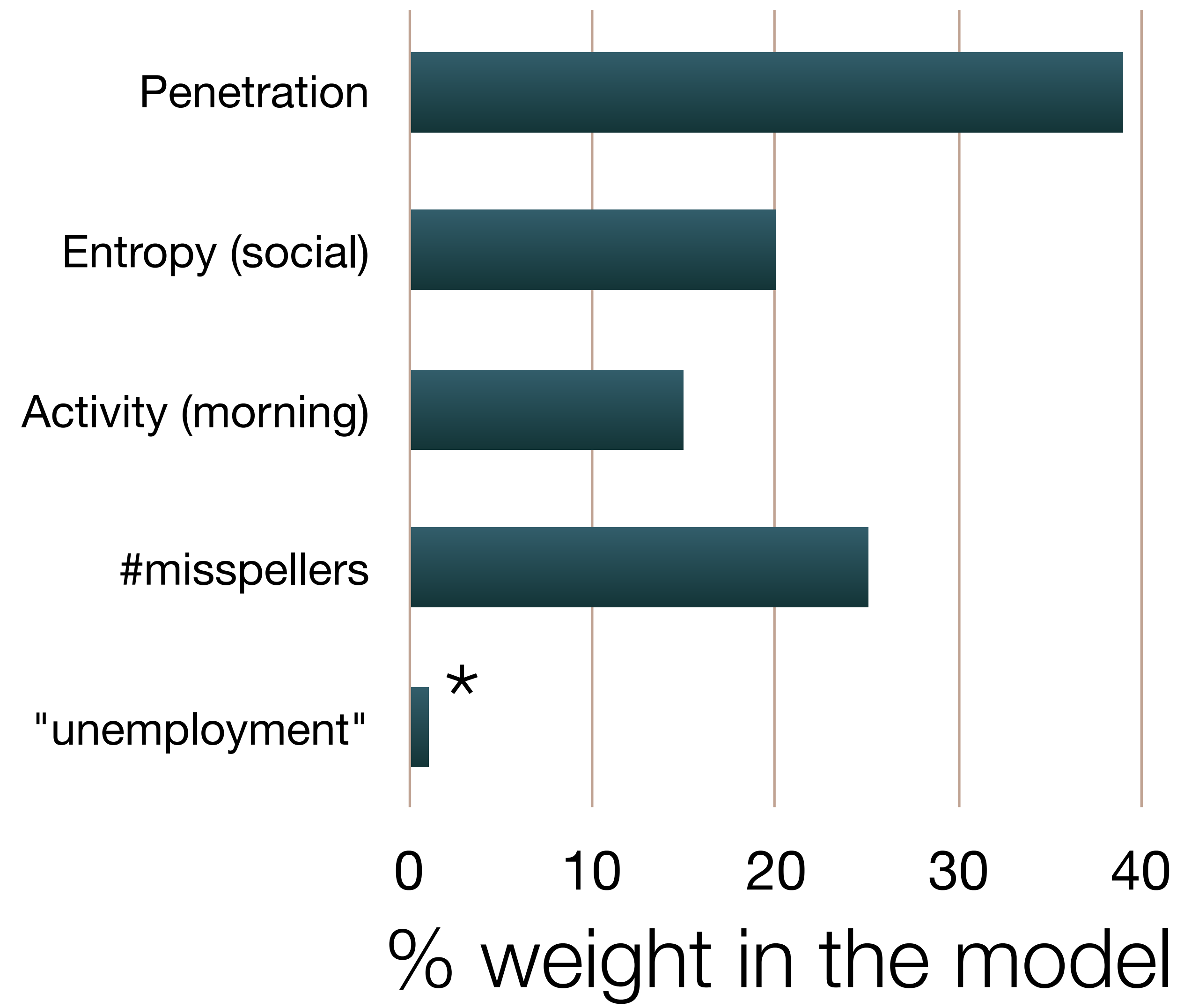


Universidad Carlos III de Madrid @estebanmoro

Explanatory power of Twitter variables

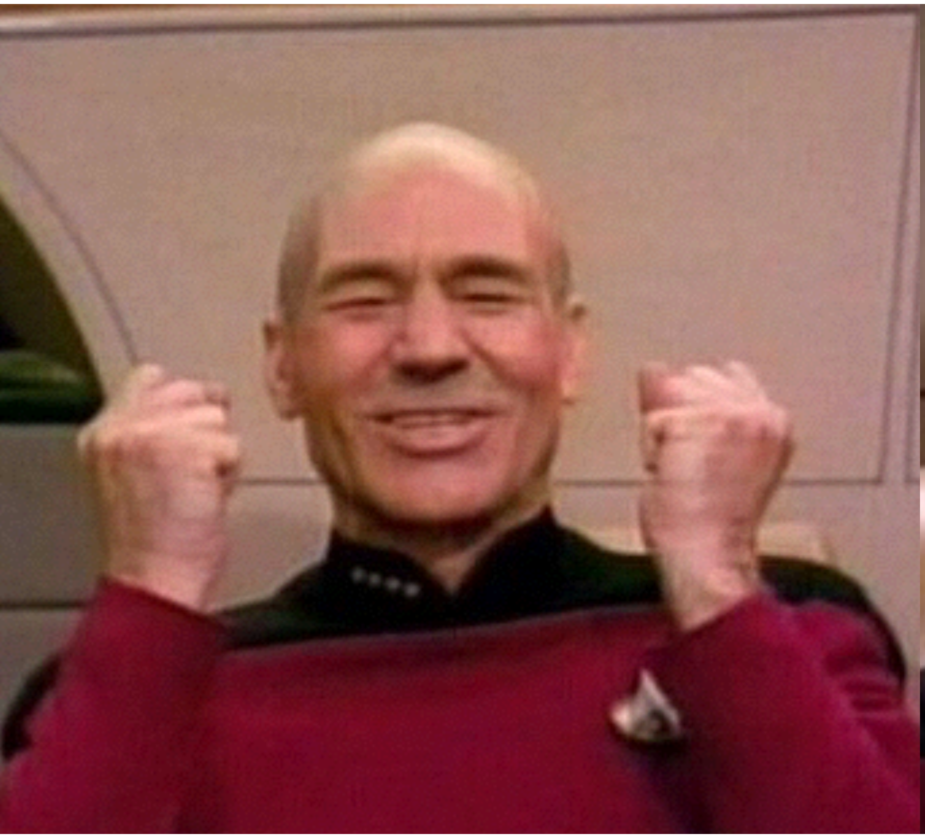
- Simple linear regression

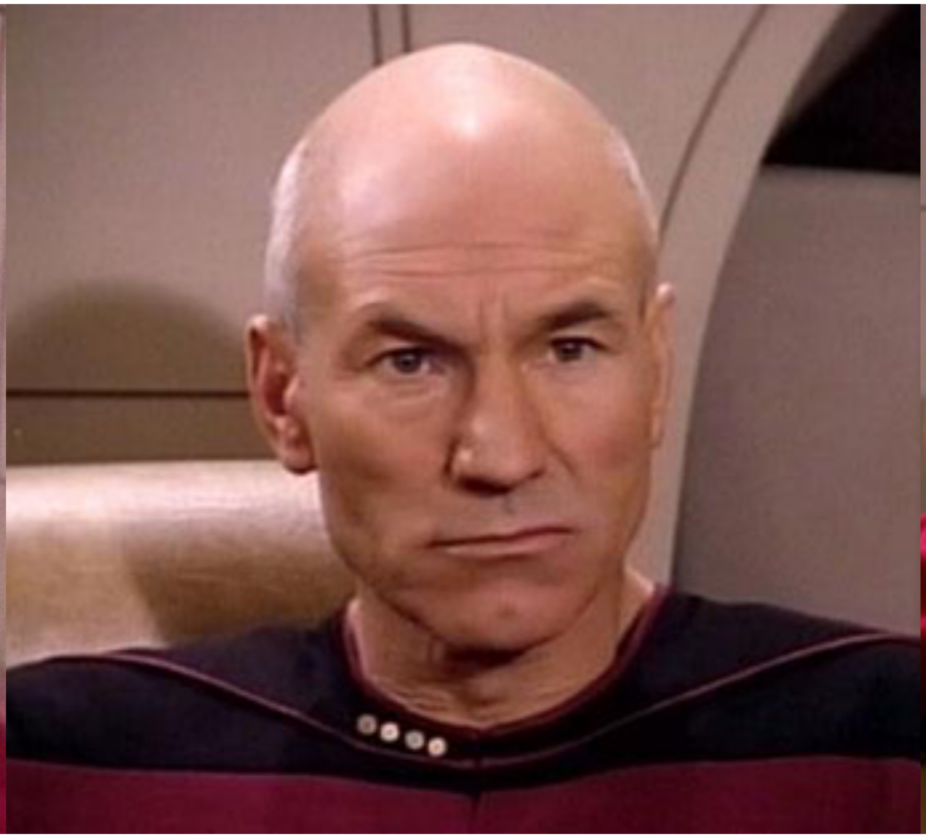


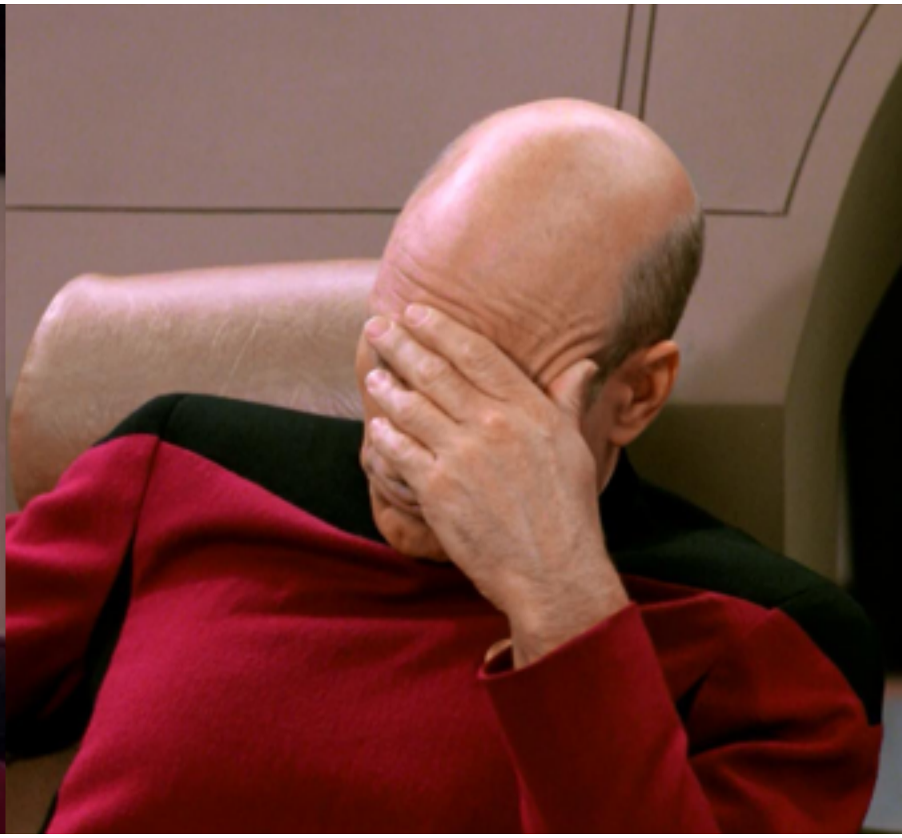


Llorente, A., Garcia-Herranz, M., Cebrian, M., & Moro, E. (2015). Social media fingerprints of unemployment. PLoS ONE, 10(5), e0128692. http://doi.org/10.1371/journal.pone.0128692

Researcher | Data scientist | Policy maker



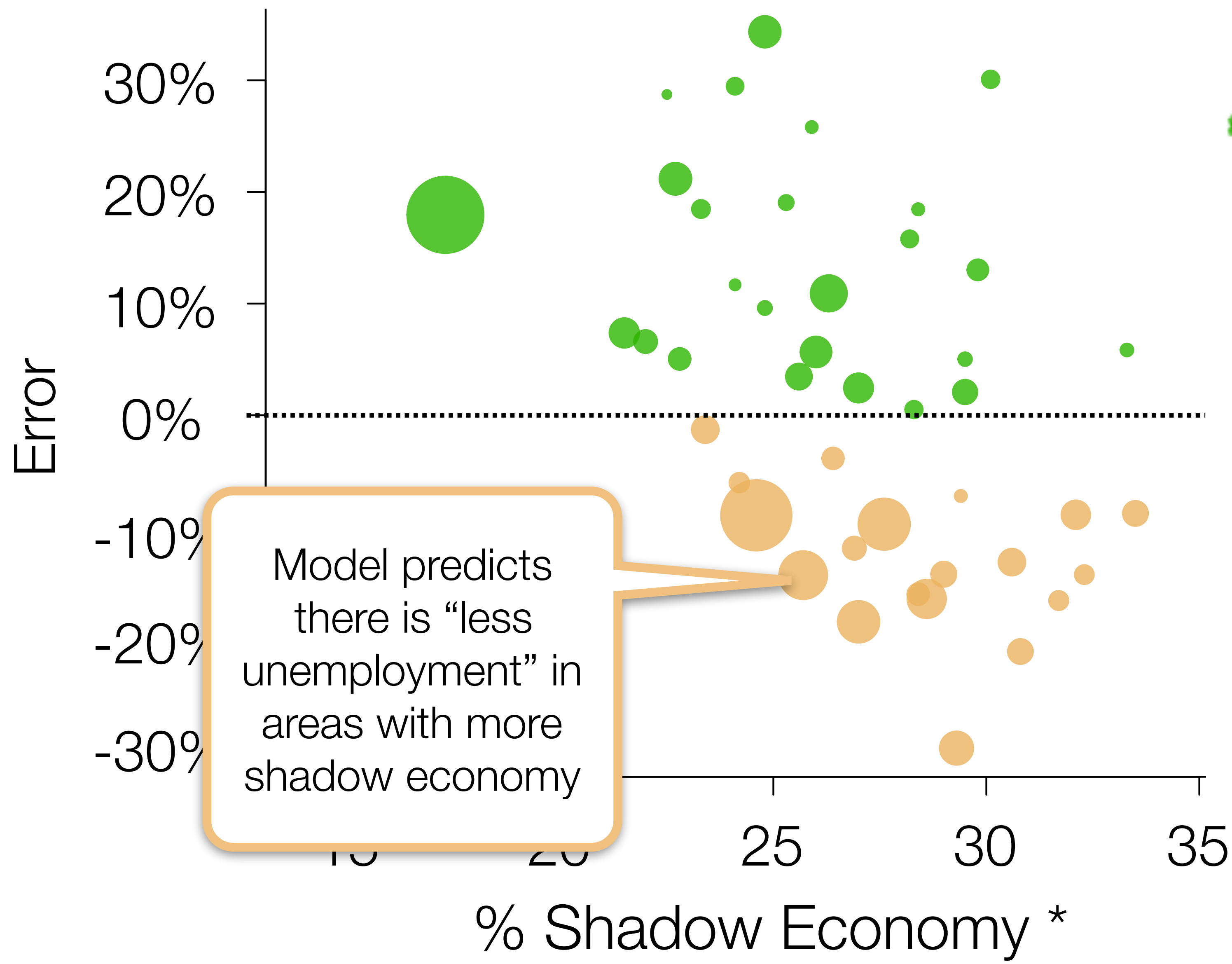




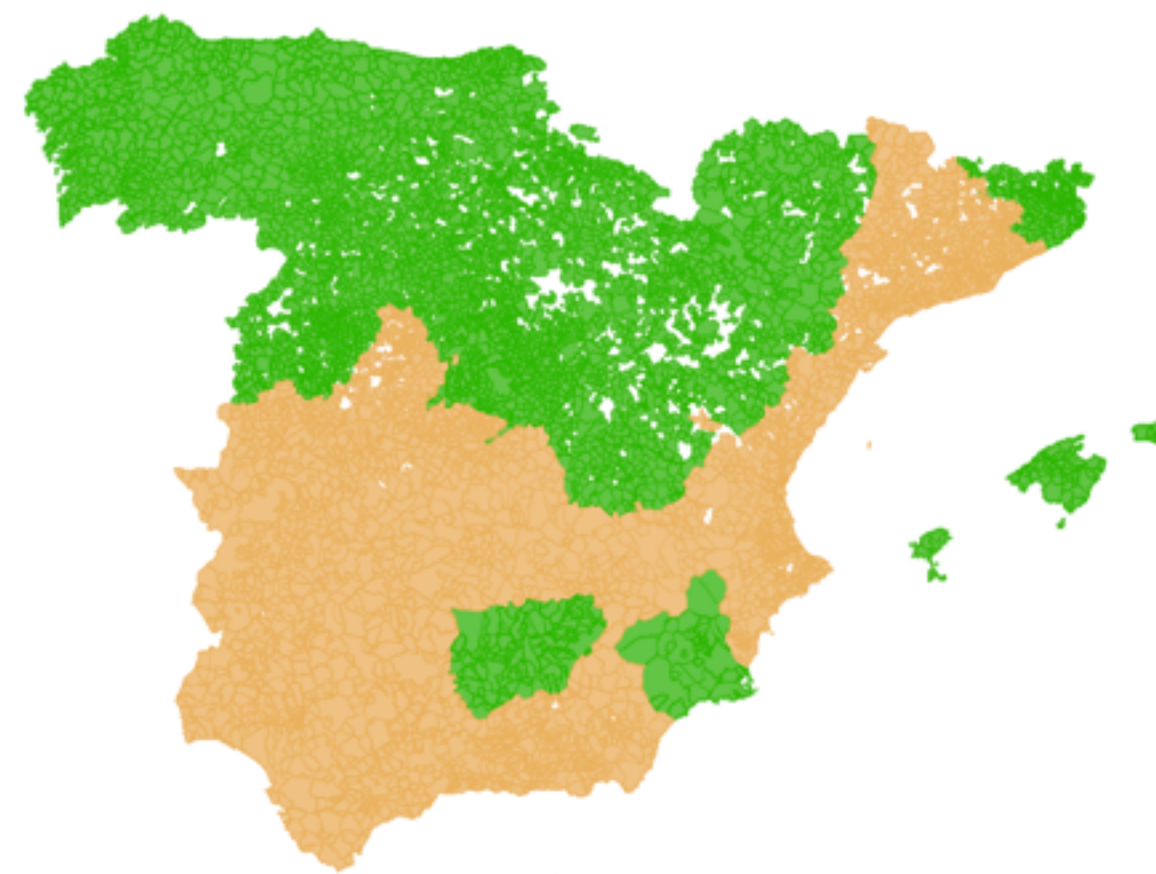
Researcher | Editor | Referee

Are we really wrong?

Model Error = Model[variables] - Official unemployment



(* GESTHA report 2012)



References

- Reviews

  - Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. EPJ Data Science, 4(1), 10. http://doi.org/10.1140/epjds/s13688-015-0046-0
  - MOBILE PHONE NETWORK DATA FOR DEVELOPMENT. (2013). UN Global Pulse
  - Saramaki, J., & Moro, E. (2015). From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. The European Physical Journal B, 88(6). http://doi.org/10.1140/epjb/e2015-60106-6
  - Naboulsi, D., Fiore, M., Ribot, S., & Stanica, R. (n.d.). Large-scale Mobile Traffic Analysis: a Survey. IEEE Communications Surveys & Tutorials, 1–1. http://doi.org/10.1109/COMST.2015.2491361

- Conferences
  - NetMob http://netmob.org
  - NetSci http://netsci2016.net
- Libraries
  - http://bandicoot.mit.edu an open-source python toolbox to analyze mobile phone metadata
  - igraph http://igraph.org (python, R, C)
  - NetworkX https://networkx.github.io (python)

Thank you!

estebanmoro.org

@estebanmoro