The emergence and evolution of strong and weak ties in social networks

Raffaella Burioni

Dipartmento di Fisica e INFN, Università di Parma

SIF Roma - September 2015

Joint work with

Marton Karsai - ENS - Lyon

Nicola Perra - Mobs Lab - Boston Northeastern

Enrico Ubaldi - Ph.D School - University of Parma

Alessandro Vezzani - CNR Nanoscience - Modena

Alessandro Vespignani - Mobs Lab Boston Northeastern

arXiv:1509.04563

"Asymptotic theory for the dynamic of networks with heterogenous social capital allocation"

models and theory in social systems?

“Par exemple, en mécanique, on néglige souvent le frottement (...). Vous, vous regardez les hommes comme infiniment égoïstes et infiniment clair-voyants. La première hypothèse peut être admise dans une première approximation, mais la deuxième nécessiterait peut-être quelques réserves.”

“For instance, friction is often neglected in mechanics (...). In your case, you consider men as infinitely selfish and infinitely clairvoyant. The first assumption may be accepted as a first approximation, but the second may call for some reservations”.

H. Poincarè, 1901, A letter to Léon Walras to “Economique et Méchanique”

- Regularities exist in large populations of social agents and in many cases they can be predicted, at least “on the average”. Forecast

models and theory in social systems?

- the "prediction" path in Statistical Physics:

Microscopic theory

- elementary interactions known
- relevant variables suggested by physical principles
- scales separations (time, energy, space)
  suggesting the right coarse graining
- Hamiltonian
- You can also be "a little" wrong

models and theory in social systems?

- the "prediction" path in Statistical Physics:

Microscopic theory

Macroscopic behavior

- elementary interactions known
- relevant variables suggested by physical principles
- scales separations (time, energy, space) suggesting the right coarse graining
- Hamiltonian
- You can also be "a little" wrong

→

Average behavior
(equilibrium,
also non equilibrium
and irreversible processes)

models and theory in social systems

- Social systems:

Microscopic theory

- elementary interactions not known in principle

- relevant variables?

- heterogeneity

- scales separations (time, energy, space)?

...

models and theory in social systems

- Social systems:

Microscopic theory

- elementary interactions not known in principle
- relevant variables?
- heterogeneity
- scales separations (time, energy, space)?

...

Datasets + models ⟶ theory and simulations
(hoping that if you are a little wrong it will work - Universality)

models and theory in social systems

- Social systems:

Microscopic theory

- elementary interactions not known in principle
- relevant variables?
- heterogeneity
- scales separations (time, energy, space)?

...

Datasets + models ⟶ theory and simulations
(hoping that if you are a little wrong it will work - Universality)

models and theory in social systems

- Social systems:

Microscopic theory → Macroscopic behavior

- elementary interactions not known in principle
- relevant variables?
- heterogeneity
- scales separations (time, energy, space)?

...

→ Average behavior

Datasets + models ⟶ theory and simulations
(hoping that if you are a little wrong it will work - Universality)

models and theory in social systems

- Social systems:

- large dataset

- modeling is a very interdisciplinary task, with no borders

- models "should be as simple as possible, but not simpler"

Very often it works! How?

an example in asymptotic theory for Time evolving networks

A"forecast" path in asymptotic networks evolution

Graphs, networks and time varying networks

Studying network evolution with "strong" and "weak" links:
Micro: A simple yet powerful model for ties reinforcement (memory) effects in the evolving networks, from dataset

Formulating and solving the analytic model:
predicting the asymptotic of the average network evolution in presence of reinforcement effects

Macro. Check analytical results vs simulations and extensive real dataset:

Forecast

The topology of interactions: networks and graph theory

i   sites, spins, fields, neurons, pc's, websites, agents, countries..

(i,j)   Links, interactions, couplings, hopping parameters, synapses, chemical bonds, routes, friendships, trades,

The most general and simple way of representing the topology of relations and interactions. On each site, there is a static or a dynamical variable, coupled to its neighbors through the links.

We are interested in
- the topological and metrical properties of the network
- the spatial and temporal behavior of the processes taking place on the network
- the link between these two .

The number of sites can also be very large, so that a statistical physics approach can be helpful

The topology of interactions: networks and graph theory

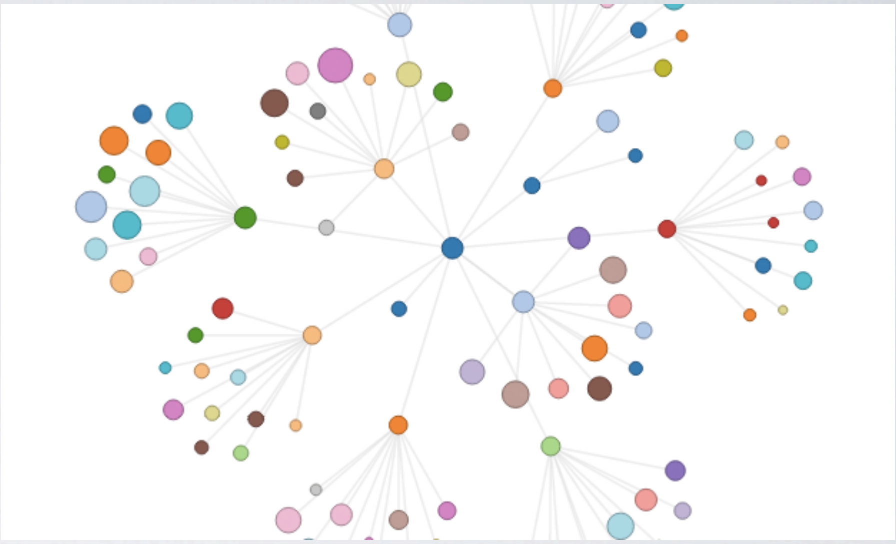
i — sites, spins, fields, neurons, pc's, websites, agents, countries..

(i,j) — Links, interactions, couplings, hopping parameters, synapses, chemical bonds, routes, friendships, trades,

- Equilibrium properties of statistical mechanics models on the network:
  symmetry breaking, multiple phases, phase diagrams, phase transitions, critical phenomena,...

- Non equilibrium and dynamical properties:
  response functions, classical and quantum transport, reaction and diffusion, spreading, synchronization,..

The topology of interactions: networks and graph theory
time varying networks



A recent new perspective on time scales of interactions

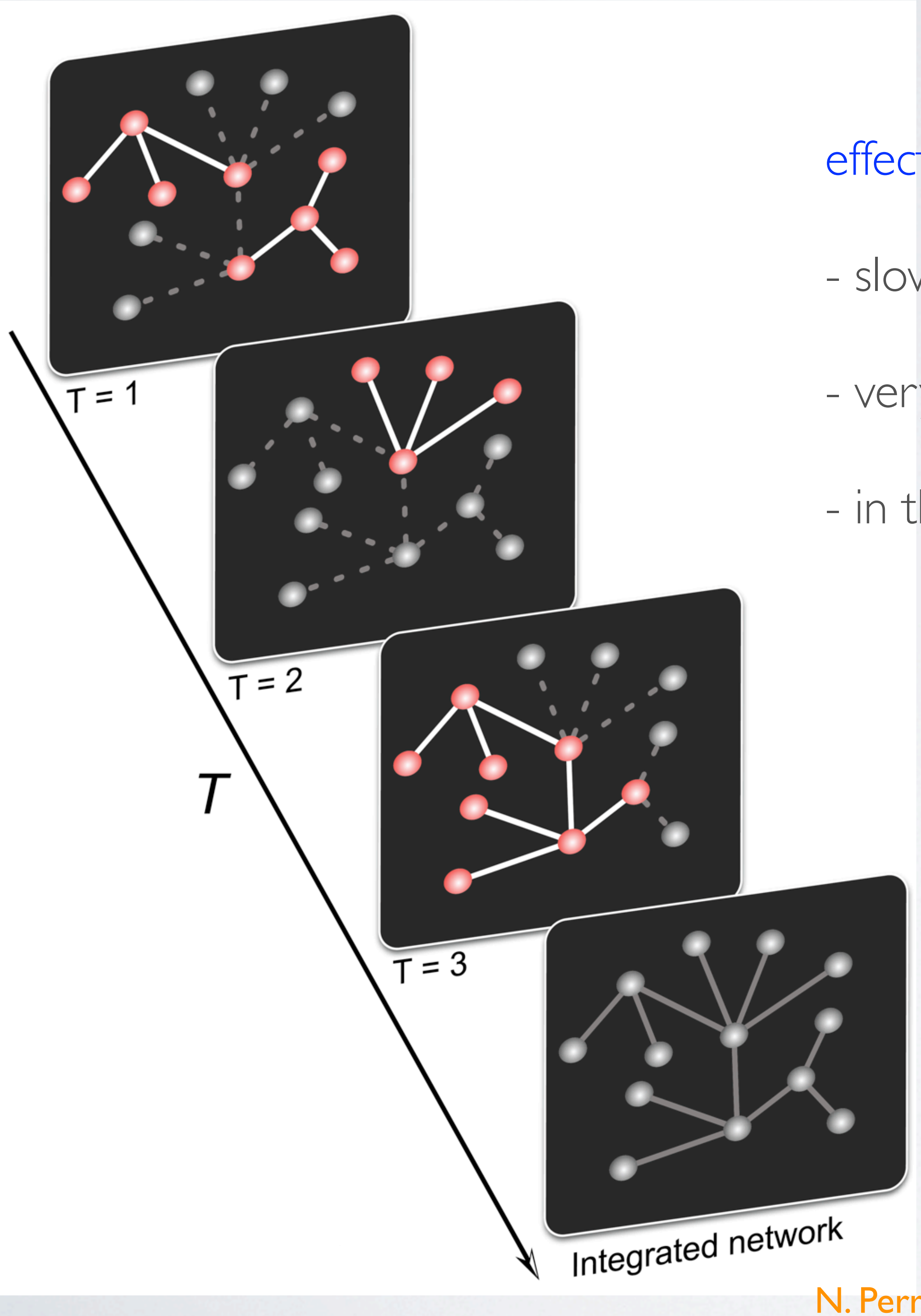
Networks are often dynamical in nature

The dynamical scale of the processes taking place on the network
are often comparable to the formation and evolution of the network itself.

The modeling of the evolving network is crucial to understand the dynamics
on the network

Many open problems, strong research: a mathematical framework is still missing

book: "Temporal Networks", Springer, (2013). P. Holme, J. Saramaki Eds
Scholtes et al (2013), Barrat et al, 2013, Lambiotte et al (2014), Holme (2015)



effects of timescales

- slow network dynamics: static picture

- very fast network dynamics: effective random coupling

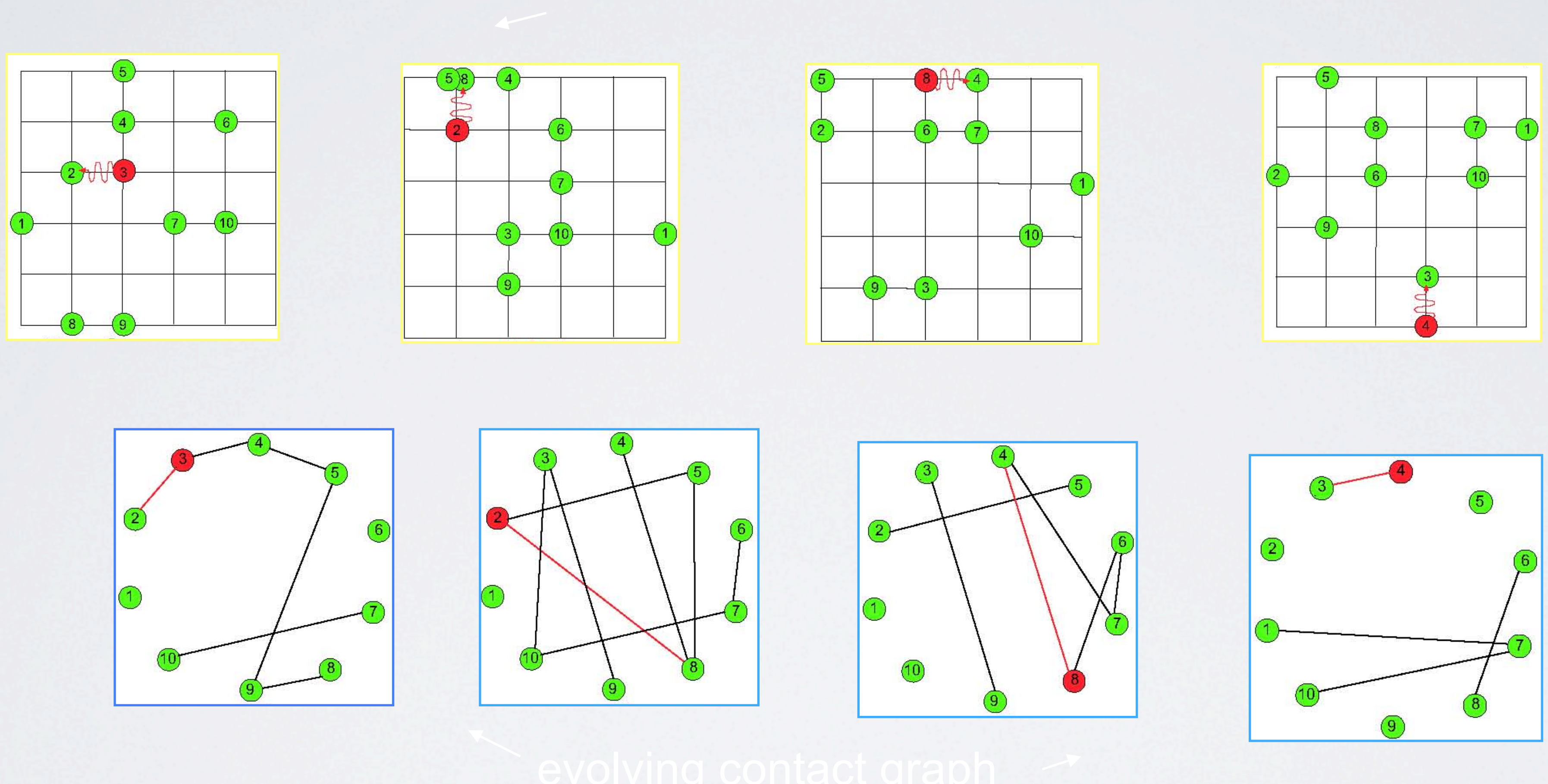
- in the middle: the most interesting and complex case

how it evolves? can we forecast
the evolution by looking at specific
micro properties?

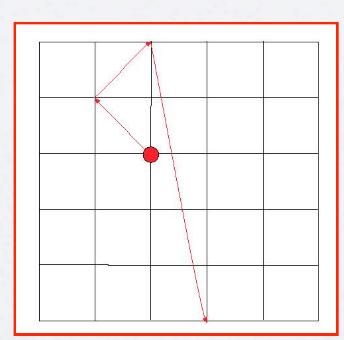
N. Perra, B. Goncalves, R. Pastor Satorras, A. Vespignani SciRep (2012)

Time Varying Networks in a reaction-diffusion problem

A dynamic contact network generated by diffusing particles + the diffusion of an excitation
Energy transfer processes



Excitation RW

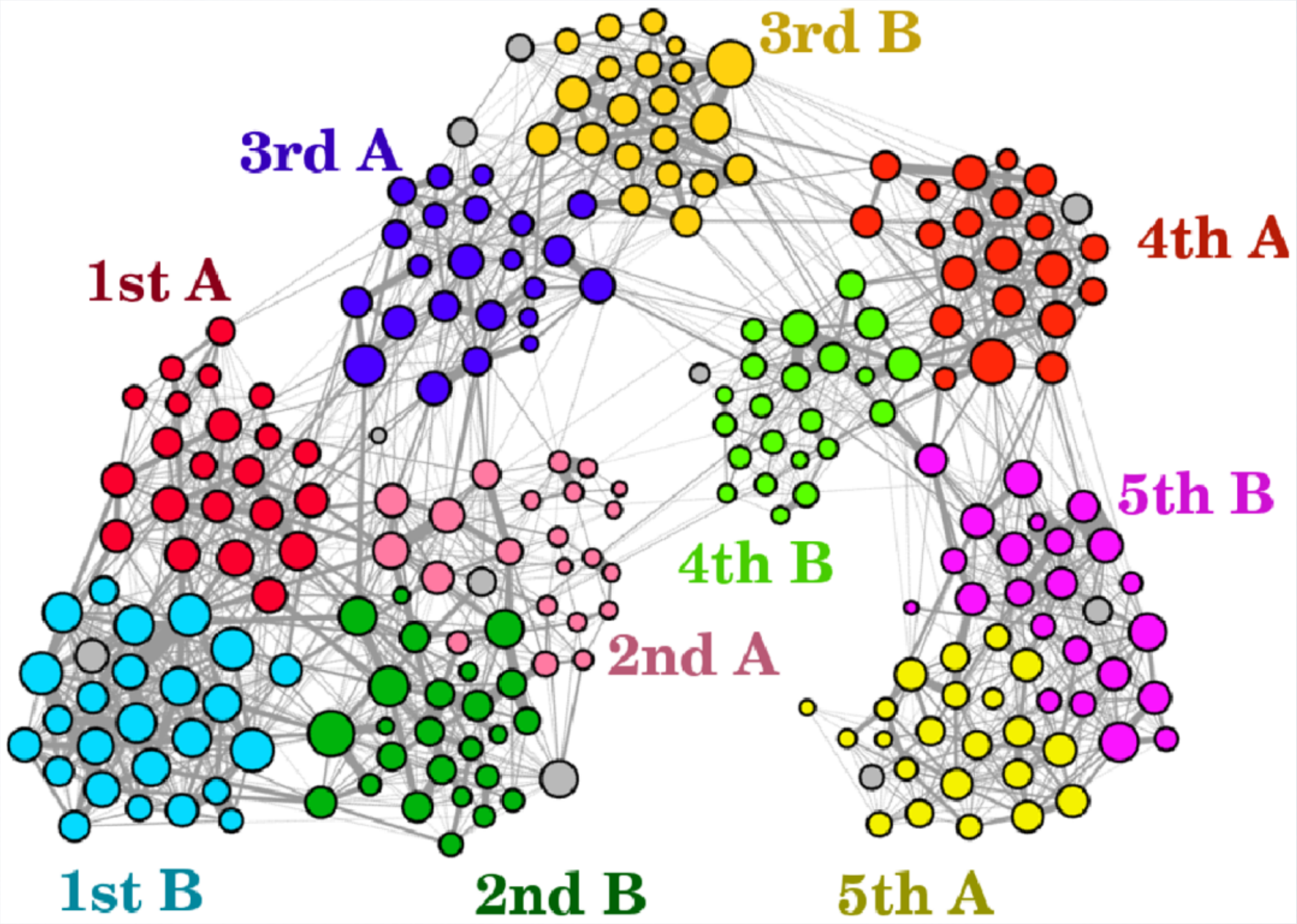


E. Agliari, R. Burioni, D. Cassi (2007, 2010, 2014)

Time Varying Networks in a social system

A dynamic contact network generated by moving agents: Sociopatterns agents wear GPS and they are linked when they are near in space

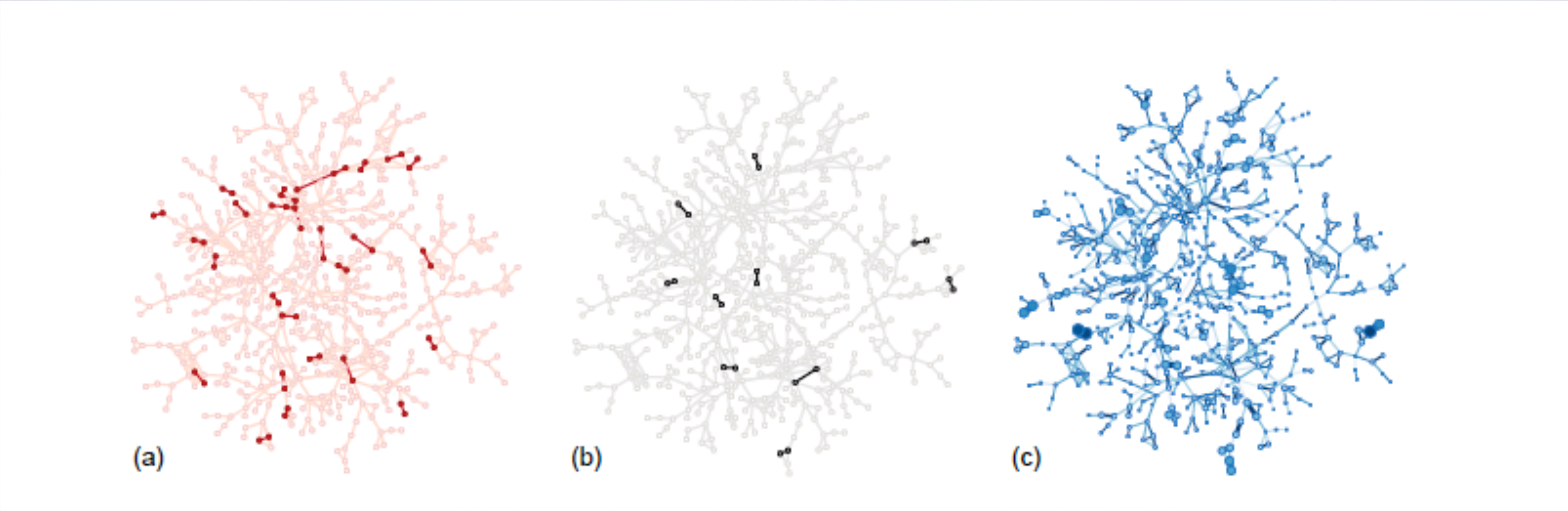
A. Barrat and C. Cattuto' s groups ISI



www.sociopatterns.org

Networks Evolution

Time varying vs integrated network: how the network evolves?

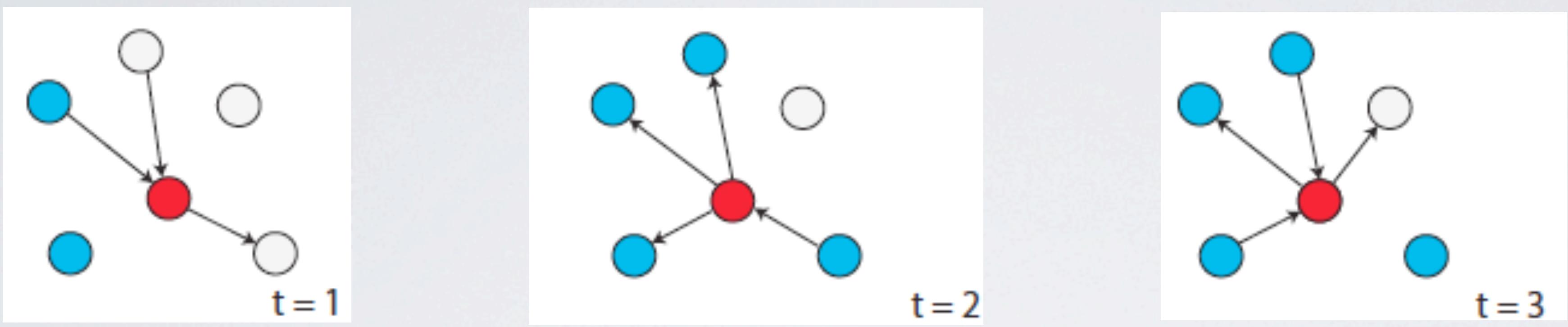


a snaphot

the final network

M. Karsai, N. Perra, A. Vespignani SciRep (2014)

Networks Evolution micro: how links grow

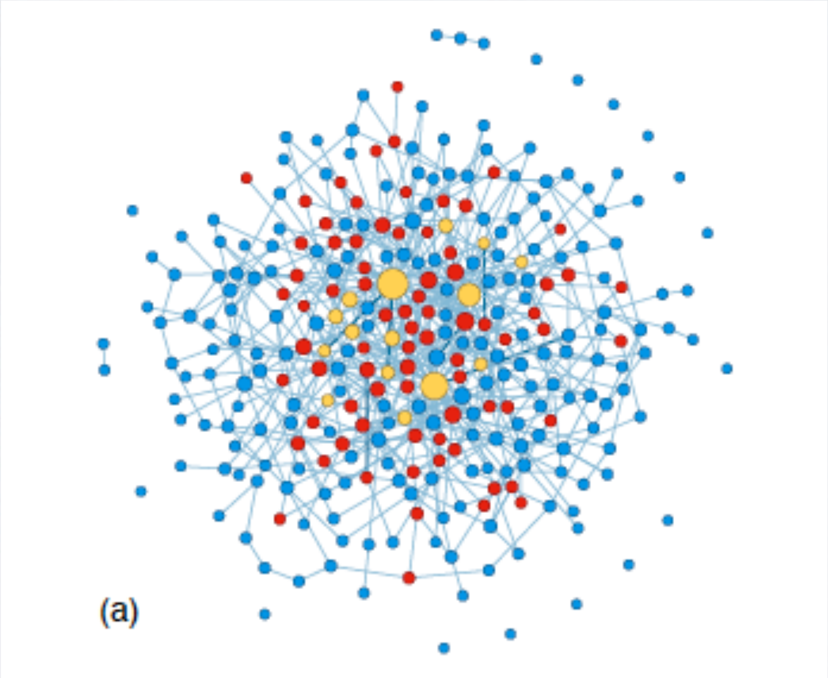


an important point in growth: strong vs weak ties

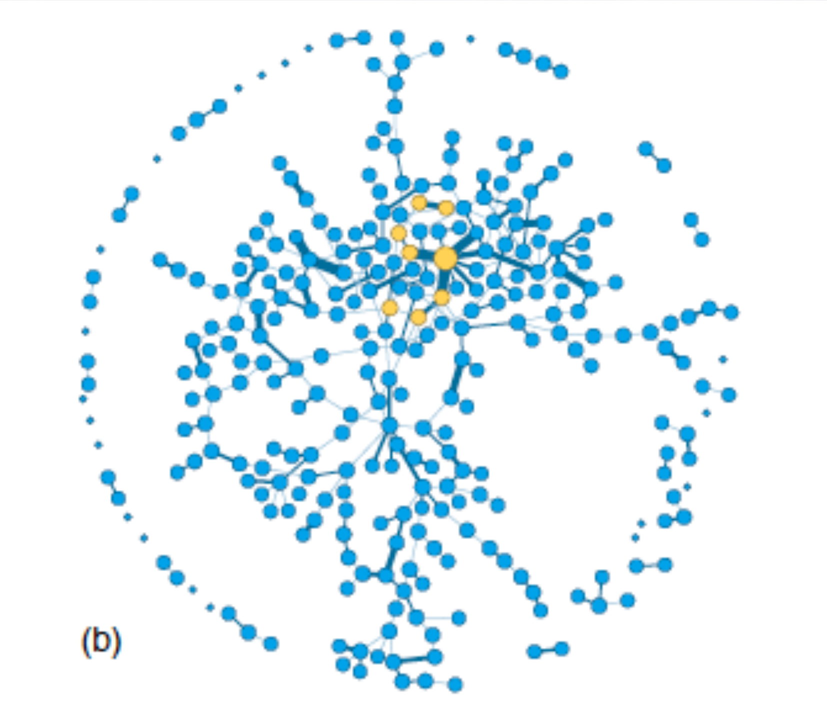
- when you activate a link, use an old link (make it stronger) or create a new one?
- can we define a probability for such events?
- what are the relevant variables that rules this probability?

Networks Evolution macro: it matters!

an "open" (less reinforced) network



a "closed" network



M. Karsai, N. Perra, A. Vespignani SciRep (2014)

Networks Evolution: the reinforcement process

A micro "memory" cost for new links attachment
measurements suggest a zero-hyp:
the relevant variable is the degree k (i.e. the total number of link) of that node at time t.

Each node has a probability to create a new random link that depends on its degree,
with a very simple form, that captures a crucial point:
adding new links costs, if you already have many.

A simple form: prob for node i to go from k to k+1

$$p_i(k) = (\frac{1}{1 + \frac{k}{c_i}})^{\beta_i}$$

prob to keep k links
and to contact an old node

$$1 - p_i(k)$$

Strong vs weak ties, in a very simplified form: beta and c, the parameters.
Distributed, data suggested and measurable from data

M. Karsai, N. Perra, A. Vespignani SciRep (2014)

E. Ubaldi,, N. Perra, M. Karsai, A. Vezzani, R. Burioni, A. Vespignani, (2015)

Networks Evolution: time scales and activity

Activity driven networks

the "nodes" of the growing network are characterized by
the number of actions (link attachment in this case) they perform in unit time. a

The activity distributions is measurable and, interestingly,
largely independent of the chosen time window.
In general, it is broadly distributed

$F(a) \sim a^{-\nu}$ at large a

$\nu \sim 2, 3$

N. Perra, B. Goncalves, R. Pastor Satorras, A. Vespignani SciRep (2012)

Networks Evolution: 7 datasets

- APS co-authorship network, Phys. Rev. A, B, D, E, L from 1st edition to 2007;

Link: collaboration

- Twitter firehose 01-09/2008 (536k users);

Link: twitter mention

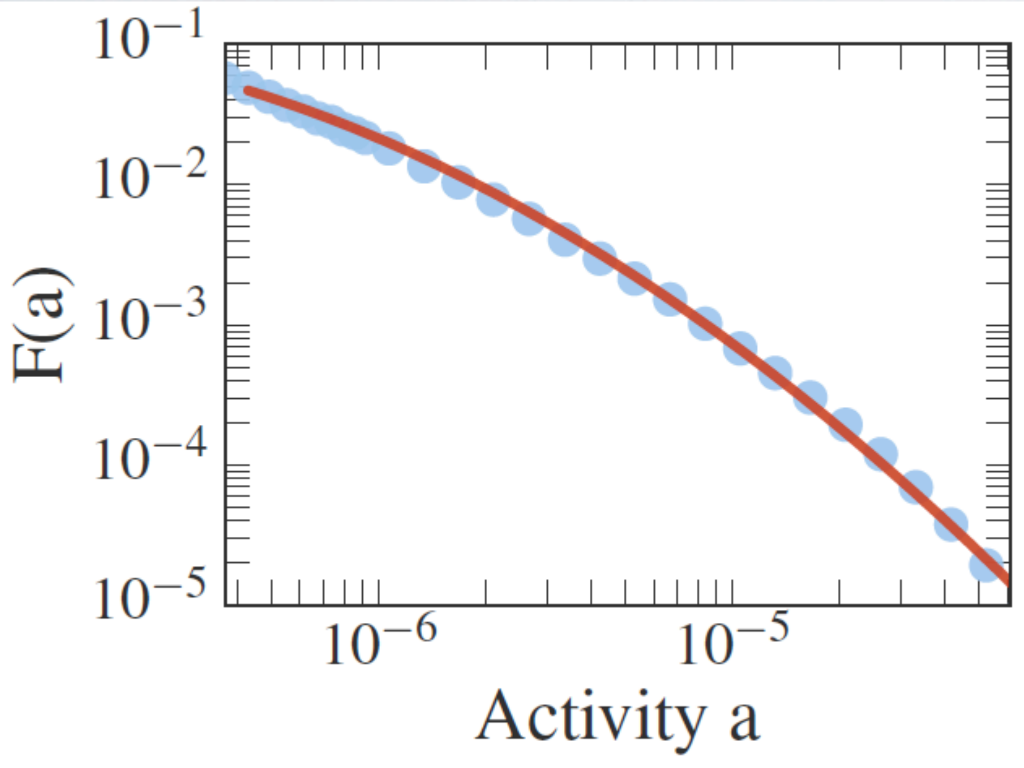
- Mobile Phone Call (6.7 million users, 7 months);

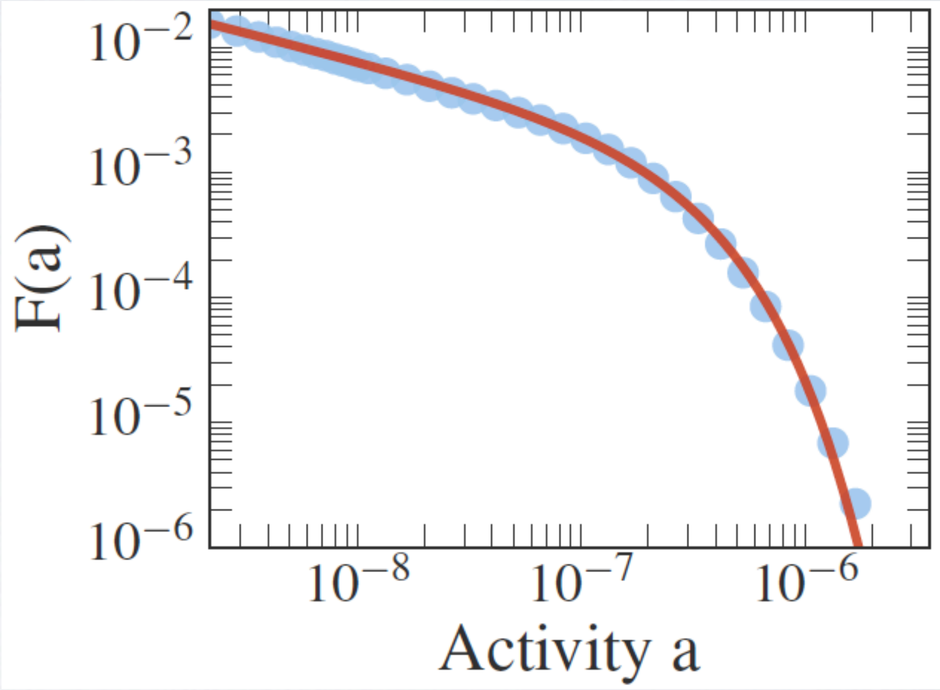
Link: phone call

Caller_ID	Called_ID	Company_Caller	Company_Called	# Event 0
Caller_ID	Called_ID	Company_Caller	Company_Called	# Event 1
Caller_ID	Called_ID	Company_Caller	Company_Called	# Event 2
. . . .	. . . .	. . . .	. . . .	. . . .

Networks Evolution: measuring micro activity parameters

Activity distributions: Fits from data and measure of ν





Truncated power law for MPC, APS

Lognormal for TWT

$F(a) \sim a^{-\nu}$ for large a

Maximum likelihood fits, Newman et al 2009, Alstott et al 2014

Networks Evolution: measuring micro reinforcement parameters

The distributions of betas and c's must be measured from real datasets and represents the microscopic input of the model, together with the activity distribution.

$$p_i(k) = (\frac{1}{1+\frac{k}{c_i}})^{\beta_i}$$

- A clever and complex averaging procedure, grouping nodes in activity classes

- Measure from large datasets

- How to use the parameters in the model

Networks Evolution: measuring micro reinforcement parameters

Results from dataset at the "microscale"

- the form of the reinforcement is simple but universal and works for all datasets

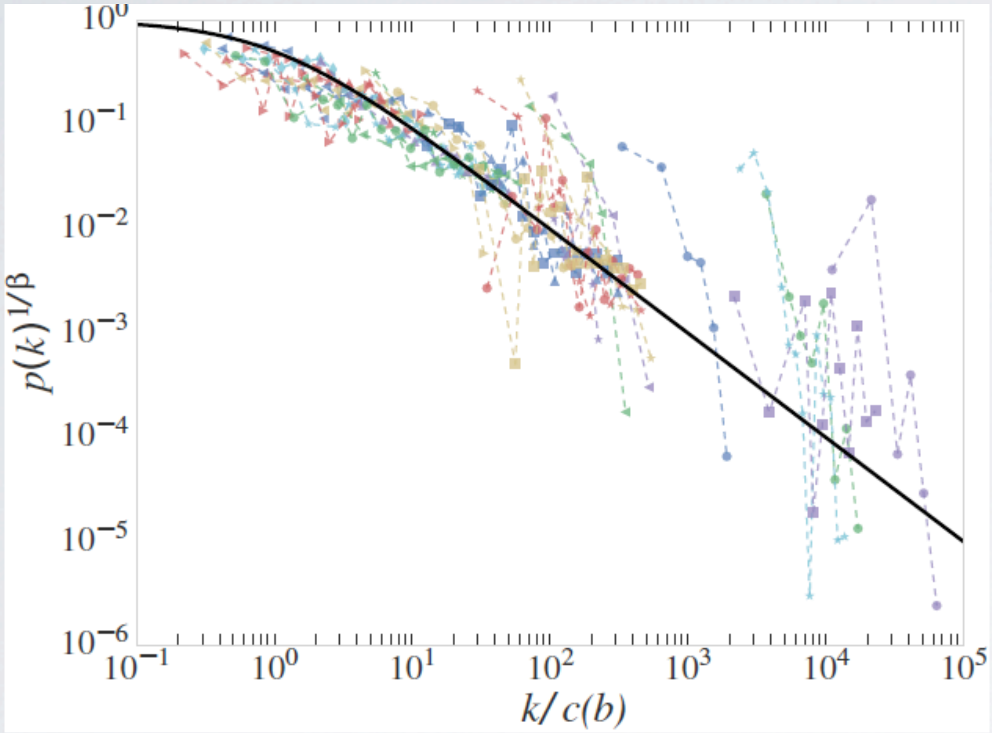
- the exponent beta has a measurable well peaked distribution

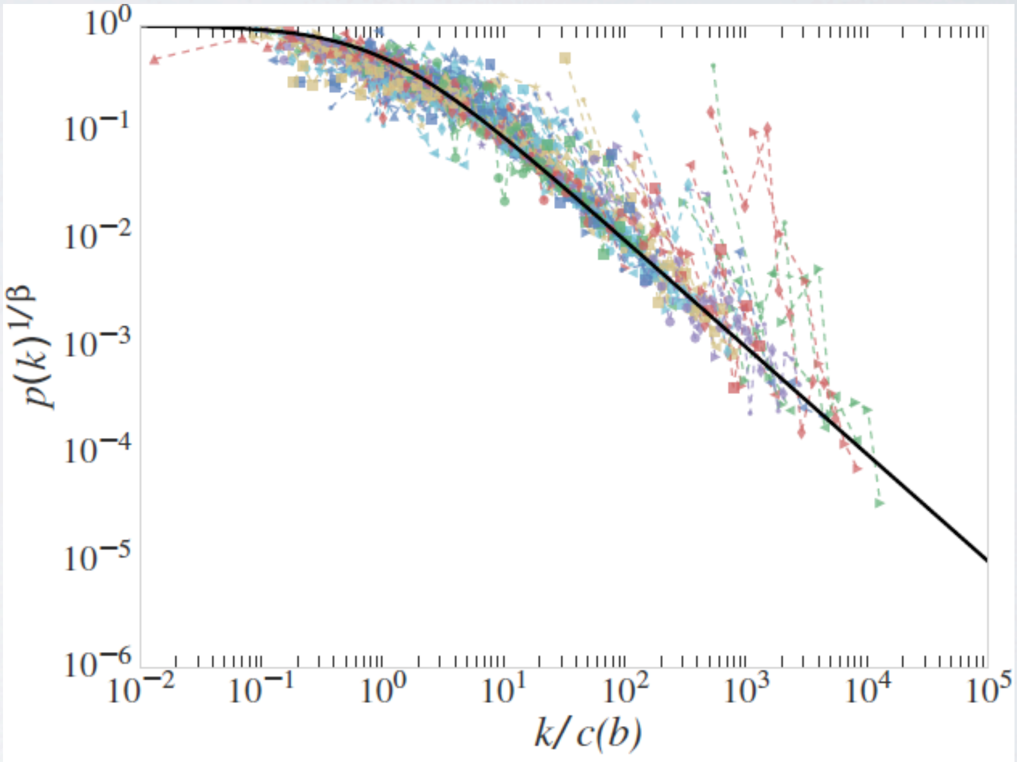
- also the coefficient c are distributed but very well peaked

- we can associate to each dataset a single value (average) of the reinforcement parameter.
  As we will see from the analytics,
  this is the relevant information for the description of the large scale evolution of the network

- Two parameters: activity distribution exponent and the average reiforcement exponent

Networks Evolution: measuring micro reinforcement parameters





APS (PRL) beta = 0.16
TWT beta = 0.5

Ex: The rescaled reinforcement probability for two dataset (a complex measure on real dataset)

Networks Evolution: analytics

- We can write and solve asymptotically at large t and large number of nodes N the master equation of the stochastic process and get the exact asymptotic scaling form for probability distribution for a node of activity a to have degree k at time t. $P(a, k, t)$

The scaling form agrees extremely well with the dataset

From this solution we obtain, as a function of the memory and activity parameters

- The growth of the average degree of the evolving network with time
- The form of the integrated degree distribution

E. Ubaldi,, N. Perra, M. Karsai, A. Vezzani, R. Burioni, A. Vespignani, (2015)

The analytic result

sketch of the analytic calculation

$$P(a_i,k,t+1) - P(a_i,k,t) = -\frac{a_i}{k^\beta}\left(P(a_i,k,t) - P(a_i,k-1,t)\right) + a_i \frac{(k+1)^\beta - k^\beta}{k^\beta (k+1)^\beta} P(a_i,k,t) - \left(P(a_i,k,t) - P(a_i,k-1,t)\right) \sum_j{}' a_j \sum_h \frac{P(a_j,h,t)}{(N-h)(h+1)^\beta}$$

large t + continuum limit + 1 << k <<N

$$\rho(a) \sim a^{-\nu}$$

$$\frac{\partial P}{\partial t} = -\frac{a}{k^\beta}\frac{\partial P}{\partial k} + \frac{a}{2k^\beta}\frac{\partial^2 P}{\partial k^2} + \frac{a\beta}{k^{\beta+1}} P(a,k,t) + \left(\frac{1}{2}\frac{\partial^2 P}{\partial k^2} - \frac{\partial P}{\partial k}\right) \int da\,\rho(a)\,a \int dh \frac{P(a,h,t)}{h^\beta}$$

good ansatz + careful estimate of the terms

The analytic result

A summary of analytic results

$$p(k) \sim (\frac{1}{1+k/c})^{\beta} \qquad \rho(a) \sim a^{-\nu}$$

↓

$$P(a,k,t) = \exp\left(-A\frac{\left(k - C(a)t^{\frac{1}{1+\beta}}\right)^2}{t^{\frac{1}{1+\beta}}}\right)$$

$$\frac{C(a)}{1+\beta} = \frac{a}{C(a)^{\beta}} + \int \frac{a\rho(a)da}{C(a)^{\beta}}.$$

$$C(a) \sim a^{1/(1+\beta)}$$

$$\langle k \rangle \simeq C(a) \cdot t^{1/(1+\beta)}$$

average degree

$$\rho(k) \sim k^{-((1+\beta)\nu - \beta)}$$

integrated degree distribution

The analytic result

A summary of analytic results   integrated degree distribution

Given the form of the activity distribution and the value of the reinforcement parameter, we can forecast the form of the degree distribution for any activity distribution

PDF	$F(a)$	$\rho(k)$
Power Law	$a^{-\nu}$	$k^{-[(1+\beta)\nu-\beta]}$
Stret. Exp.	$a^{\nu-1}\exp\left[-\lambda a^{\nu}\right]$	$k^{[(1+\beta)(\nu-1)+\beta]}\exp\left[-\tau k^{(1+\beta)\nu}\right]$
Trunc. PL	$a^{-\nu}\exp\left[-\lambda a\right]$	$k^{-[(1+\beta)\nu-\beta]}\exp\left[-\tau k^{(1+\beta)}\right]$
Log-Normal	$\frac{1}{a}\exp\left[-\frac{(\ln(a)-\mu)^2}{2\sigma_a^2}\right]$	$\frac{1}{k}\exp\left[-\frac{(\ln(k)-\gamma)^2}{2\left(\frac{\sigma_a}{1+\beta}\right)^2}\right]$

The analytic result

An insight in the case without memory from our equations: let us put beta=0

$$P(a,k,t) = (2\pi(a+\langle a\rangle)t)^{-\frac{1}{2}}\exp(-\frac{(k-(a+\langle a\rangle)t)^2}{2t(a+\langle a\rangle)})$$

$$P(a,k,t) \sim \delta(k-(a+\langle a\rangle)t)$$

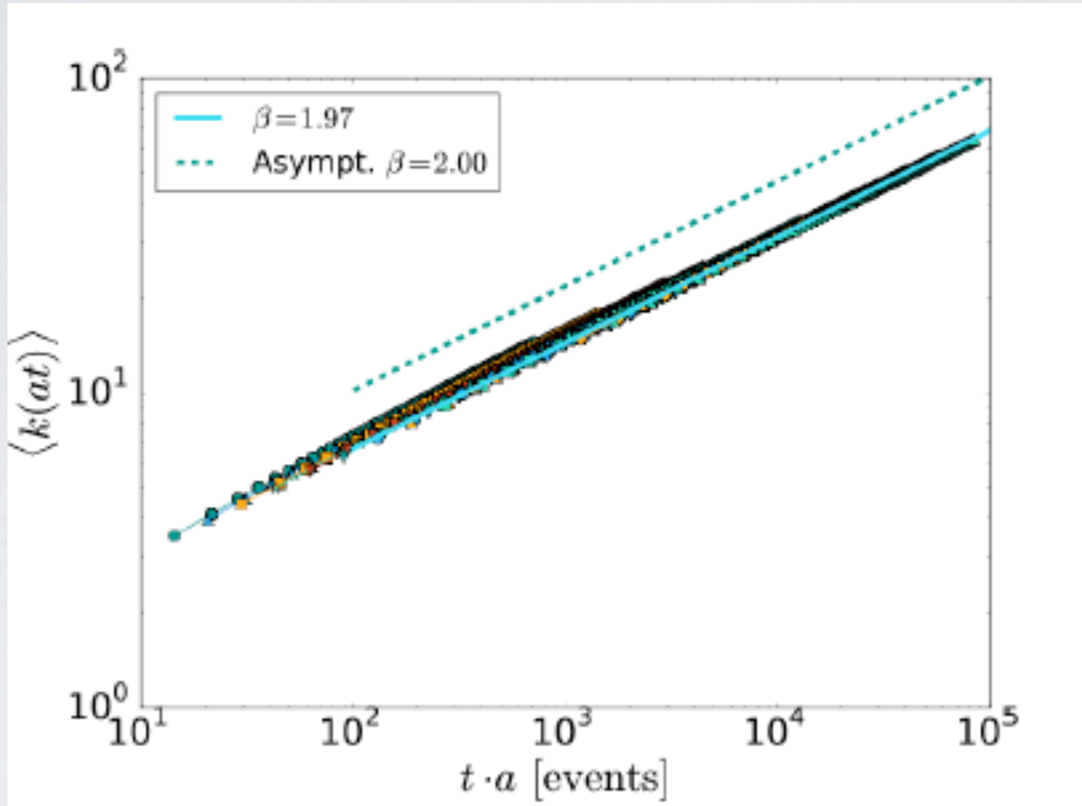
large times

$$< k >= (a+\langle a\rangle)t$$

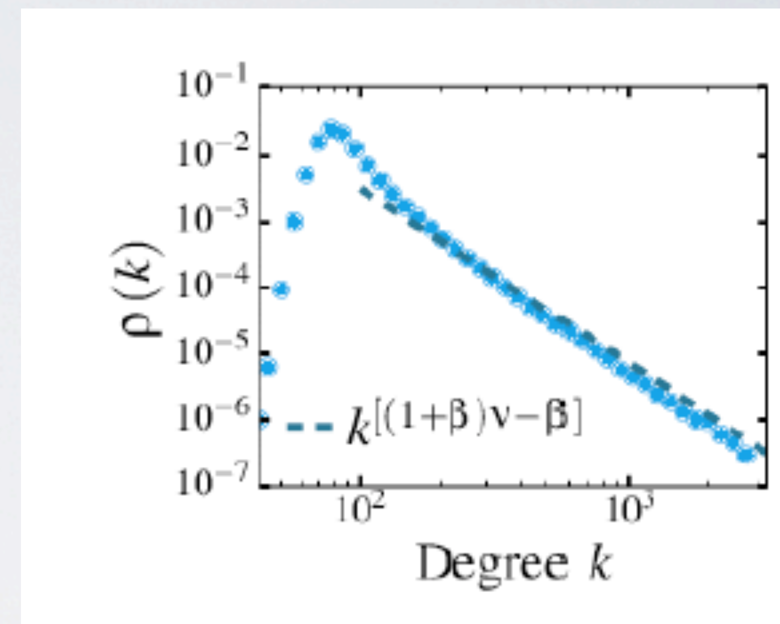
Starnini, Pastor Satorras 2013

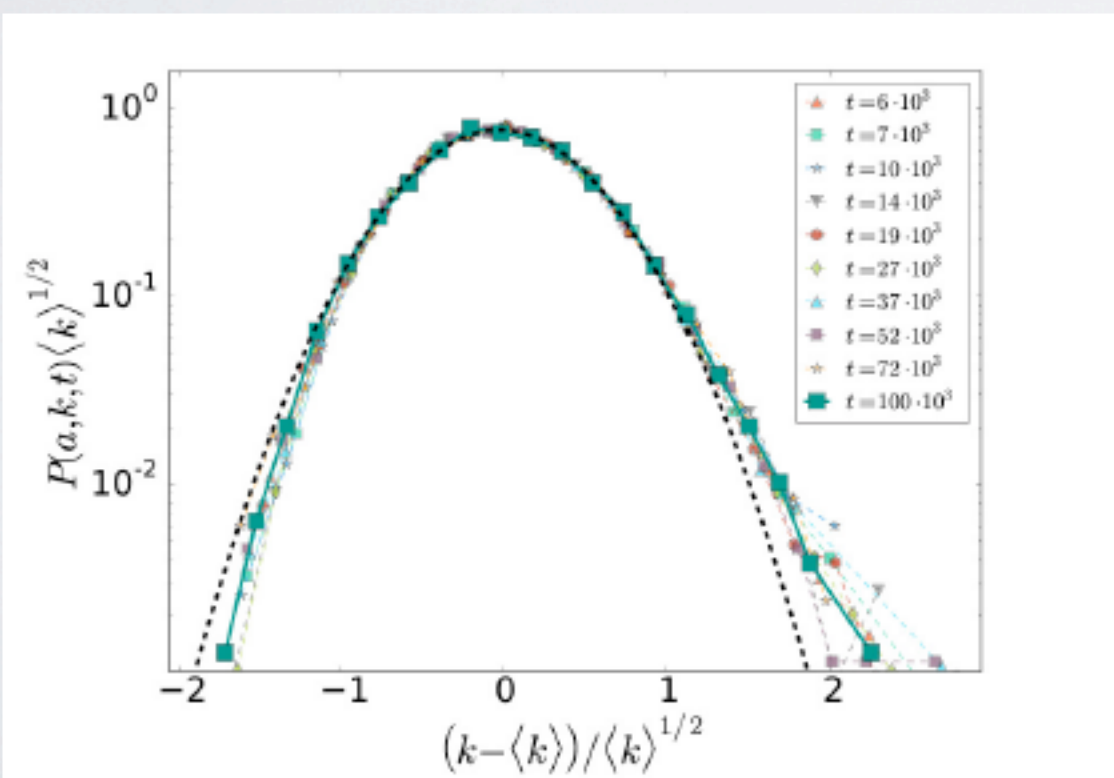
Checks

Simulations



$$\langle k \rangle \simeq C(a) \cdot t^{1/(1+\beta)}$$

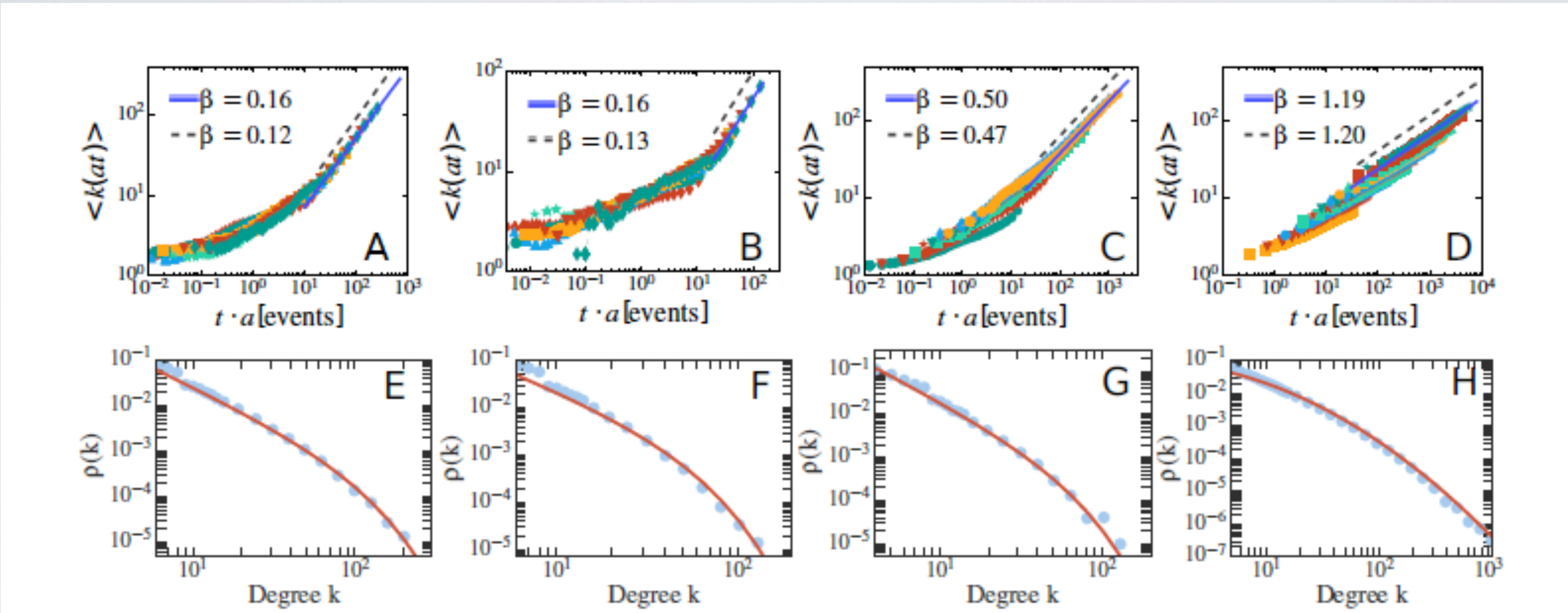




$$P(a,k,t) = \exp\left(-A\frac{\left(k - C(a)t^{\frac{1}{1+\beta}}\right)^2}{t^{\frac{1}{1+\beta}}}\right)$$

and Real data!

MPC



$$\langle k \rangle \simeq C(a) \cdot t^{1/(1+\beta)}$$

Blue fit, dashed analytics, different colors are different activity classes (same growth!)

$$\rho(k) \sim k^{-((1+\beta)\nu-\beta)}$$

Red analytics, blue points data

- Variables: Activity, reinforce

- Measure the micro "reinforcement" and the activity from large statistics

  = get the large scale evolution of the network and the distributions in very different datasets. Universal Mechanism

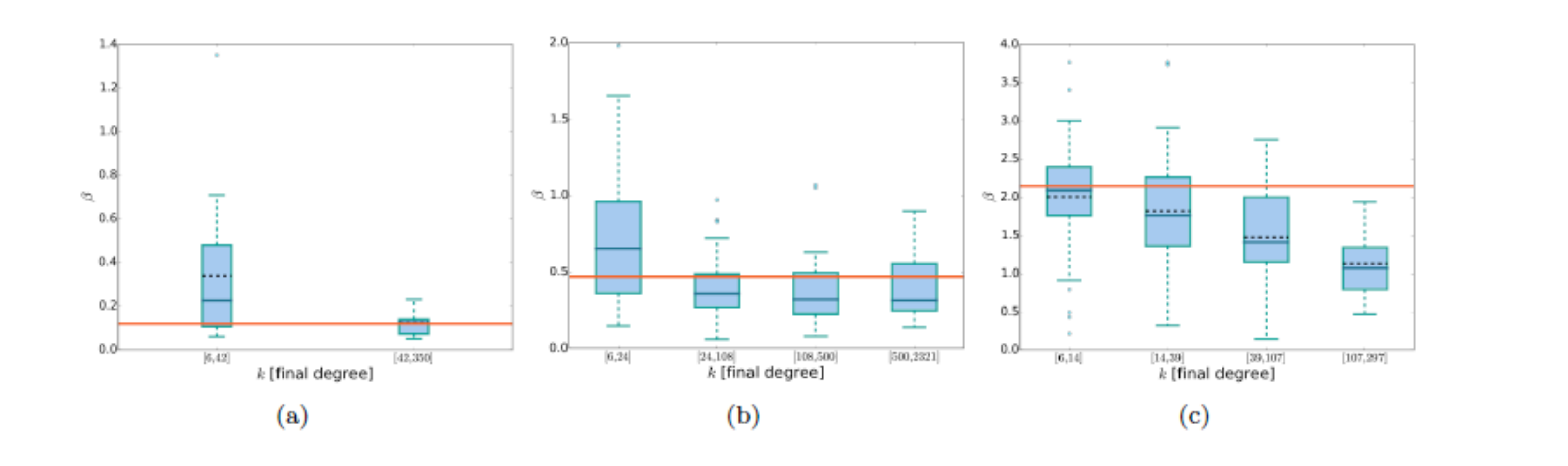
- A first step in the description of the network growth: the backbone
- Evolution of reciprocated ties vs simple ties?
- Wide distribution of intertime events
- forecast for the networks in the transient (a lot of work going on)

- Models: Would it be possible to measure the memory in "controlled conditions" in social experiments?

- A simple mechanism for the simple reinforcement functions?
  The "adjacent possible" and the emergence of correlated novelties with V. Loreto and F. Tria.

F. Tria, V. Loreto, V.D.P. Servedio, S. Strogatz 2014 Talk Francesca Tria

Real data: distribution of beta



Real data: distribution of C's