# Old and New Ties in Social Networks Evolution: Novelties Exploration in Human Interactions

Raffaella Burioni

Statistical Physics & Complex Systems
Department of Mathematics, Physics and Computer Science
and INFN, University of Parma - Italy

Roma - 2017

# Joint work with

Enrico Ubaldi - ISI Institute - Torino Italy

Vittorio Loreto - Sapienza Università Roma

Francesca Tria - Sapienza Università Roma

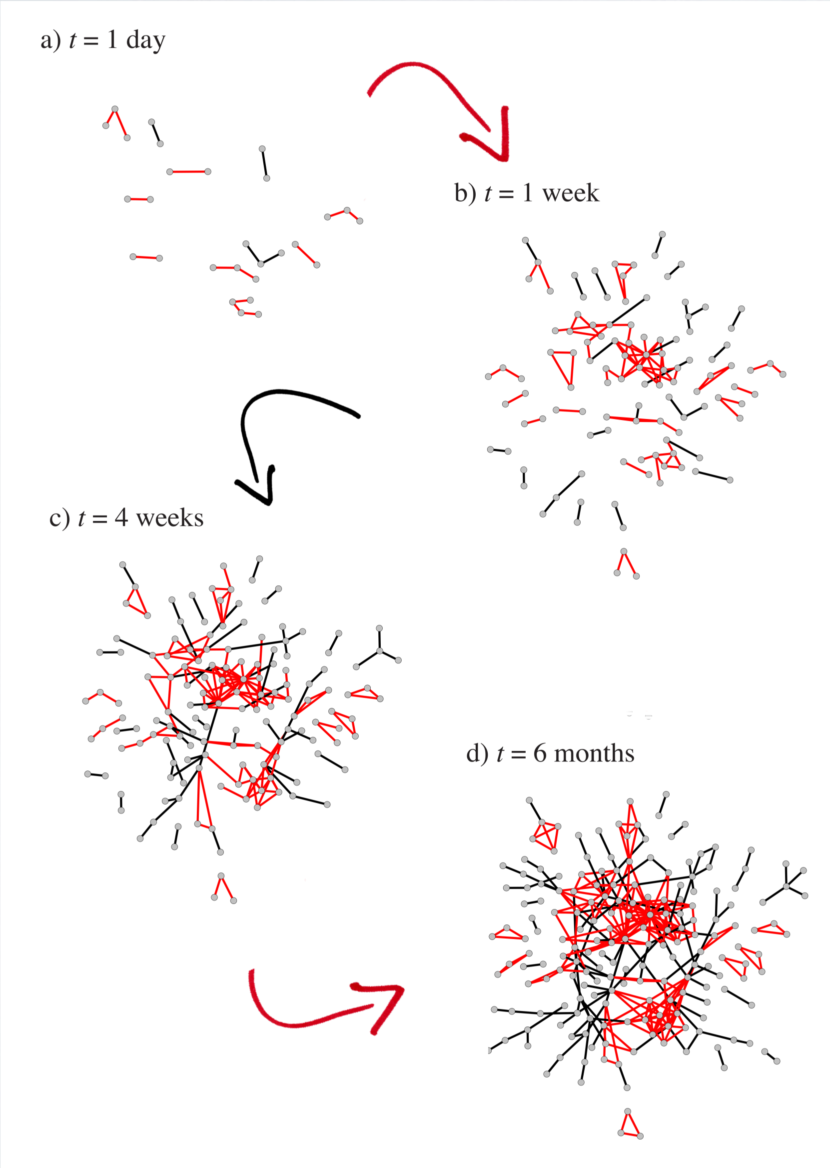
Marton Karsai - ENS - Lyon

Nicola Perra - University of Greenwich - UK

Alessandro Vezzani - IMEM CNR - Parma

Alessandro Vespignani - Mobs Lab - Boston Northeastern

# Time Varying Networks: novelties in social networks




people ○

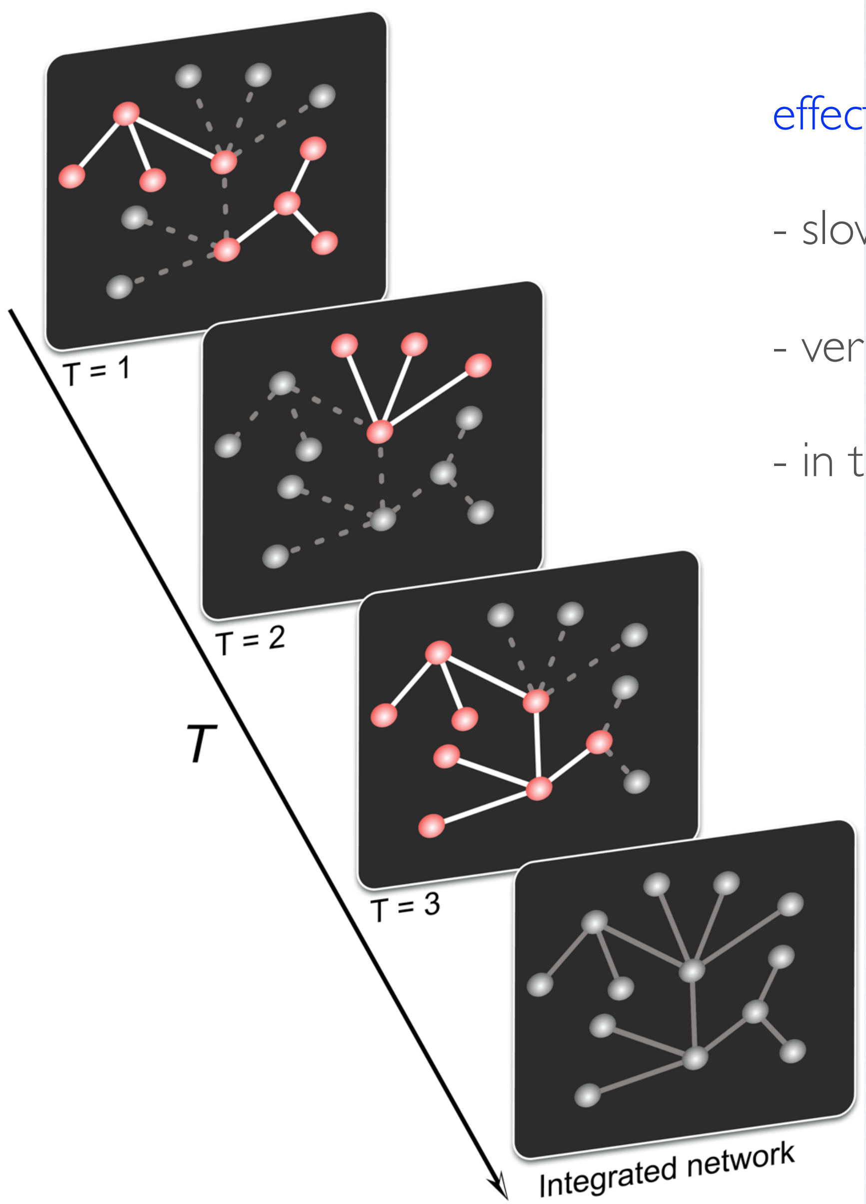
new contact

old contact

Novelties makes Networks dynamical
and put memory in the evolution

Time Varying networks: Many open problems, strong research going on:

"Temporal Networks", Springer, (2013). P. Holme, J. Saramaki Eds




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 some specific properties?

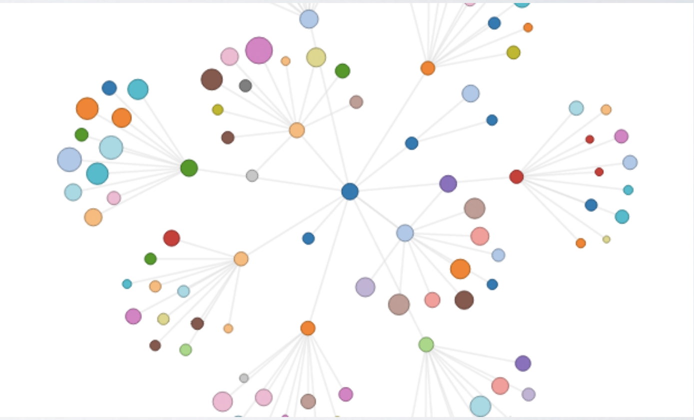
- if there are rules for ties formations, where do they come from?

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

# Time Varying Social Networks and novelties discovery

new vs old: memory effects

- sometimes people contact people from their circle (no new link added)
- sometimes they contact someone new (new link added)



-What are the most important mechanisms driving the evolution, the old/new link attachment rules?

- Is there something similar to a discovery of novelties?

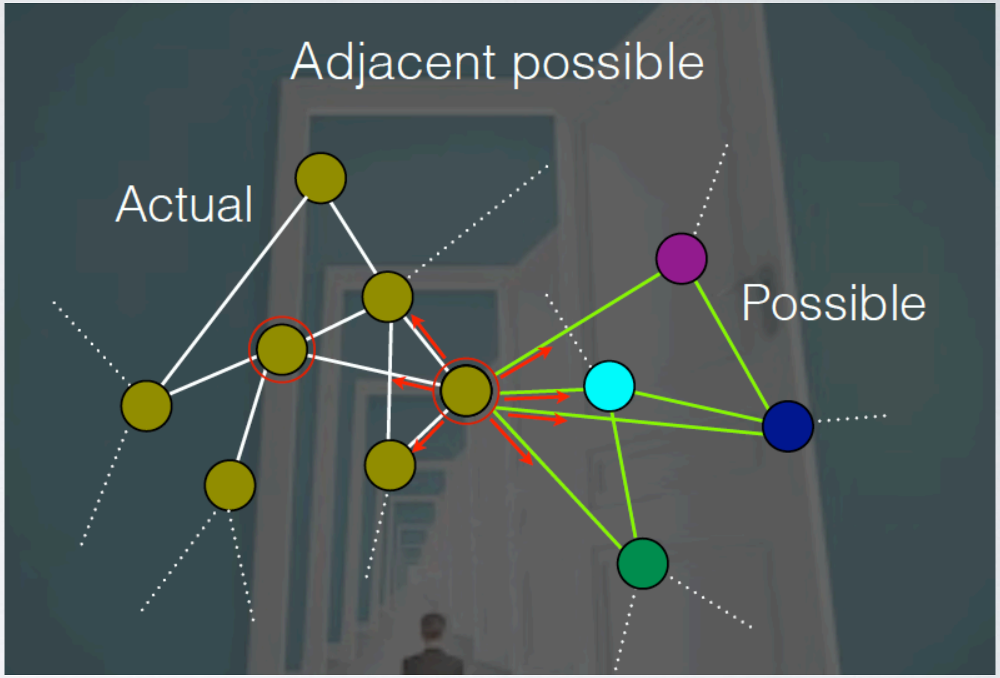
- While the social circle of people we contact enlarges, is there any triggering due to exploration of novelties?

- What are the signature of the adjacent possible here?

# The adjacent possible

We continuously experience novelties
in very diverse scenarios:

Once in contact with such novelties,
as our knowledge expands,
we are naturally setting the stage
for other novelties to be experienced
in the future.
Correlated novelties.




F. Tria

This way of thinking goes under the Kauffman's theory
of the adjacent-possible , which has recently been shown to catch the individuals' knowledge space exploration.

Is this working when looking at the growth of the individual's set of social contacts?

Tria, Loreto, Servedio, S.H. Strogatz (2014)

# Time Varying Social Networks and novelties discovery

Many effects are working when networks grows. We want to

- detect some of the main mechanisms that drives the process of Social Circles Growth
- measure them from large dataset
- use them to build evolution equations and hopefully solve them
- forecast the evolution

Match these mechanism with a theory
and a model of adjacent possible - the Urn model

# Basic mechanisms in social networks evolution

Measure from datasets: two mechanisms

- Activity: i.e rate of link formation from dataset

- Ties selection rules from dataset:
  what is the probability to select an old or a new link?
  To discover a new contact?

+ Burstiness effects in social actions and large distributions of interevent times

R. Burioni, G. Gradenigo, A. Sarracino, A. Vezzani, A. Vulpiani (2013)
E. Ubaldi, N. Perra, M. Karsai, A. Vezzani, R. Burioni, (2017)
R. Burioni, E. Ubaldi, A. Vezzani (2017)

# Networks Evolution I: when links grow

Activity driven networks:

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

The activity distribution is
- measurable
- largely independent of the chosen time window.
- in general broadly distributed. Node i is assigned an activity $a_i$ ,

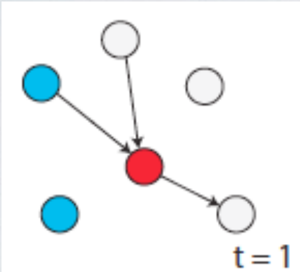
and $a_i \Delta t$ is the prob to get active in $\Delta t$

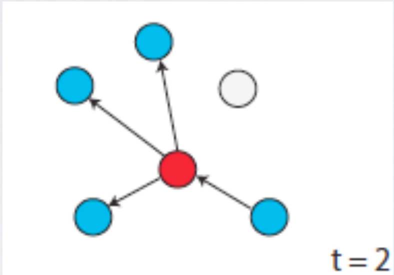
$F(a) \propto a^{-(\nu+1)}$ at large a

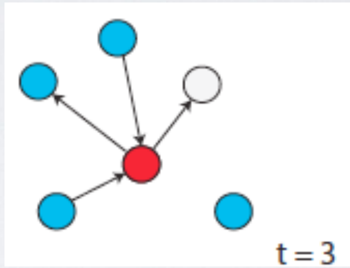
$\nu \sim 1, 2$

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

# Networks Evolution: how links grow





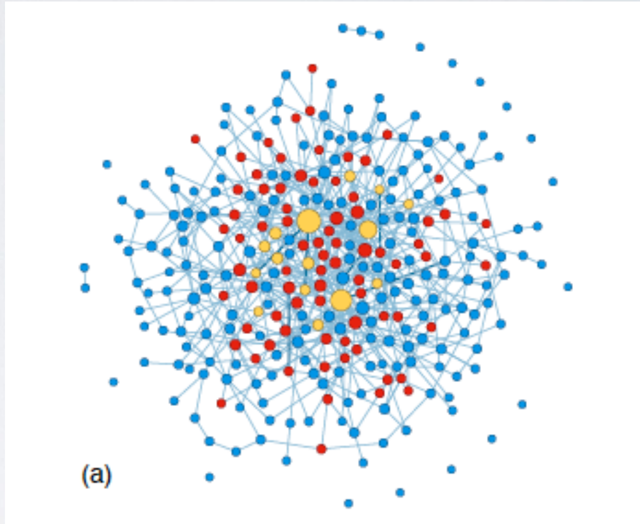



Old or new ties?

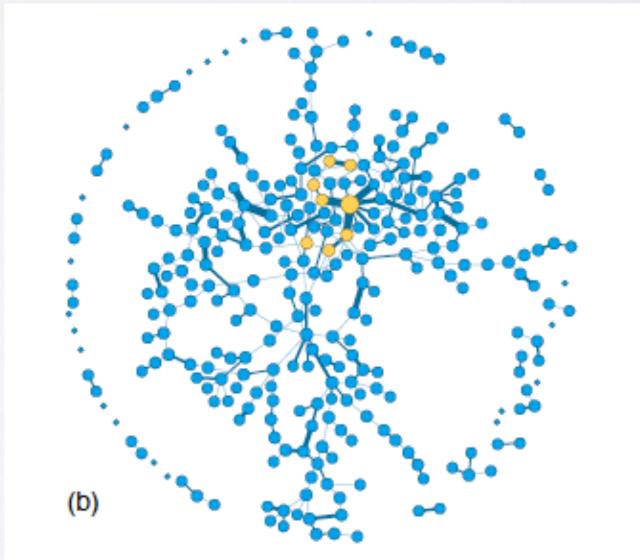
- when you activate a link, use an old link
  or enlarge your social circle, create a new link, contacting a new node?
- can we define a probability for such events?
- what are the relevant variables that rules this probability?

# Networks Evolution: how links grows matters

an “open” (easy to contact new people) network




a “closed” network




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

# Network Evolution: the selection process

Data inspired:

Each node has a probability to create a new random link that depends on its degree, (the number of already contacted nodes up to that time) with a 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 $$1 - p_i(k)$$

Very simplified form: beta and c, the parameters

Distributed, data suggested and measurable from data

E. Ubaldi, N. Perra, M. Karsai, A. Vezzani, R. Burioni, A. Vespignani, (2016)(2017)

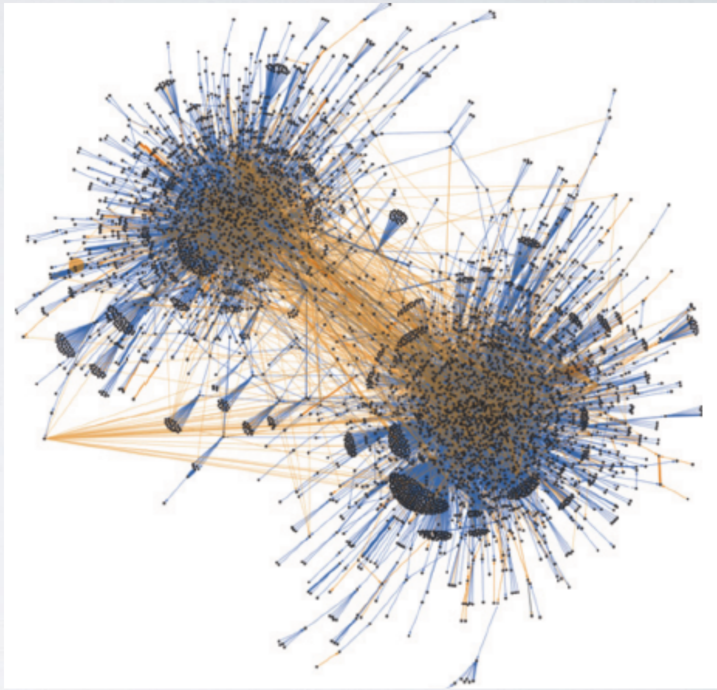
# Networks Evolution: 7 datasets

- APS co-authorship network, Phys. Rev. A, B, D, E, L from 1st edition to 2007;
- Twitter firehose 01-09/2008 (536k users);
- [redacted] Mobile Phone Call (6.7 million users, 7 months);

Link: collaboration

Link: twitter mention

Link: phone call



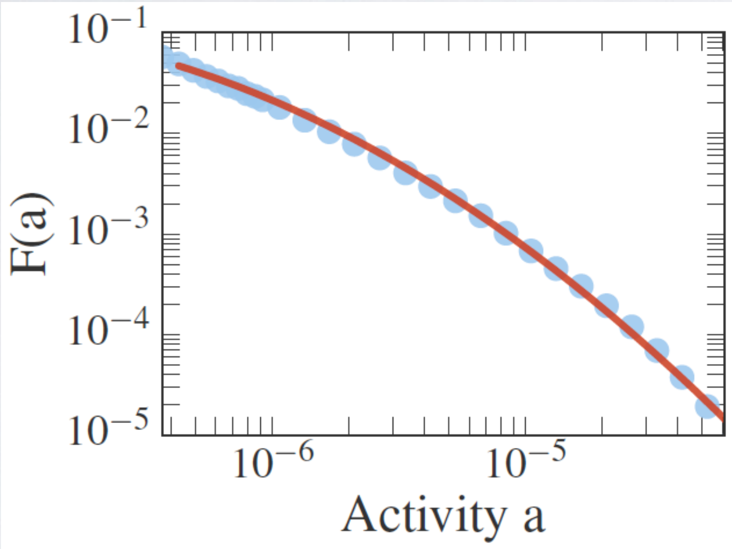
From A. Vespignani, (2012)

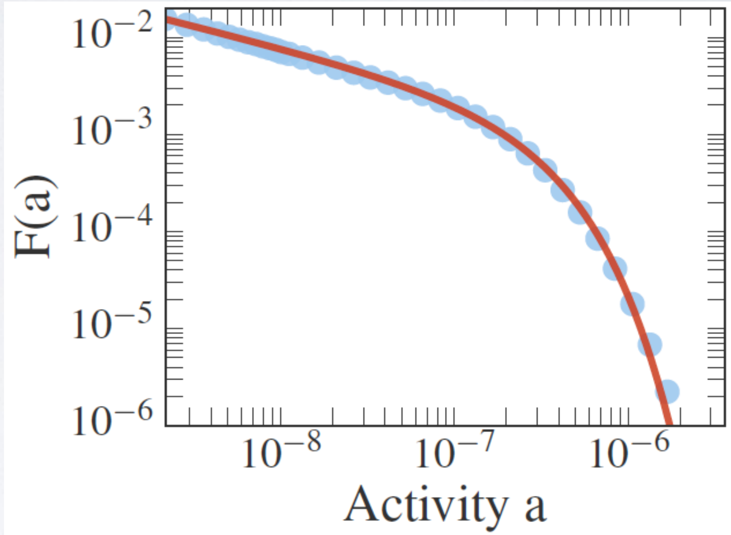
```
Citer_ID_00    Cited_ID_00    # Event 0
Citer_ID_01    Cited_ID_01    # Event 1
Citer_ID_02    Cited_ID_02    # Event 2
```

```
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 activity parameters

Activity distributions: Fits from data and measure of $\nu$




Truncated power law for MPC, APS $\quad F(a) \sim a^{-\nu} \quad$ for large a

Lognormal for TWT

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

# Networks Evolution: measuring ties selection parameters

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

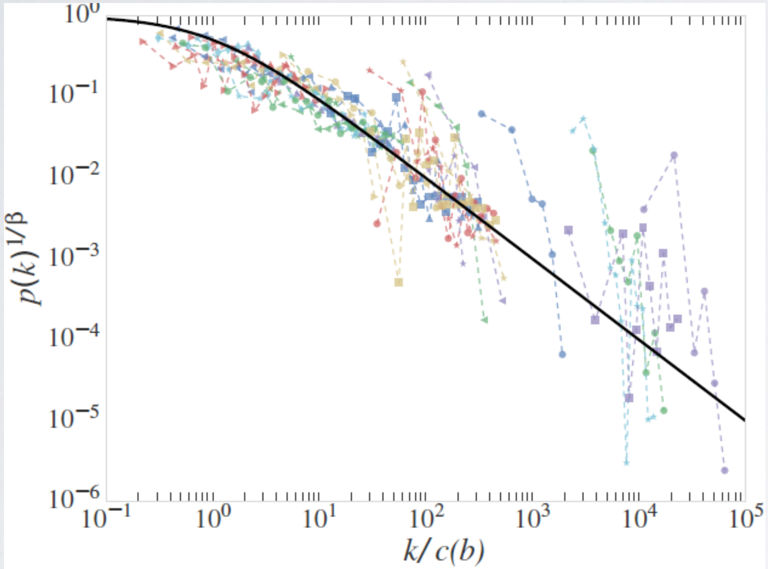
$$p_i(k) = (\frac{1}{1+\frac{k}{c_i}})^{\beta_i}$$

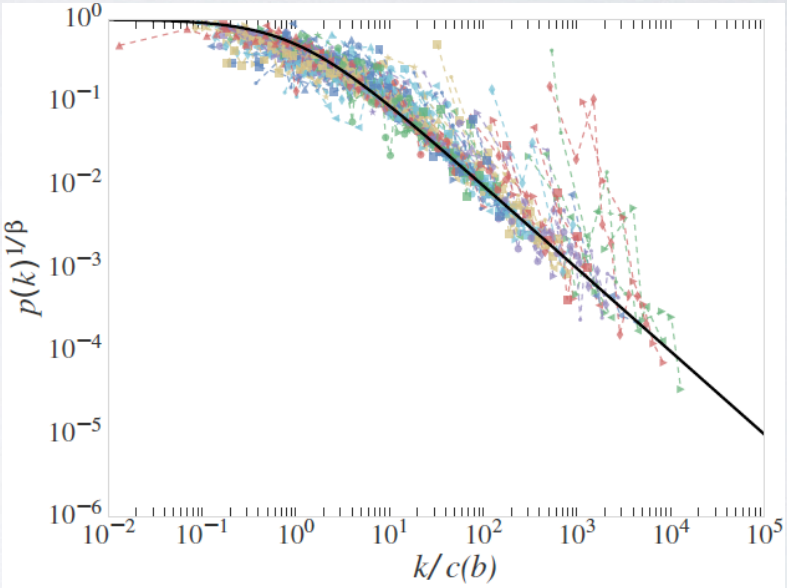
- A clever and complex averaging procedure, grouping nodes in activity classes

- Measure from large datasets

- the form of the memory is simple but works for all datasets
- the exponent beta has a measurable well peaked distribution
- also the coefficient c are distributed but very well peaked

# Networks Evolution: measuring ties selection parameters




APS (PRL) beta = 0.16
TWT beta = 0.5

The rescaled reinforcement probability for two datasets

# Networks Evolution: statistical physics approach and analytics

- Parameters: activity distribution and the memory exponent

- 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 to have degree k (already contacted nodes) at time t. $P(k,t)$

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, (2016)(2017)

# The analytic result:

A summary of analytic results

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

$$\downarrow$$

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

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

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

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

average degree growth

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

integrated degree distribution

# The analytic result: reinforcement only

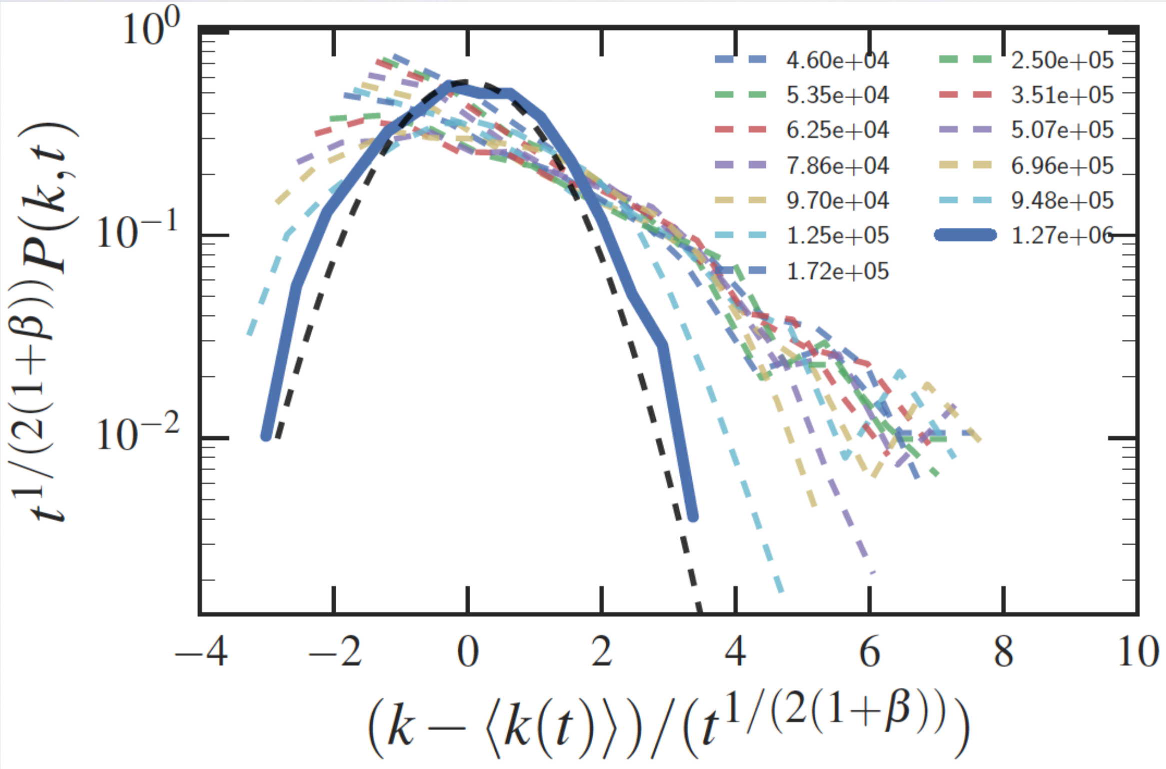
A summary of analytic results          integrated degree distribution

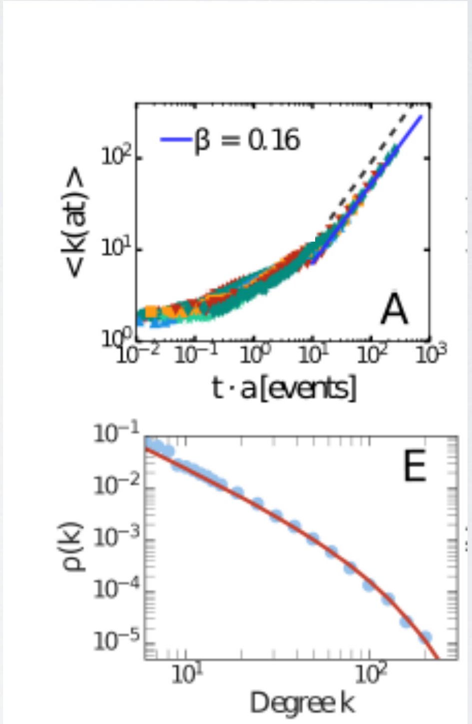
Given the form of the activity distribution and the value of the reinforcement parameter, we can forecast the form of the degree distribution for any activity distribution

| 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 | $\frac{1}{a}\exp\left[-\frac{(\ln(a)-\mu)^2}{2\sigma_a^2}\right]$ | $\frac{1}{k}\exp\left[-\frac{(\ln(k)-\gamma)^2}{2\left(\frac{\sigma_a}{1+\beta}\right)^2}\right]$ |

# and Real data: an example

APS: $\alpha \sim 2.1 \qquad \beta \sim 0.16$



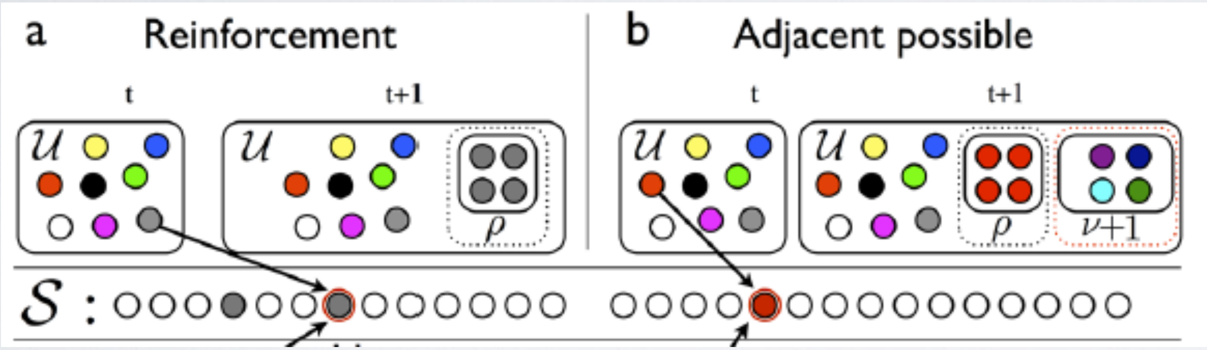
- Variables: Activity, memory
- Measure the "memory" and activity from large statistics

= get the large scale evolution of the network

Can we justify these mechanisms from the point of view of a novelty discovery?

Yes. Something analogous to the adjacent possible mechanism
seems to act a the level of old/new ties choices
and it drives the enlargement of our Social Horizon

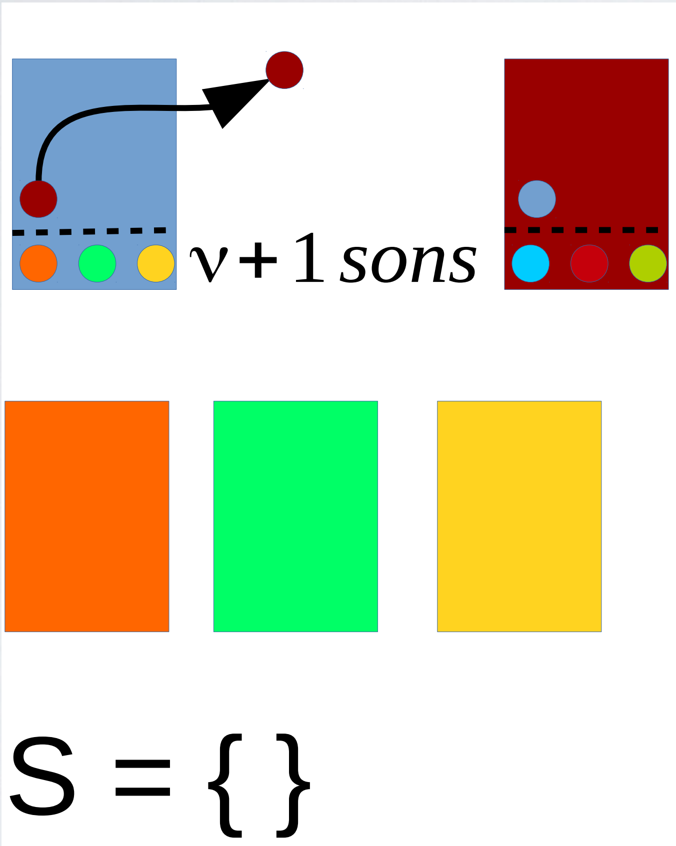
# Urn models and the adjacent possible:




Can we generalize it to model old/new links activation?

Tria, Loreto, Servedio, S.H. Strogatz (2014)

# Link activation and novelties discovery

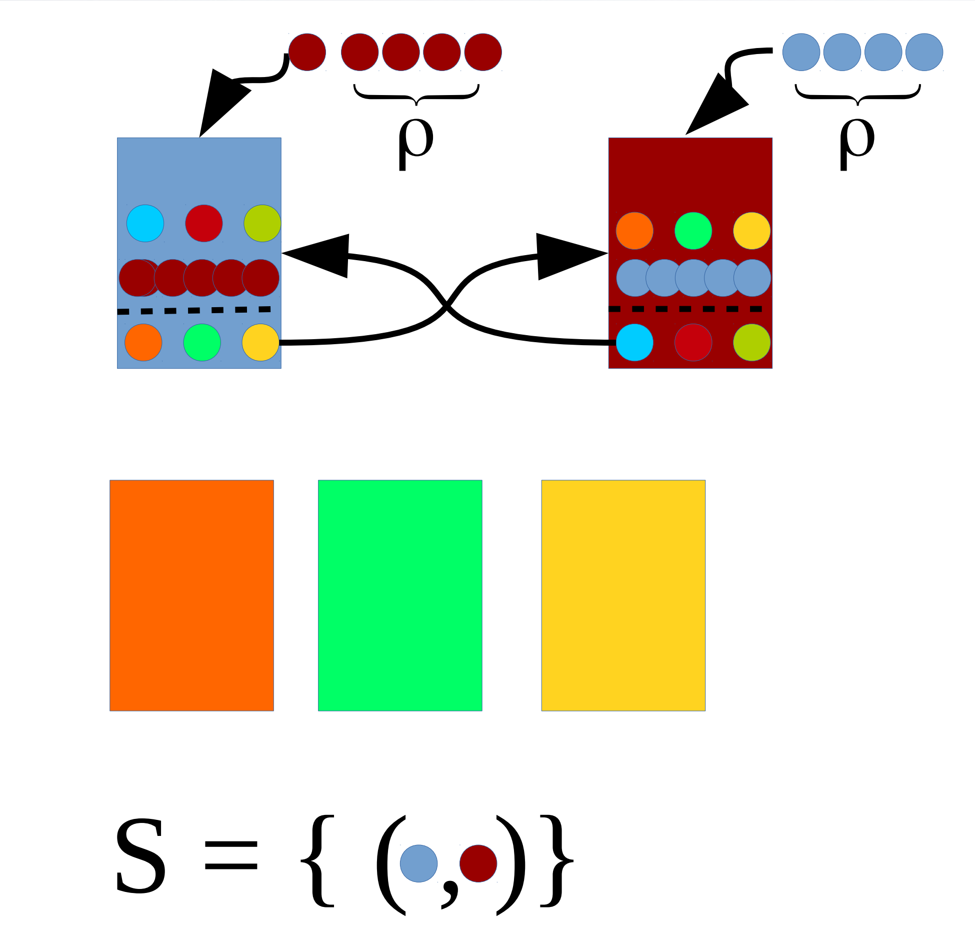



- two full urns (cyan and red) containing a copy of them and their $\nu$ + 1 sons.

-The sons of the cyan are initially empty urns.

- The extraction selects the cyan urn as the calling one. From this urn we withdraw the red.
- The first contact is then (cyan, red) .

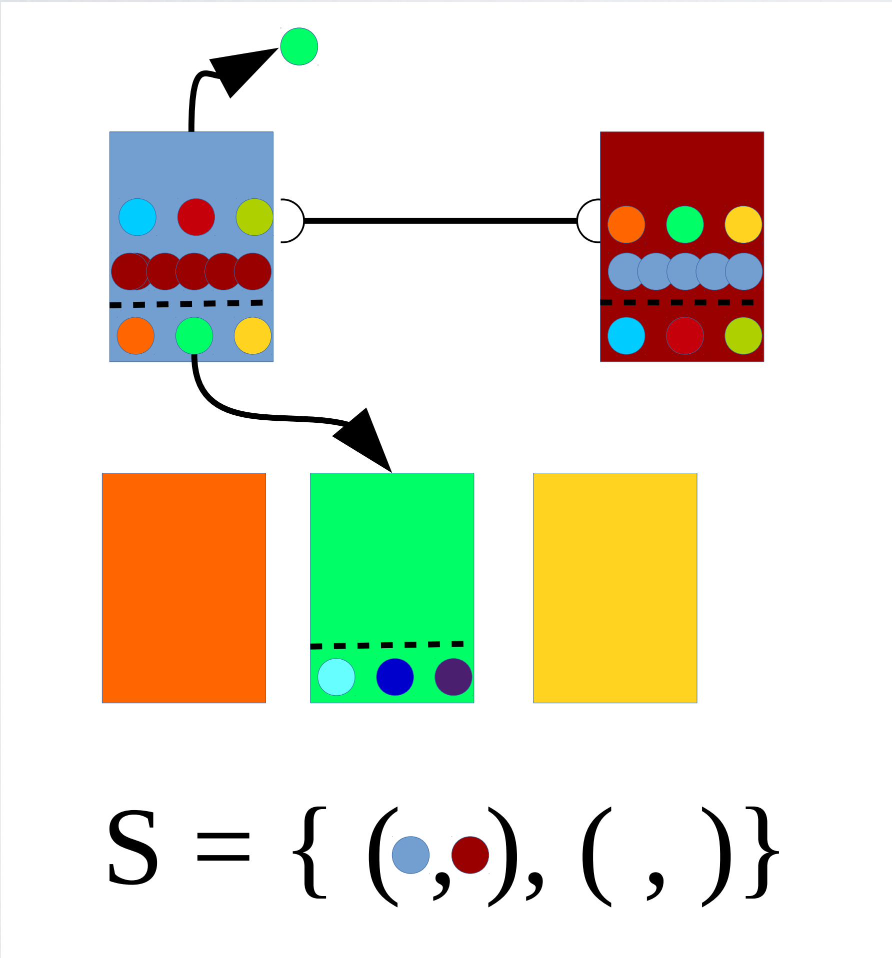
# Link activation and novelties discovery




- put back $\rho$ copies of the red ID into the cyan urn and vice-versa.
- first contact between the two urns, then we also exchange their $\nu + 1$ sons

- Future contacts between these two urns will only result in their reinforcement.

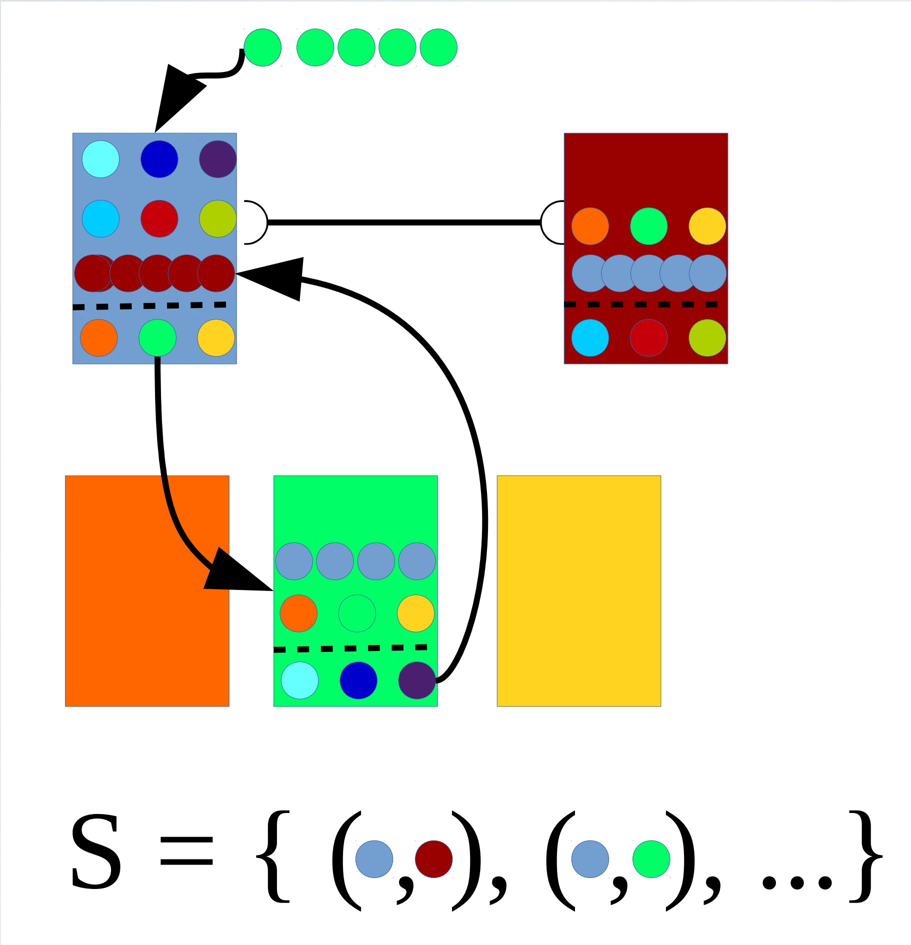
# Link activation and novelties discovery




-The cyan urn is selected again and it now to withdraws its green son, that was never called before.

- As the son's urn is empty, it creates its $\nu + 1$ sons elements together with their empty
In practice, the green enter the systems.
with all his sons.

# Link activation and novelties discovery




The (cyan, green) event is recorded in the sequence and the two urns repeat the reinforcement and sons exchange steps.

# Link activation and novelties discovery

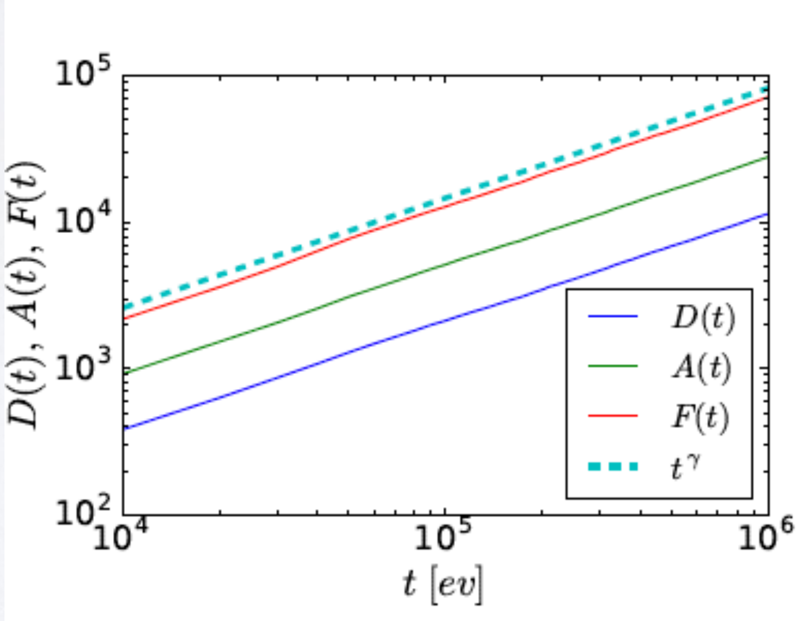
Analytics on the dynamical process:

Number of distinct agents in the sequence

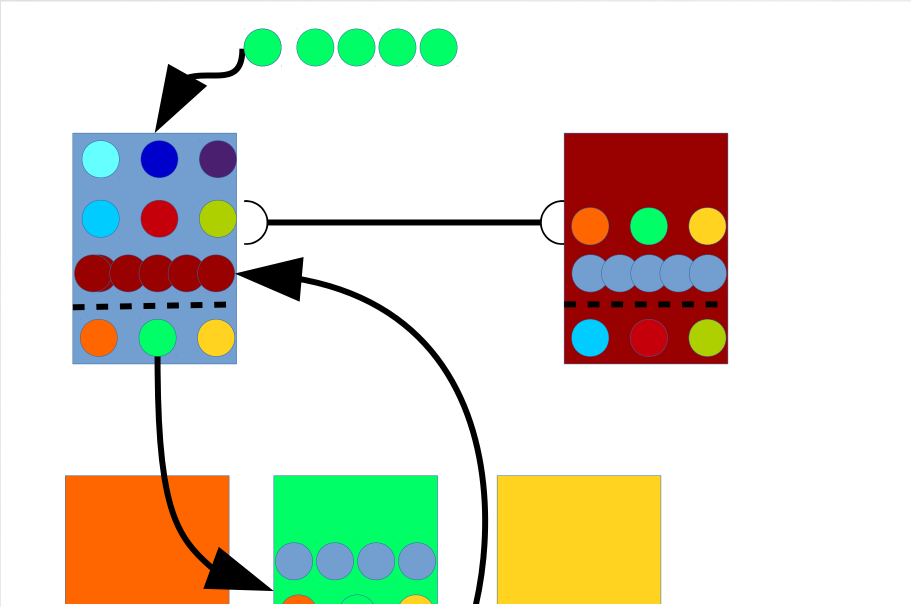
$$D(t) \sim t^{\gamma} \qquad \gamma = \frac{3}{2}\frac{\nu}{\rho}$$

hold for $\qquad \dfrac{\rho}{\nu} > 3/2$

$A(t)$ number of links



# Link activation and novelties discovery



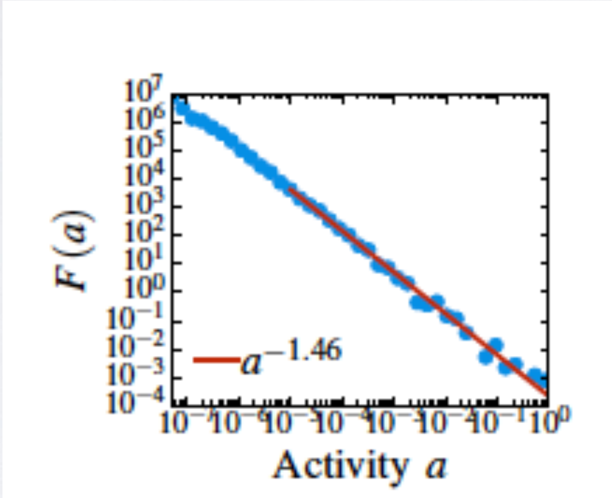
| | | | | |
|---|---|---|---|---|
| Caller_ID | Called_ID | Company_Caller | Company_Called | # Event 0 |
| Caller_ID | Called_ID | Company_Caller | Company_Called | # Event 1 |
| Caller_ID | Called_ID | Company_Caller | Company_Called | # Event 2 |
| . . . | . . . | . . . | . . . | . . . |

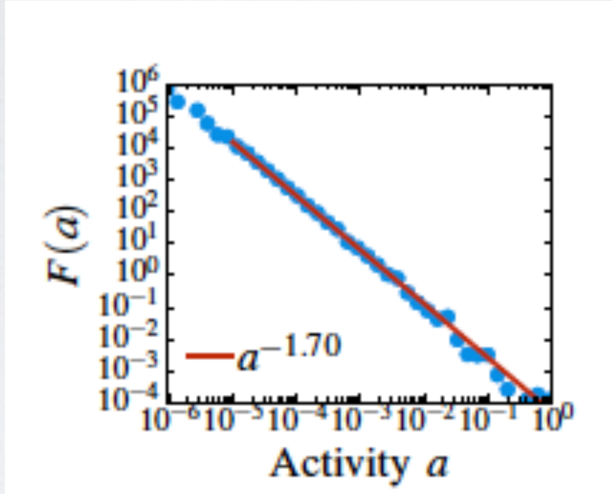
$$S = \{ (\bullet,\bullet), (\bullet,\bullet), \ldots \}$$

Now take this sequence of links and analyze them as if they were real data

# Basic mechanisms in social networks evolution & novelties

The activity distribution

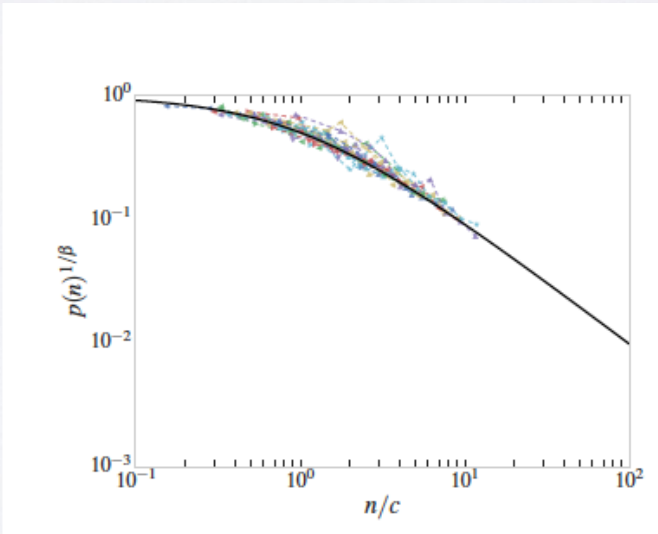


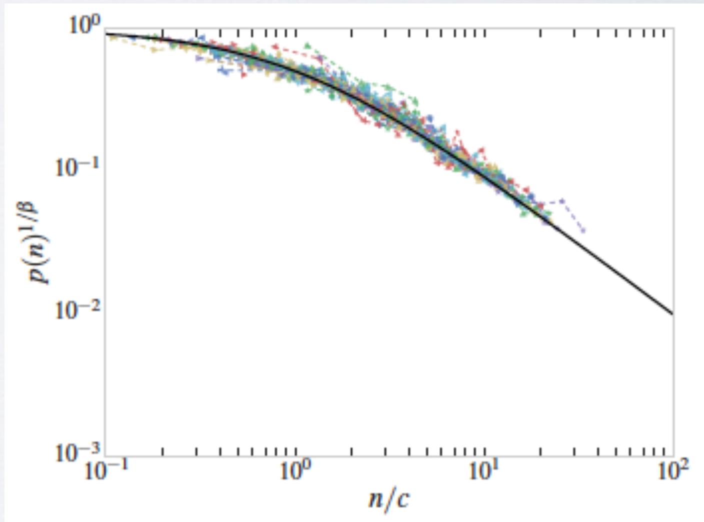




$\rho = 9, \nu = 3$

$\rho = 10, \nu = 5$

The link activation probability, rescaled

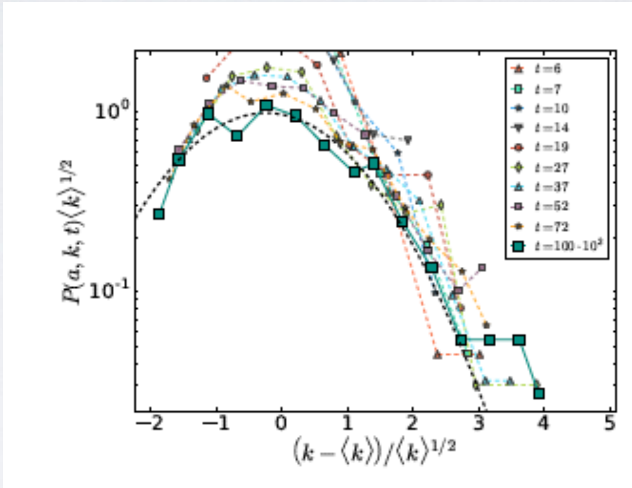


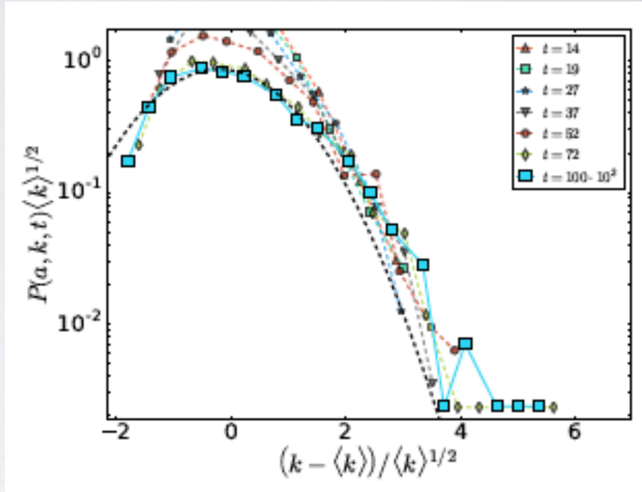




as in real datasets
Measure of beta and nu

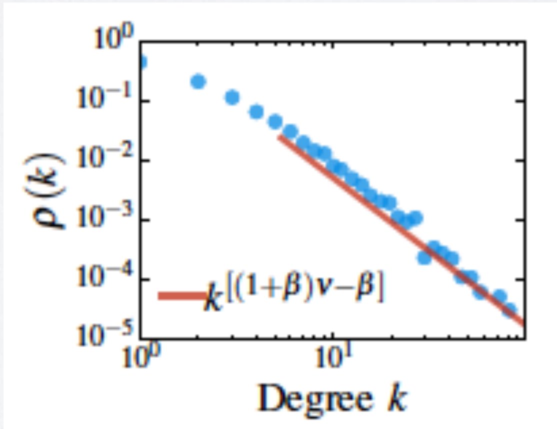
# Basic mechanisms in social networks evolution & novelties

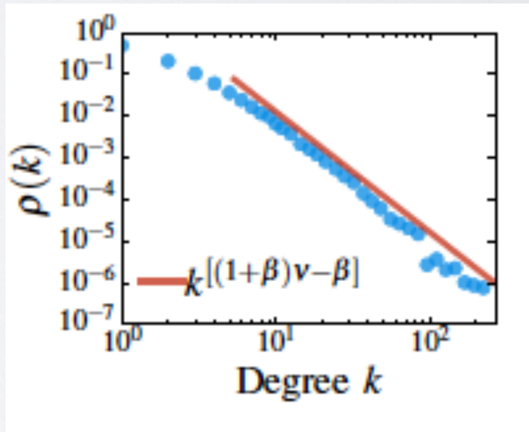
The degree probability distribution, rescaled




The integrated degree distribution

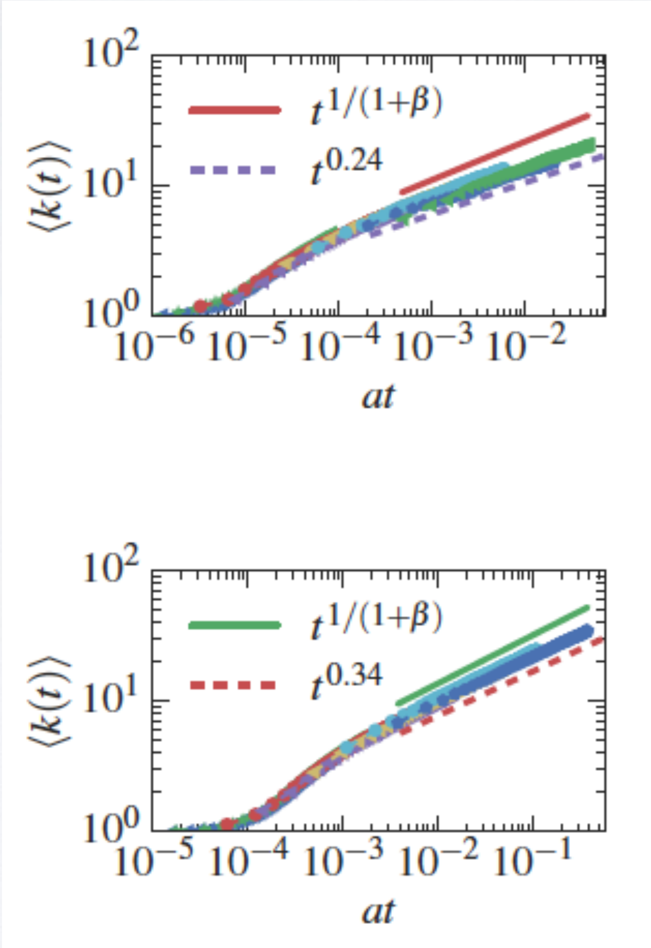



lines, analytical predictions: as in datasets

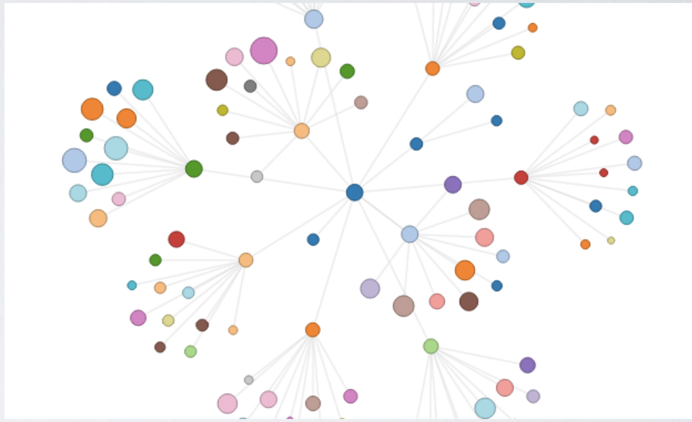
# Basic mechanisms in social networks evolution & novelties

The average degree




red, analytical prediction. Not bad but not perfect

# Social networks evolution with Heterogeneous Activity + old/new links selections



- Measured from large statistics + analytics solved

= get the large scale evolution of the network

The data driven (unexplained) forms of the Heterogeneous Activity + old/new links selections rule is well reproduced by a mechanism of Adjacent Possible

Something analogous to the adjacent possible mechanism seems to act a the level of old/new ties choices and activity distribution and seems to drives the enlargement of our Social Horizon

# Refs:

"Asymptotic theory of time varying networks with heterogenous activity and tie allocation"
E. Ubaldi,, N. Perra, M. Karsai, A. Vezzani, R. Burioni, A.Vespignani, Nat. Sci. Rep. (2016)

"Burstiness and ties activation strategies in time varying social networks"
E. Ubaldi, N. Perra, M. Karsai, A. Vezzani, R. Burioni, Nat. Sci. Rep. (2017)

Social networks evolution: when one (new) thing leads to another
E.Ubaldi, F Tria, R.Burioni, V. Loreto
In preparation