

# Modeling large-scale networks

**Fabien Tarissan**

CNRS

École Normale Supérieure Paris-Saclay

PARCOURS IA – ML&IL – COURS 2

# Plan

## ① Summary

## ② Network models

- Random graphs (Erdős-Rényi)

- Small-world model (Watts-Strogatz)

- Scale-free model (Barabási and Albert)

- Configuration Model

- Bipartite model

## ③ Dynamic networks

- Mobile networks

- Markovian model

- Empirical study

*Summary*

# Networks

Definition : Collection of entities related by means of interactions.

**Large complex networks** :

- Computer science : Internet, Peer-to-peer, Web
- Biology : Gene regulatory networks, Protein-protein interaction networks
- Social science : friendship relations, co-authors networks
- And a lot more : economy, linguistic, . . .

**Some properties are shared by a lot of networks**

⇒ leads to reconsider traditional approaches

Should lead to common solutions

# Problematic

## Measurement :

how to acquire information on a network ?

- Real graph  $\mapsto$  partial view
- Representative sample ? Bias ?

## Analysis :

how to describe a network ?

- Representation of data
- Relevant metrics
- Shared by all kind of networks ?

## Modelling :

how generate a network ?

- Random generation of similar structures (with observed properties)
- Underlying mechanisms
- Support for simulations
- Explanation of the observed properties

## ... and a lot more :

algorithmic, dynamics, ...

# Definitions

## Networks as graphs

A graph  $G = (V, E)$  is a couple of sets.

- $V$  is the set of *vertices* (or *nodes*),  $n = |V|$  is the number of nodes
- $E \subseteq (V \times V)$  is the set of *edges* (or *links*),  $m = |E|$  is the number of links

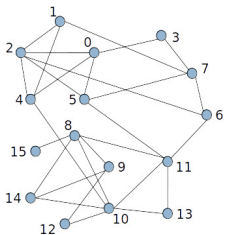
## Notions

- degree, average degree of the graph, density of the graph, ...
- path, distance, connected component, average length, diameter, ...
- directed vs. undirected graphs, weighted vs. unweighted networks, ...
- one-mode, two-mode networks, ... multi-level networks, multiplex networks
- clustering coefficient, transitive ratio, community, modularity, ...

## Data encoding

- Adjacency matrices, adjacency lists, ...

# Communities



**Goal** : Identify automatically *relevant groups*.

**Challenges** :

- Unknown number of communities
- Unknown sizes of communities
- Scalability ?

## Algorithms

- A lot of different approaches : percolation, random walk, k-core, ...
- Louvain algorithm : efficient, scalable. Based on **modularity** :

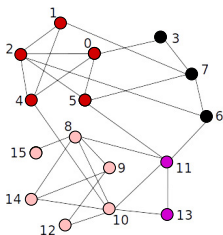
$$Q = \frac{1}{2m} \sum_C e_C - \frac{a_C^2}{2m}$$

$e_C$  : links  $\in C$ ,  $a_C$  : links with one end  $\in C$

Related to a mini-project !

<http://tarissan.complexnetworks.fr/iam1/community.pdf>

# Communities



**Goal** : Identify automatically *relevant groups*.

**Challenges** :

- Unknown number of communities
- Unknown sizes of communities
- Scalability ?

## Algorithms

- A lot of different approaches : percolation, random walk, k-core, ...
- Louvain algorithm : efficient, scalable. Based on **modularity** :

$$Q = \frac{1}{2m} \sum_C e_C - \frac{a_C^2}{2m}$$

$e_C$  : links  $\in C$ ,  $a_C$  : links with one end  $\in C$

Related to a mini-project !

<http://tarissan.complexnetworks.fr/iam1/community.pdf>



# *Network Models*

# A network science

## Common properties

- Networks from different context share similar structural properties
- Dynamic processes driving the formation of the networks **can not be related** to a particular context.
- Needs to seek explanations **regardless** of the real nature of the networks.

⇒ **New research questions**

## Highlighting common properties

## Searching for explanatory models

- **How** do the networks organize ?
- **Why** do they organize in this particular shape ?

Contrast between **globales** properties and **local** interactions !

⇒ **Emergent properties**

# A network science

## Common properties

- Networks from different context share similar structural properties
- Dynamic processes driving the formation of the networks **can not be related** to a particular context.
- Needs to seek explanations **regardless** of the real nature of the networks.

⇒ **New research questions**

## Highlighting common properties

- short distances
- low density
- high local density
- heterogeneous degree distribution
- ...

## Searching for explanatory models

- **How** do the networks organize ?
- **Why** do they organize in this particular shape ?

Contrast between **globales** properties and **local** interactions !

⇒ **Emergent properties**

2 examples :

- 1 Small-world networks (*Nature* 1998)
- 2 Scale-free networks (*Science* 1999, *Nature* 2000)

# Random graphs – Motivation

## Understand the structure

Are the observed properties **normal** ?

**Answer** : compare to a **synthetic random graph**

Draw randomly (**uniform probability**) in the set of graphs (**of a given size**)

→ **common** properties to the large majority of graphs

→ **expected** properties

## Simulate processes

**Note** : also possible with a non-random generative model

# Erdős-Rényi model

 $G_{n,p}$ 

- $n$  nodes
- Any edge exists with a given probability  $p$

# Erdős-Rényi model

$G_{n,p}$

- $n$  nodes
- Any edge exists with a given probability  $p$

Complexity :  $\mathcal{O}(n^2)$

# Erdős-Rényi model

 $G_{n,m}$ 

- $n$  nodes
- $m$  edges randomly chosen

# Erdős-Rényi model

 $G_{n,m}$ 

- $n$  nodes
- $m$  edges randomly chosen

Complexity :  $\mathcal{O}(m)$



# Equivalence between $G_{n,p}$ and $G_{n,m}$

$p$  is the **density**

$$p = \frac{2m}{n(n-1)}$$

$G_{n,m}$  and  $G_{n,p}$  are equivalent if  $p$  and  $m$  verify this relationship

# Double edges

$G_{n,m}$  : non-zero probability to draw **double** edges

## Hard to detect

- suppose we write the graph without storage
- how to proceed ?

## In practice

- few double edges
- do not change dramatically the properties observed

→ often considered as normal edges  
but avoid **loops**

# Notion of expected property

**Example** : random graph,  $n = m = 4950$

**Observation** : clique of 100 nodes (other nodes with degree 0)

Surprising ?

# Notion of expected property

**Example** : random graph,  $n = m = 4950$

**Observation** : clique of 100 nodes (other nodes with degree 0)

Surprising ?

Probability to have degree 0 :  $q = (1 - p)^{n-1} \sim 0.14$ .

⇒ Expected number of degree 0 nodes :

$$nq \sim 683$$

# Notion of expected property

**Example** : random graph,  $n = m = 4950$

**Observation** : clique of 100 nodes (other nodes with degree 0)

Surprising ?

Probability to have degree 0 :  $q = (1 - p)^{n-1} \sim 0.14$ .

⇒ Expected number of degree 0 nodes :

$$nq \sim 683$$

$$683 \neq 4850$$

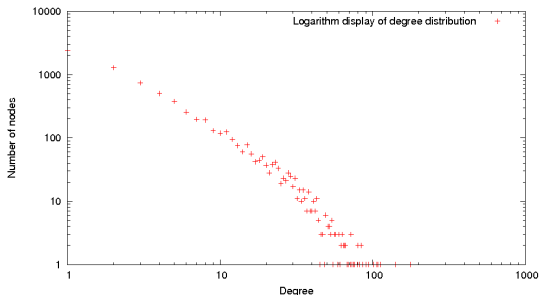
→ seem very unlikely with a random process  
(other process involved)

# Notion of expected property

## Real example

Graph properties from file "PPI-dr"

Nodes: 7362      Edges: 22453      Density: 0.0008  
Average degree: 6.0997      Highest degree: 176  
Number of component: 79      Highest size: 7202  
Transitive Ratio : 0.0196      Clustering Coefficient : 0.0184



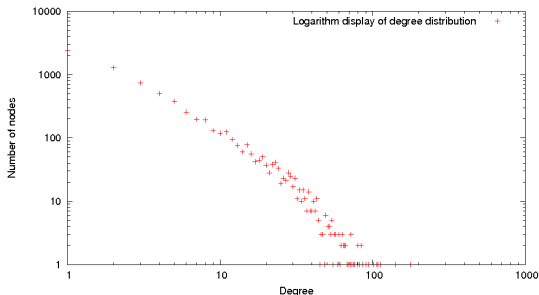
## Observation :

# Notion of expected property

## Real example

Graph properties from file "PPI-dr"

Nodes: 7362      Edges: 22453      Density: 0.0008  
 Average degree: 6.0997      Highest degree: 176  
 Number of component: 79      Highest size: 7202  
 Transitive Ratio : 0.0196      Clustering Coefficient : 0.0184



**Observation** : existence of high degree nodes ( $\geq 100$ )

Surprising ?

# Notion of expected property

## Real example

**Observation** : existence of high degree nodes ( $\geq 100$ )

Surprising?

Probability to have degree  $k$  ? (suppose  $n = 5000$  and  $m = 10000$ )



# Notion of expected property

## Real example

**Observation** : existence of high degree nodes ( $\geq 100$ )

Surprising ?

Probability to have degree  $k$  ? (suppose  $n = 5000$  and  $m = 10000$ )

$$p(k) = C_k^n * p^k * (1-p)^{n-k} < C_k^n * p^k$$

$$\text{Here, } p^k \equiv (1/10^3)^k$$

$$C_k^n = \frac{n!}{k!(n-k)!} = \frac{n*(n-1)*\dots*(n-k+1)}{k!} < \frac{n^k}{k!}$$

$$\text{But } k! \equiv \sqrt{2\pi k} \left(\frac{k}{e}\right)^k \text{ (Stirling)}$$

$$\text{Thus, } p(k) < \frac{(5000)^k}{\text{cst} - \text{gt} - 1 * 50^k} * \left(\frac{1}{10^3}\right)^k$$

$$p(k) < (10^2)^k * \left(\frac{1}{10^3}\right)^k$$

$$\text{It turns out that } p(k) < \left(\frac{1}{10}\right)^k \text{ (very unlikely!)}$$

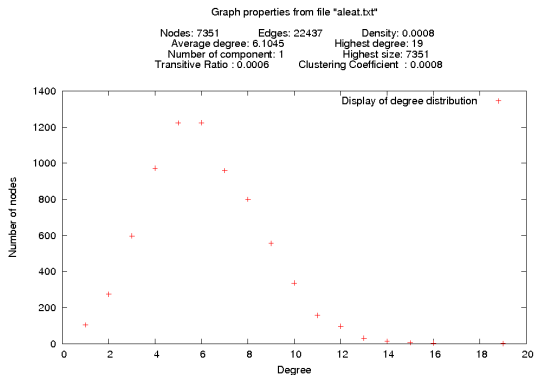
# Notion of expected property

## Real example

**Observation** : existence of high degree nodes ( $\geq 100$ )

Surprising ?

**Random** :



# Properties of Erdős-Rényi graphs

- Density
- Connectedness
  
- Average distance, diameter
  
- Degree distribution
- Clustering coefficient

# Properties of Erdős-Rényi graphs

- Density **set by operator**
- Connectedness **giant component, size  $\mathcal{O}(n)$**
- Average distance, diameter  $\sim \log(n)$
- Degree distribution **homogeneous**
- Clustering coefficient  $\simeq$  **density**

(if  $m \geq \mathcal{O}(n)$ )

(for  $m \geq \mathcal{O}(n)$ )

# Properties of Erdős-Rényi graphs

	real	random
density	low	low
connectedness	giant comp.	giant comp.
distances	low	low
degree distrib.	heterogeneous	homogeneous
clustering	high	low
communities	yes	no

# Conclusion on Erdős-Rényi graphs

Real-world complex networks are very different from random Erdős-Rényi graphs

## Consequences

- Resemblances (connectedness, distances) are actually meaningful
- Not a good model for simulations, proofs ...

→ [Other models?](#)

# Swall-World networks

A study from Duncan Watts (sociologist) and Steven Strogatz (mathematician)

Empirical study of 3 networks of different nature

- biological network : neural network (*C. Elegans* worm)
- (human) infrastructure : power-grid network of (part of) the US
- social network : collaboration networks between movie actors

	n	m	$L_{actual}$	$L_{random}$	$C_{actual}$	$C_{random}$
Film actors	225 226	6 869 393	3.65	2.99	0.79	0.00027
Power grid	4 941	6 596	18.7	12.4	0.080	0.005
<i>C. elegans</i>	282	1974	2.65	2.25	0.28	0.05

# Swall-World networks

A study from Duncan Watts (sociologist) and Steven Strogatz (mathematician)

Empirical study of 3 networks of different nature

- biological network : neural network (*C. Elegans* worm)
- (human) infrastructure : power-grid network of (part of) the US
- social network : collaboration networks between movie actors

	n	m	$L_{actual}$	$L_{random}$	$C_{actual}$	$C_{random}$
Film actors	225 226	6 869 393	3.65	2.99	0.79	0.00027
Power grid	4 941	6 596	18.7	12.4	0.080	0.005
<i>C. elegans</i>	282	1974	2.65	2.25	0.28	0.05

The networks all have short distances and a high local density  
 ⇒ "Small-world" networks



# Which driving mechanisms ?

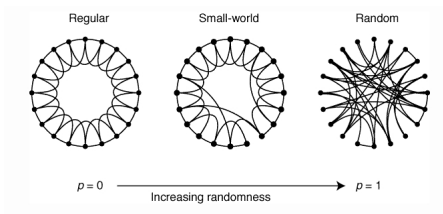
Small-world = short distances and high clustering ... **incompatible** properties !

Standard models :

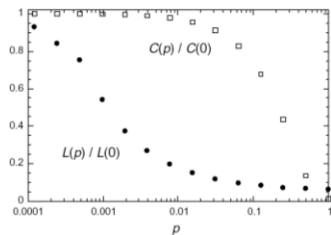
- Erdős-Rényi (random) : **short distances** but **low clustering**
- $k$ -regular : **high clustering** but **high distances**

Watts-Strogatz model (*Nature*, 1998)

From a  $k$ -regular network, random reconnections of edges with probability  $p$  (parameter of the model)

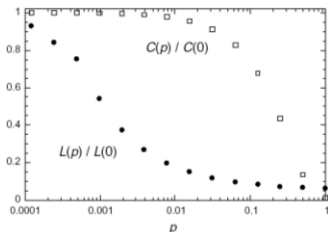


# Watts-Strogatz model



## Results

# Watts-Strogatz model



## Results

With a very low value of  $p \in [0.001 : 0.01]$  (ie. small number of random rewirings) one can generate graphs with both properties (small-world graphs).

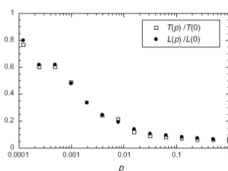
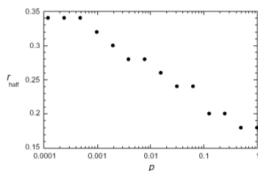
## Interpretation

- the links organize primarily locally ( $\mapsto$  hence a **high local density**)
- random links have the ability to create bridges between distant regions ( $\mapsto$  hence **low average distances**)

# What are the benefits ?

## Diffusion models (SIR model)

The diffusion spreads nodes by nodes according to the infection rate (parameter).  
 → Study of the impact of structural properties on the diffusion

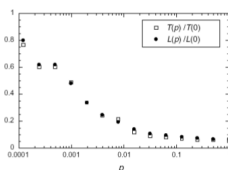
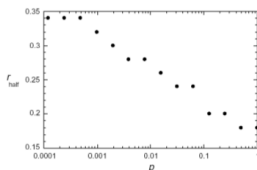


## Results

# What are the benefits ?

## Diffusion models (SIR model)

The diffusion spreads nodes by nodes according to the infection rate (parameter).  
 → Study of the impact of structural properties on the diffusion



## Results

- ① The more random links there are, the weakest the infection rate needs to be
- ② For a given infection rate, the diffusion is more efficient when there are random links

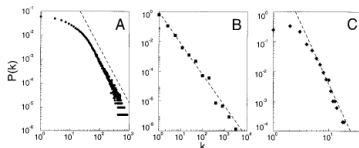
Duncan J. Watts and Steven H. Strogatz, "Collective dynamics of 'small-world' networks", *Nature*, vol. 393, n° 6634, 1998, p. 440-442.

# Scale-free networks

A study from Albert-Lázló Barabási and Réka Albert (physicists)

All nodes have the same degree ... **is this realistic ?**

- collaborations between actors
- the Web
- the american power-grid network



## Results

- 1 **heterogeneous** degree distribution (close to a power-law)
- 2 Most of the nodes have a very small degree
- 3 Existence of **hubs**

⇒ **Scale-free** networks

# Preferential attachment principle

A model to explain this scale-free nature ?

Main flaws of existing models : they are **static** !

- random graphs
- k-regular graphs
- Watts-Strogatz model

But most of networks grow in time (web, scientific collaborations, ...)

⇒ **How do new nodes link to existing ones ?**

## Barabási-Albert model

Simple (but realistic) hypothesis : graph built according to a **preferential attachment** principle :

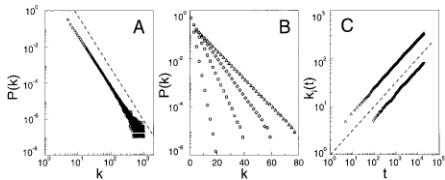
$$P(k_i) = \frac{k_i}{\sum_j k_j}$$

# Barabási-Albert model

## Virtuous effect of the model

The more a node has a high degree, the more it attracts new nodes  
 ... hence an even **higher degree** it gets!

**Justification** : *"rich gets richer"* rule (or Merton's *"Matthew's effect"*)



## Result

This model generates graphs with a **power-law degree distribution** (not proved here)

**Albert-László Barabási and Réka Albert**, "Emergence of scaling in random networks", *Science*, vol. 286, n° 5439, 1999, p. 509-512.



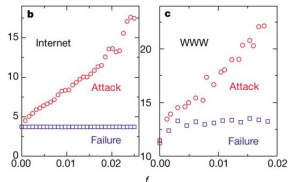
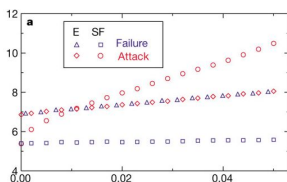
# What role do the *hubs* play ?

## Different perturbations

- failure : one removes nodes randomly
- attack : one removes nodes with high degree first

## Network model

- random networks
- scale-free networks



## Results

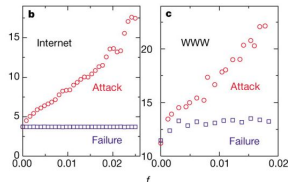
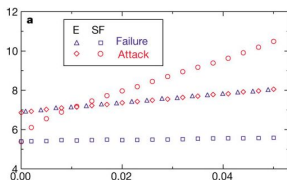
# What role do the *hubs* play ?

## Different perturbations

- failure : one removes nodes randomly
- attack : one removes nodes with high degree first

## Network model

- random networks
- scale-free networks



## Results

The presence of *hubs* make networks :

- more **robust** regarding random failures
- but **vulnerable** to targeted attacks

Réka Albert, Hawoong Jeong and Albert-László Barabási, "Error and attack tolerance of complex networks", *Nature*, vol. 406, n° 6794, 2000, p. 378-382.

# Configuration model

Degree distribution

$p_1, p_2, p_3, \dots$

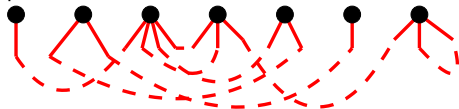
Draw nodes degree according to the distribution

1      2      4      3      2      1      3

Associate to any node half-links (or stubs)



Draw randomly pairs of stubs



# Configuration model (implem)

Table : node  $i$  occurs exactly  $\delta(i)$  times

0	1	1	2	2	2	2	3	3	3	4	4	5	6	6	6
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

---

**Algorithm 1:** Generating a graph with fixed degree distribution

---

```

begin
  Choose a random pair of stubs
   $i = 2m$ 
  while  $i > 0$  do
     $u = \text{random}(0, i - 1)$ 
    swap boxes  $u$  and  $i - 1$ 
     $v = \text{random}(0, i - 2)$ 
    swap boxes  $v$  and  $i - 2$ 
     $i = i - 2$ 
    // edge  $(u, v)$  created
  end
end

```

---

# Switching method

## Principle

- we **must** have a graph having the degree distribution
- iterate **switching of edges ends**
- after a *sufficient amount* of switches, the graph produced is a **random element of the set of graphs**



# Switching method

## Why does it work ?

- The degree of any node remains unchanged
- The process is a Markov chain

can be seen as a random walk in the set of graphs defined by this degree distribution

after a while, we visit all elements with the same probability (not proved here)

## How can we know that enough switches have been made ?

Measuring some features (ex : clustering) during the process  
when these features do not evolve any more ?

# Properties – Comparison

	real	Erdős-Rényi	fixed d.d.
density	low	low	low
connectedness	giant comp.	giant comp.	giant comp.
distances	low	low	low
degree	heterogeneous	homogeneous	heterogeneous
clustering	high	low	low
communities	yes	no	no

→ clustering **is not a consequence** of heterogeneous degree

# Bipartite Graph

Newman, Watts and Strogatz - *PNAS*, 2002

Example of the *Internet Movie Data Base* : what means a link between two actors ?  
**Richer representation** : network actor/movie

## Vocabulary

- **bipartite graph** :  
2 subsets of nodes A and B,  
links only connect nodes in A to nodes in B
- the actor network is a **projection** of this network with **less information**



# Bipartite case

The direct generative method (as well as the switching method) can be applied :

- using two degree distributions (for nodes A and B)
- connecting only nodes of A to nodes of B

## Results

- explains clustering and degree in projections for some graphs  
in Newman *et al.* : *coboarding* ok, not in collaboration networks
- no large-scale structure (communities)

# Conclusion – Properties

## A language to describe networks : graphs

- nodes, links
- degree, density
- path, length, diameter, connected component
- local density, clustering coefficient, transitive ratio
- community

## Properties of the networks

Most networks share **common properties**.

One needs **models** to explain the **emergence** of those properties.

	Network	random	k-regular	CM	WS	AB
density	low	low	low	low	low	low
connect.	giant comp.	giant comp.	giant comp.	giant comp.	giant comp.	giant comp.
distances	short	short	long	short	short	short
degree	heterogeneous	homogeneous	homogeneous	heterogeneous	homogeneous	heterogeneous
clustering	fort	low	high	low	high	low

# Conclusion – Network Science

## A new direction of research

- 1 Search for common properties of the networks
- 2 Identification of mechanisms leading to the emergence of those properties
- 3 Identification of the benefits of those properties for networks

## Small-world networks

- Small-world property : short distances and high local density
- Watts-Strogatz model : few random links in a  $k$ -regular graph
- Benefits : fast diffusion of information

## Réseaux sans échelle

- Scale-free property : heterogeneous degree distribution
- Barabási-Albert model : preferential attachment principle
- Benefits : robust regarding random failures

# *Mobile Networks*

# Dynamical aspect of networks

## Motivation

- Development of wireless devices
- A lot of new open dataset
- Dynamics ON and **OF** the network
- New structural properties
- Redefining usual metrics (graphs)

## Issues

- How acquire knowledge from this object ? (**measure**)
- Which notable properties ? (**analyze**)
- Which models best capture those properties ? (**modelling**)

# Models for evolving graphs

## Background :

- **Evolving graph** model : recent [FER02]
- Evolving graph = Succession of distinct graphs  $G_0, G_1, \dots$  with  $V$  given
- Capture all types of dynamics

# Models for evolving graphs

## Background :

- **Evolving graph** model : recent [FER02]
- Evolving graph = Succession of distinct graphs  $G_0, G_1, \dots$  with  $V$  given
- Capture all types of dynamics

## Variant of edge-markovian evolving graph :

- Temporal dependency in the evolution of the graph
- $G_{t+1}$  determined by  $G_t$  and 2 parameters :
  - $p$  : probability of creation of a *non-existing* link
  - $d$  : probability of deletion of an *existing* link

# Example

Example with 4 nodes,  $p = 0.3$ ,  $d = 0.2$  and 5 time steps.

1	3	2	3
1	4	1	3
2	3	1	2
2	3	4	4
2	4	1	2
2	4	4	4
3	4	1	3

- 1st and 2nd column : identifiers of nodes involved in the contact
- 3rd column : starting time of contact
- 4th column : ending time of contact



# Advantages / drawbacks

Interest is twofold :

- $\forall G_0, p, d$  : converge towards an Erdős Rényi graph with  $\hat{p} = \frac{p}{p+d}$
- Few parameters  $\implies$  theoretical results

But it is also its weakness :

# Advantages / drawbacks

Interest is twofold :

- $\forall G_0, p, d$  : converge towards an Erdős Rényi graph with  $\hat{p} = \frac{p}{p+d}$
- Few parameters  $\implies$  theoretical results

But it is also its weakness :

- 2 parameters to rule **all** creations/deletions
- Suppose that those 2 values are representative for the l'**entire** evolution of the de network

# Methodology

## Goal :

Conduct a study to see if it is true.

- Analyze properties of the dynamics as observed in several dataset
- Comparison with the markovian model

## *Elements of response*

- Yes for [WH11] (and [VOJ11]) but ...
- ... study over 1 dataset
- ... the criteria is weak : time needed to flood the network

# Rollernet

- Rollerblade tour in Paris
- Date : August 2006.
- Duration : 3h with a break (30 min) couvering approx. 30km,
- Location : street of Paris
- Technology : *iMotes* (bluetooth)
- Size : 62 participants
- Frequency : every 15s.

# Infocom06

- Experiment made during Infocom conference at Barcelona.
- Date : April 2006
- Duration : 3 days
- Technology : *iMote*
- Size : 98 iMotes (78 participants, 17 static, and 3 in elevators)
- Frequency : every 120s.

# Sociopattern

- Exhibition in at a gallery (diseases propagations).
- Date : 2009
- Duration : 3 months
- Technology : *radio badges*
- Size : 88 to 410 (depends on the day)
- Frequency : every 20s.

## 6 case studies

Dataset	RollerNet	Infocom05	Infocom06	HT09	Socio	PMTR
Duration	3 hours	4 days	4 days	2,5 days	1 day	10 days
Participants	62	41	98	113	151	44
Contacts	60 146	17 682	148 784	9 865	2 051	11 895
Frequency (sec.)	15	120	120	20	20	1

*For each :*

- "Physical" contact network among individuals
- Each individual is equipped with a sensing device
- Detection between devices if proximity between individuals (2 to 10 m.)
- Frequency of detection varies, as well as duration of the experiments

*In the rest of the presentation, 3 dataset only :*

- RollerNet
- Infocom06 : similar to Infocom05
- SocioPattern : similar to HT09 and PMTR

# Methodology



# Methodology

*For each dataset and for each time step*

- Fraction of created links (over possible new links)
- Fraction of deleted links (over existing links)

Corresponds to the parameters  $p$  and  $d$  of the model

*Analyze :*

- Evolution over time
- Distribution of the values
- Generation of artificial graphs according to the markovian model
- Comparison between real/artificial graph

# Created links

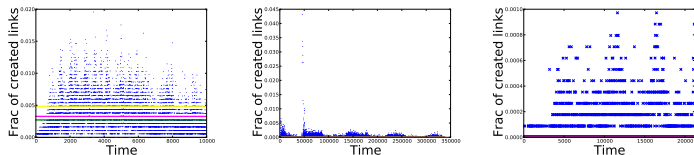


FIGURE – Evolution of the proportion de created links over time

## Results

# Created links

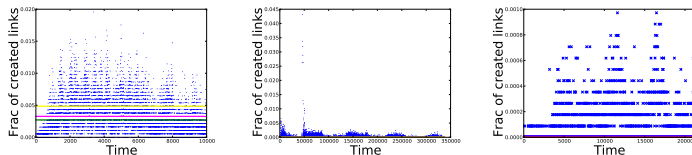


FIGURE – Evolution of the proportion de created links over time

## Results

- RollerNet : notion of average is relevant
- Infocom06, SocioPattern : wide range of values
- Infocom06, SocioPattern : average, median and 75th percentile overcome by weak values
- $\Rightarrow$  Infocom06, SocioPattern : non realistic

# Deleted links

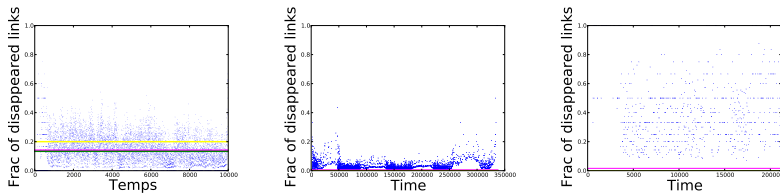


FIGURE – Evolution of the proportion of deleted links over time

## Results

# Deleted links

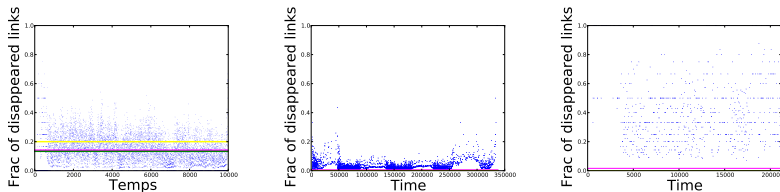


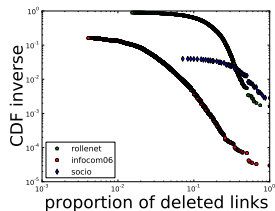
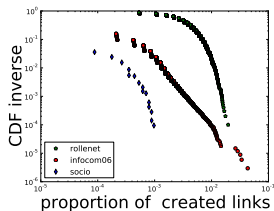
FIGURE – Evolution of the proportion of deleted links over time

## Results

- Same observation but amplified
- Range of values is covered ( $[0 : 1]$ )
- Particular case for  $d = 1$

# Distribution of $p$ and $d$ values

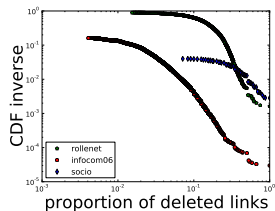
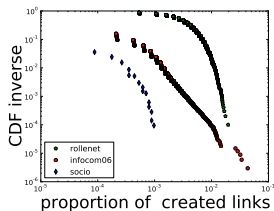
Dataset	RollerNet	Infocom06	Socio
Fractions of created links (average)	$3.2 (10^{-3})$	$9.5 (10^{-5})$	$9 (10^{-6})$
Fractions of deleted links (average)	$1.4 (10^{-1})$	$4.5 (10^{-3})$	$1.6 (10^{-2})$



## Results

# Distribution of $p$ and $d$ values

Dataset	RollerNet	Infocom06	Socio
Fractions of created links (average)	$3.2 (10^{-3})$	$9.5 (10^{-5})$	$9 (10^{-6})$
Fractions of deleted links (average)	$1.4 (10^{-1})$	$4.5 (10^{-3})$	$1.6 (10^{-2})$



## Results

- Clearly heterogeneous for Infocom06
- and on several order of magnitudes
- RollerNet : sudden slope around the average value

# Methodology

*So far :*

- Studied the dynamics related to creation and delation of links
- Provided evidences that the models *is probably not suited* to particular dataset

*How to demonstrate that the model is not pertinent ?*



# Methodology

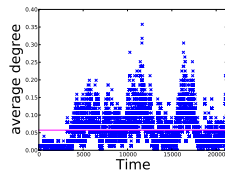
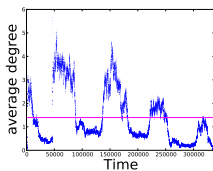
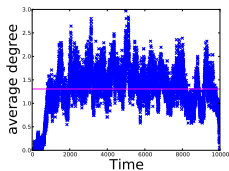
## *So far :*

- Studied the dynamics related to creation and delation of links
- Provided evidences that the models *is probably not suited* to particular dataset

## *How to demonstrate that the model is not pertinent ?*

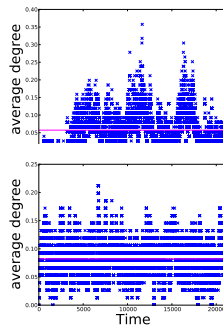
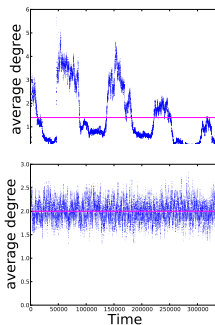
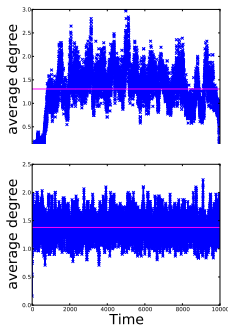
- Choose an external criteria (ie not the fraction of created and deleted links) ...
- ... but close enough the meaning of  $p$  and  $d$  (for fairness)
- Compute the value of the criteria for the real and the artificial graphs.
- Comparison between real/artificial graph.

# Evolution of mean degree



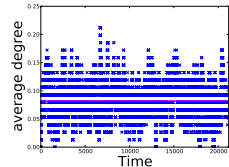
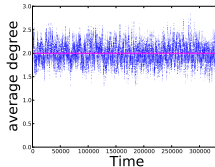
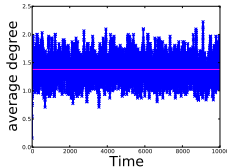
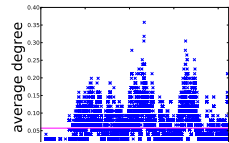
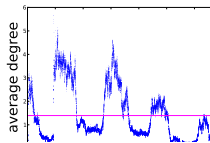
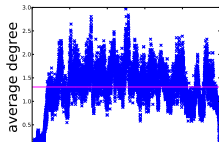
## Results

# Evolution of mean degree



## Results

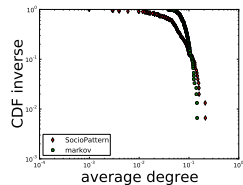
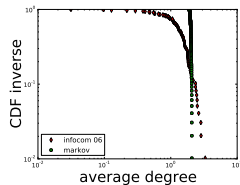
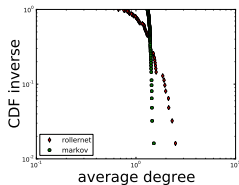
# Evolution of mean degree



## Results

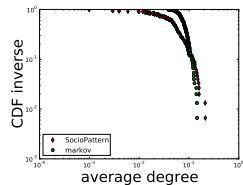
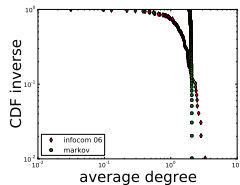
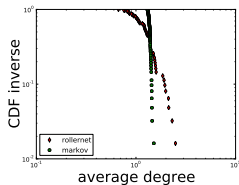
- "Uniformization" for Infocom06 and SocioPattern (not the same range of values !)
- Seems to have little impact on RollerNet
- Except at the beginning (expected)

# Average degree distribution



## Results

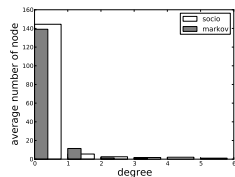
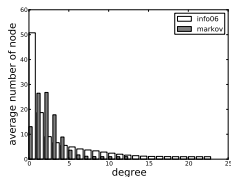
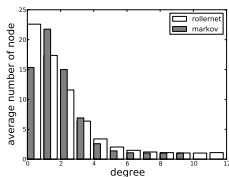
# Average degree distribution



## Results

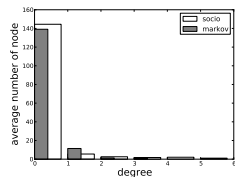
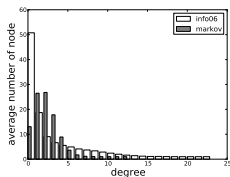
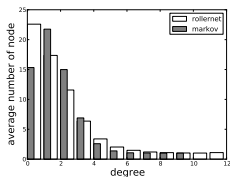
- Infocom06 : clear differences between model and real data (expected)
- RollerNet and SocioPattern : also different, although less obvious

# Degree distribution



## Results

# Degree distribution

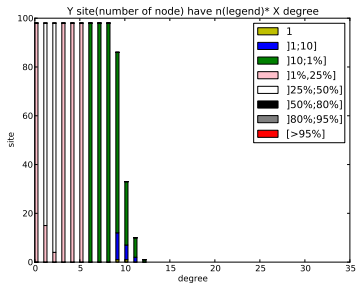
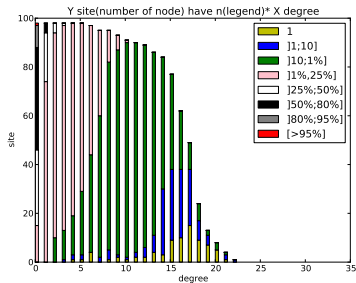


## Results

average value relevant  $\nRightarrow$  the model reproduces well the global properties of the networks

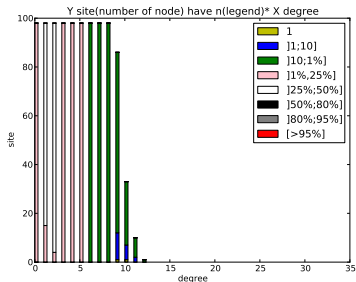
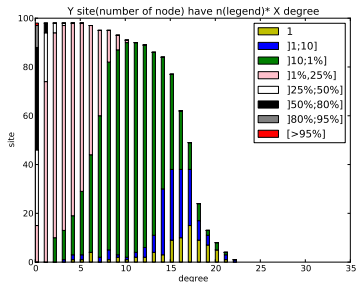


# Distribution and frequency of the degrees (Infocom06)



## Results

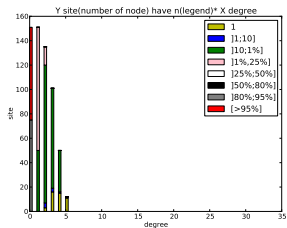
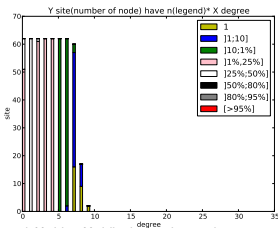
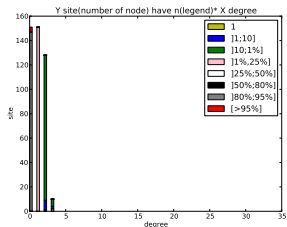
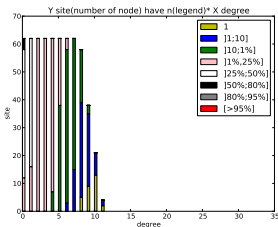
# Distribution and frequency of the degrees (Infocom06)



## Results

- Nodes are more degree-stable in real networks
- Small degrees are over-represented
- *No node with the same degree more than 50 % of the time in the model*

# Distribution and frequency of the degrees (RollerNet, SocioPattern)



# Conclusions and perspectives

## Conclusions

- Confrontation markovian model vs. real data
- Hypothesis of homogeneity does not stand in most of the cases
- Even in favourable case, it **does not reproduce the dynamics**
- *Still remain useful* : cf [WHI11, VOJ11]

## Perspectives

- Consider other way to define  $p$  and  $d$  (following an heterogeneous distribution? different for each nodes? depending on the graph state? ...)
- Study refined properties (distribution of connexions)
- Analyze correlation between creations and deletions
- Take into account the local density

Related to a mini-project !

<http://tarissan.complexnetworks.fr/iaml/mobile.pdf>

# What next ?

<http://tarissan.complexnetworks.fr/iaml.html>

- 1 Practical session on (static) network models  
[http://tarissan.complexnetworks.fr/iaml/tp\\_mlia.pdf](http://tarissan.complexnetworks.fr/iaml/tp_mlia.pdf)
- 2 Discuss the community detection mini-project  
<http://tarissan.complexnetworks.fr/iaml/community.pdf>
- 3 Discuss the mobile mini-project  
<http://tarissan.complexnetworks.fr/iaml/mobile.pdf>