Networks Structure and Dynamics
4. Random graph models

Lionel Tabourier, Fabien Tarissan

LIP6 – CNRS and Université Pierre et Marie Curie

first_name.last_name@lip6.fr

October 11^{th} 2016

Motivations and context
Graphs with given density
Non-uniformly random networks
Graphs with fixed degree distribution

Outline

Motivations and context
Graphs with given density
Non-uniformly random networks
Graphs with fixed degree distribution

Random graphs – Motivation

1: 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

2: simulate processes

Note: also possible with a non-random generative model

Motivations and context
Graphs with given density
Non-uniformly random networks
Graphs with fixed degree distribution

Random graphs – Motivation

1: 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

2: simulate processes

Note: also possible with a non-random generative model

Random graphs – Motivation

1: 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

2: simulate processes

Note: also possible with a non-random generative model

Outline

Erdös-Rényi model

$$G_{n,p}$$

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

Motivations and context
Graphs with given density
Non-uniformly random networks
Graphs with fixed degree distribution

Erdös-Rényi model

$$G_{n,p}$$

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

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

Erdös-Rényi model

$$G_{n,m}$$

- n nodes
- m edges randomly chosen

Motivations and context
Graphs with given density
Non-uniformly random networks
Graphs with fixed degree distribution

Erdös-Rényi model

$$G_{n,m}$$

- n nodes
- m edges randomly chosen

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

Motivations and context
Graphs with given density
Non-uniformly random networks
Graphs with fixed degree distribution

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

p is the density

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

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

Motivations and context
Graphs with given density
Non-uniformly random networks
Graphs with fixed degree distribution

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

$\rightarrow$ often considered as normal edges
but avoid loops

Notion of expected property

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

Result : clique of 100 nodes (other nodes with degree 0)

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

$\Rightarrow$ Expected number of degree 0 nodes:
$nq \sim 683$ to be compared with 4850...

$\rightarrow$ seem very unlikely with a random process
(other process at stake)

Notion of expected property

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

Result : clique of 100 nodes (other nodes with degree 0)

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

$\Rightarrow$ Expected number of degree 0 nodes:
$nq \sim 683$ to be compared with 4850...

$\rightarrow$ seem very unlikely with a random process
(other process at stake)

Notion of expected property

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

Result : clique of 100 nodes (other nodes with degree 0)

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

$\Rightarrow$ Expected number of degree 0 nodes:
$nq \sim 683$ to be compared with 4850...

$\rightarrow$ seem very unlikely with a random process
(other process at stake)

Properties of Erdös-Rényi graphs

- Density
- Connectedness
- Average distance, diameter

Properties of Erdös-Rényi graphs

- Density set by operator
- Connectedness giant component, size $\mathcal{O}(n)$
  (if $m \geq \mathcal{O}(n)$)
- Average distance, diameter $\sim \log(n)$
  (for $m \geq \mathcal{O}(n)$)

Properties of Erdös-Rényi graphs

- Degree distribution
- Clustering coefficient

Properties of Erdös-Rényi graphs

- Degree distribution homogeneous
- Clustering coefficient $\simeq$ density

Properties of Erdös-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 Erdös-Rényi graphs

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

Consequences

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

$\rightarrow$ Other models?

Outline

Swall-World networks

Watts et Strogatz - *Nature, 1998*

Small-world: small average distance, high clustering

From a regular network, random reconnections of edges with probability p:

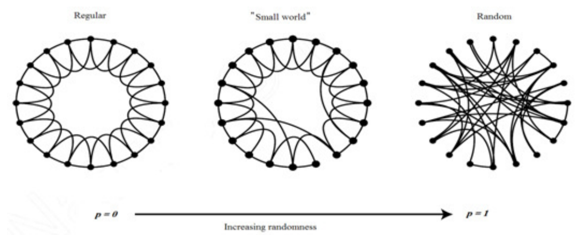
Produce high clustering but distribution non heterogeneous...

Swall-World networks

Watts et Strogatz - *Nature, 1998*

Small-world: small average distance, high clustering

From a regular network, random reconnections of edges with probability p:



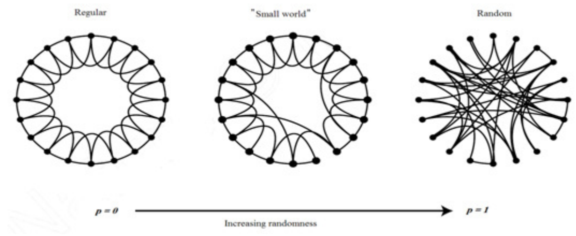
Produce high clustering but distribution non heterogeneous...

Swall-World networks

Watts et Strogatz - *Nature, 1998*

Small-world: small average distance, high clustering

From a regular network, random reconnections of edges with probability p:



Produce high clustering but distribution non heterogeneous...

Scale-Free model

fr: modèle sans échelle

Barabási et Albert - *Science, 1999*

Graph built according to the **preferential attachment** law:

probability for a node *i* to be connected to a new comer proportional to degree $\delta(i)$

Justification: this process leads to a power-law degree distribution (not proved here)

Other justification (?): generative process in agreement with the *"rich gets richer"* rule (or Merton's *"Matthew's effect"*)
but is it really relevant ? ...

Scale-Free model

fr: modèle sans échelle

Barabási et Albert - *Science, 1999*

Graph built according to the **preferential attachment** law:

probability for a node *i* to be connected to a new comer proportional to degree $\delta(i)$

Justification: this process leads to a power-law degree distribution (not proved here)

Other justification (?): generative process in agreement with the *"rich gets richer"* rule (or Merton's *"Matthew's effect"*)
but is it really relevant ? ...

Scale-Free model

fr: modèle sans échelle

Barabási et Albert - *Science, 1999*

Graph built according to the **preferential attachment** law:

probability for a node *i* to be connected to a new comer proportional to degree $\delta(i)$

Justification: this process leads to a power-law degree distribution (not proved here)

Other justification (?): generative process in agreement with the *"rich gets richer"* rule (or Merton's *"Matthew's effect"*)
but is it really relevant ? ...

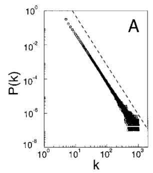
Scale-Free model

Algorithm 1: BA graph generation, with n nodes and $n.\alpha$ links

Parameters: n , α
Initial conditions: graph G with n_0 nodes , $i = n_0$
while $i < n$ **do**
  add node i to G
  $num\_links\_i = 0$
  while $num\_links\_i < \alpha$ **do**
    draw j randomly in $\{0, \ldots, i-1\}$
    $P(i,j) = \frac{\delta(j)}{\sum_q \delta(q)}$
    draw r randomly in $[0; 1]$
    if $r \leq P(i,j)$ *and* $(i,j) \notin G$ **then**
      add link (i,j) to G
      num_links_i ++
    end
  end
  $i = i + 1$
end

Scale-Free model

- heterogeneous degree distribution: power-law

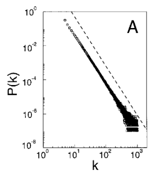


- but hidden properties : very few cycles
  ex: for $\alpha = 1$, it's a tree

Not at all a random element of the set of graphs with a power-law degree distribution...
Not a realistic generative model

Scale-Free model

- heterogeneous degree distribution: power-law



- but hidden properties : very few cycles
  ex: for $\alpha = 1$, it's a tree

Not at all a random element of the set of graphs with a power-law degree distribution...
Not a realistic generative model

Outline

Generative method: configuration model

Degree distribution

$p_1, p_2, p_3, \ldots$

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

Notice possible loops or multilinks

Generative method: configuration model

Degree distribution

$p_1, p_2, p_3, \ldots$

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

Notice possible loops or multilinks

Generative method: configuration model

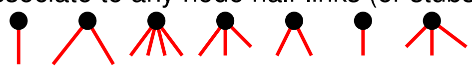
Degree distribution

$p_1, p_2, p_3, \ldots$

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

Notice possible loops or multilinks

Generative method: configuration model

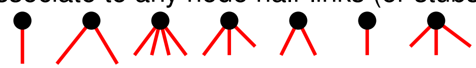
Degree distribution

$p_1, p_2, p_3, \ldots$

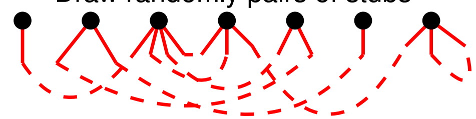
Draw nodes degree according to the distribution

1 2 4 3 2 1 3

Associate to any node half-links (or stubs)



Draw randomly pairs of stubs



Notice possible loops or multilinks

Generative method: configuration model

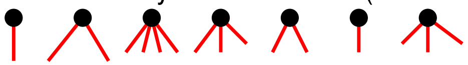
Degree distribution

$p_1, p_2, p_3, \ldots$

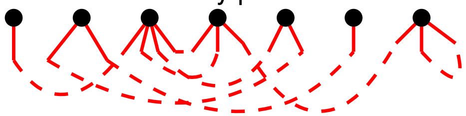
Draw nodes degree according to the distribution

1 2 4 3 2 1 3

Associate to any node half-links (or stubs)



Draw randomly pairs of stubs



Notice possible loops or multilinks

Implementing the configuration model

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

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

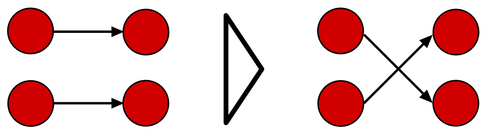
Algorithm 2: Generating a graph with fixed degree distribution

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

Other implementation: switching method

Principle

- we must start from 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



Other implementation: switching method

Why does it work?

- The degree of any node remains unchanged

  so we keep the distribution 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)

  if we make enough switches, we obtain a random graph with this degree distribution

Other implementation: switching method

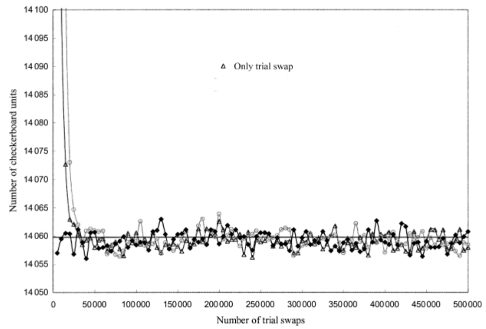
Why does it work?

- The degree of any node remains unchanged
  so we keep the distribution 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)
  if we make enough switches, we obtain a random graph with this degree distribution

Other implementation: switching method

When to stop switchings?

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



Properties – Comparison

	real	Erdös-Rényi	fixed d.d.
density	low	low	low
connectedness	giant comp.	giant comp.	giant comp.
distances	low	low	low
degree	heterogene	homogene	heterogene
clustering	high	low	low
communities	yes	no	no

→ clustering is not a consequence of heterogeneous degree

Graph with fixed degree distribution: the bipartite case

Newman, Watts et 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: *fr: graphe biparti*
  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

Graph with fixed degree distribution: the 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)