MU5IN075
Network Analysis and Mining
4. Random Graph Models I

Esteban Bautista-Ruiz, Lionel Tabourier

LIP6 – CNRS and Sorbonne Université

first_name.last_name@lip6.fr

October 5, 2021

Outline

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
→ observe common properties to the large majority of graphs
→ they are the expected properties

2: simulate processes

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
→ observe common properties to the large majority of graphs
→ they are the expected properties

2: simulate processes

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
→ observe common properties to the large majority of graphs
→ they are the expected properties

2: simulate processes

Outline

Erdős-Rényi model

version 1: $G_{n,p}$

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

Exercise: write a pseudocode to generate $G_{n,p}$

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

Erdős-Rényi model

version 1: $G_{n,p}$

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

Exercise: write a pseudocode to generate $G_{n,p}$

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

Erdős-Rényi model

version 2: $G_{n,m}$

- n nodes
- m edges chosen uniformly at random

Exercise: write a pseudocode to generate $G_{n,m}$

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

Erdős-Rényi model

version 2: $G_{n,m}$

- n nodes
- m edges chosen uniformly at random

Exercise: write a pseudocode to generate $G_{n,m}$

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

Erdős-Rényi model

version 2: $G_{n,m}$

- n nodes
- m edges chosen uniformly at random

Exercise: write a pseudocode to generate $G_{n,m}$

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

What about multiple edges and self-loops?

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

In a $G_{n,p}$ ER graph, probability p is the density (δ)

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

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

(note: strictly speaking, the 2 models are not equivalent)

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

In a $G_{n,p}$ ER graph, probability p is the density (δ)

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

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

(note: strictly speaking, the 2 models are not equivalent)

Degree distributions of ER graphs

Quiz: what is the expected degree distribution?

- probability of having exactly degree k for one node?

$$\mathcal{P}(k) = \binom{n-1}{k} p^k \cdot (1-p)^{n-1-k} \rightarrow \text{ binomial law}$$

- average number of neighbors $\langle k \rangle$?

$$\langle k \rangle = \sum_{k=0}^{n-1} \mathcal{P}(k).k = \ldots = (n-1)p$$

- standard deviation σ_k of the degree distribution?

$$\sigma_k = \sqrt{\langle k^2 \rangle - \langle k \rangle^2} = \ldots = \sqrt{\langle k \rangle}$$

Degree distributions of ER graphs

Quiz: what is the expected degree distribution?

- probability of having exactly degree k for one node?

$$\mathcal{P}(k) = \binom{n-1}{k} p^k \cdot (1-p)^{n-1-k} \rightarrow \text{ binomial law}$$

- average number of neighbors $\langle k \rangle$?

$$\langle k \rangle = \sum_{k=0}^{n-1} \mathcal{P}(k).k = \ldots = (n-1)p$$

- standard deviation σ_k of the degree distribution?

$$\sigma_k = \sqrt{\langle k^2 \rangle - \langle k \rangle^2} = \ldots = \sqrt{\langle k \rangle}$$

Degree distributions of ER graphs

Quiz: what is the expected degree distribution?

- probability of having exactly degree k for one node?

$$\mathcal{P}(k) = \binom{n-1}{k} p^k \cdot (1-p)^{n-1-k} \rightarrow \text{ binomial law}$$

- average number of neighbors $\langle k \rangle$?

$$\langle k \rangle = \sum_{k=0}^{n-1} \mathcal{P}(k).k = \ldots = (n-1)p$$

- standard deviation σ_k of the degree distribution?

$$\sigma_k = \sqrt{\langle k^2 \rangle - \langle k \rangle^2} = \ldots = \sqrt{\langle k \rangle}$$

Degree distributions of ER graphs

Quiz: what is the expected degree distribution?

- probability of having exactly degree k for one node?

$$\mathcal{P}(k) = \binom{n-1}{k} p^k \cdot (1-p)^{n-1-k} \rightarrow \text{ binomial law}$$

- average number of neighbors $\langle k \rangle$?

$$\langle k \rangle = \sum_{k=0}^{n-1} \mathcal{P}(k).k = \ldots = (n-1)p$$

- standard deviation σ_k of the degree distribution?

$$\sigma_k = \sqrt{\langle k^2 \rangle - \langle k \rangle^2} = \ldots = \sqrt{\langle k \rangle}$$

Notion of expected property

Example: Erdős-Rényi random graph $G_{n,m}$, $n = m = 4950$

Result: clique of 100 nodes and other 4850 nodes degree 0

Surprising?

Probability to have degree 0: $\mathcal{P}(k=0) = (1-p)^m \sim 0.135$.

$\Rightarrow$ Expected number of degree 0 nodes:

$\mathcal{P}(k=0) \cdot n \sim 670$ to be compared with 4850…

$\rightarrow$ seems very unlikely

how unlikely? statistical tests, out of NAM scope ($\rightarrow$ NDA)

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Notion of expected property

Example: Erdős-Rényi random graph $G_{n,m}$, $n = m = 4950$

Result: clique of 100 nodes and other 4850 nodes degree 0

Surprising?

Probability to have degree 0: $\mathcal{P}(k=0) = (1-p)^m \sim 0.135$.

$\Rightarrow$ Expected number of degree 0 nodes:

$\mathcal{P}(k=0) \cdot n \sim 670$ to be compared with 4850…

$\rightarrow$ seems very unlikely

how unlikely? statistical tests, out of NAM scope ($\rightarrow$ NDA)

9/30

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Notion of expected property

Example: Erdős-Rényi random graph $G_{n,m}$, $n = m = 4950$

Result: clique of 100 nodes and other 4850 nodes degree 0

Surprising?

Probability to have degree 0: $\mathcal{P}(k=0) = (1-p)^m \sim 0.135$.

$\Rightarrow$ Expected number of degree 0 nodes:

$\mathcal{P}(k=0) \cdot n \sim 670$ to be compared with 4850…

$\rightarrow$ seems very unlikely

how unlikely? statistical tests, out of NAM scope ($\rightarrow$ NDA)

9/30

Notion of expected property

Example: Erdős-Rényi random graph $G_{n,m}$, $n = m = 4950$

Result: clique of 100 nodes and other 4850 nodes degree 0

Surprising?
Probability to have degree 0: $\mathcal{P}(k = 0) = (1 - p)^m \sim 0.135$.

$\Rightarrow$ Expected number of degree 0 nodes:
$\mathcal{P}(k = 0) \cdot n \sim 670$ to be compared with 4850...

$\rightarrow$ seems very unlikely
how unlikely? statistical tests, out of NAM scope ($\rightarrow$ NDA)

Connectedness of ER graphs

Do ER graphs exhibit a giant component?

Experimental approach: the percolation phenomenon

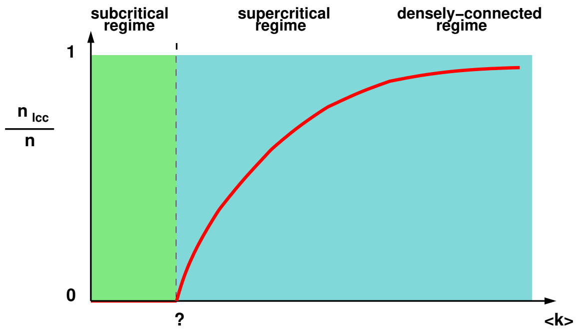
`https://www.complexity-explorables.org/explorables/the-blob/`

Connectedness of ER graphs

Do ER graphs exhibit a giant component?

Experimental approach: the percolation phenomenon

`https://www.complexity-explorables.org/explorables/the-blob/`



Connectedness of ER graphs

Do ER graphs exhibit a giant component?

Experimental approach: the percolation phenomenon

`https://www.complexity-explorables.org/explorables/the-blob/`

Some real graphs:

- Internet AS level (CAIDA data): $\langle k \rangle \simeq 4.03$
- US power grid (T. Opsahl data): $\langle k \rangle \simeq 2.67$
- Scientific collaborations (GaTech data): $\langle k \rangle \simeq 3.75$
- Yeast metabolic network (Y. Moreno data): $\langle k \rangle \simeq 2.44$
- Actor collaborations (Notre-Dame data): $\langle k \rangle \simeq 173.3$

$\Rightarrow$ Existence of a giant component explained by density alone

Connectedness of ER graphs

Do ER graphs exhibit a giant component?

Experimental approach: the percolation phenomenon

`https://www.complexity-explorables.org/explorables/the-blob/`

Some real graphs:

- Internet AS level (CAIDA data): $\langle k \rangle \simeq 4.03$
- US power grid (T. Opsahl data): $\langle k \rangle \simeq 2.67$
- Scientific collaborations (GaTech data): $\langle k \rangle \simeq 3.75$
- Yeast metabolic network (Y. Moreno data): $\langle k \rangle \simeq 2.44$
- Actor collaborations (Notre-Dame data): $\langle k \rangle \simeq 173.3$

$\Rightarrow$ Existence of a giant component explained by density alone

Clustering of ER graphs

Quiz: What clustering coefficient do you expect for ER graphs?

clue: think of the probabilistic interpretation of the cc

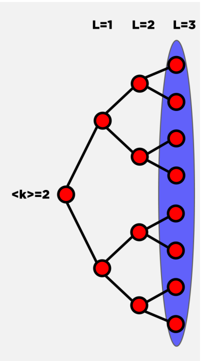
Clustering of ER graphs

Quiz: What clustering coefficient do you expect for ER graphs?

clue: think of the probabilistic interpretation of the cc

Distances in ER graphs

Rough hypothesis: no cycles...

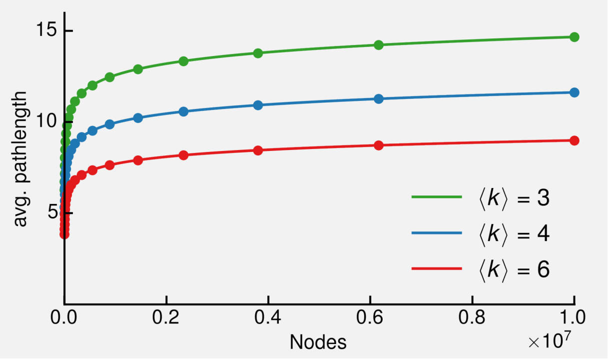


credits image: V.Gauthier

$\Rightarrow$ Short distances explained by density alone

Distances in ER graphs

Rough hypothesis: no cycles...



credits image: V.Gauthier

$\Rightarrow$ Short distances explained by density alone

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Summary of ER graphs properties

- Density
- Connectedness
- Average distance, diameter

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Summary of ER graphs properties

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

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Summary of ER graphs properties

- Degree distribution
- Clustering coefficient
- Communities

Summary of ER graphs properties

- Degree distribution homogeneous
- Clustering coefficient $\simeq$ density
- Communities no

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Summary of ER graphs properties

	real	ER
density	low	?
connectedness	giant comp.	?
distances	low	?
degree distrib.	heterogeneous	?
clustering	high	?
communities	yes	?

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Summary of ER graphs properties

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

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Conclusion on Erdős-Rényi graphs

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

Consequences

- Resemblances (connectedness, distances) can be explained with this simple model
- In general, not a good model for simulations, proofs ...

$\rightarrow$ Other models?

Outline

General idea of the Small-World model

Reminder: Milgram's Small-World experiment

Small-world: small average distance, high clustering

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

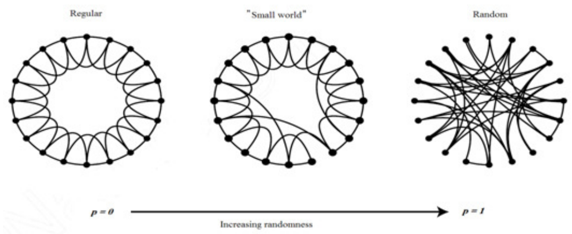
Watts et Strogatz - *Nature, 1998*

General idea of the Small-World model

Reminder: Milgram's Small-World experiment

Small-world: small average distance, high clustering

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



Watts et Strogatz - *Nature, 1998*

Degree distributions of SW graphs

Quiz: what is the expected degree distribution?

- regular graph?
- random graph?
- intermediary case?

Degree distributions of SW graphs

Quiz: what is the expected degree distribution?

- regular graph? all nodes have the same degree
- random graph?
- intermediary case?

Degree distributions of SW graphs

Quiz: what is the expected degree distribution?

- regular graph? all nodes have the same degree
- random graph? Erdős-Rényi case: binomial
- intermediary case?

Degree distributions of SW graphs

Quiz: what is the expected degree distribution?

- regular graph? all nodes have the same degree
- random graph? Erdős-Rényi case: binomial
- intermediary case? somewhere inbetween

$\Rightarrow$ homogeneous

Clustering in SW graphs

Quiz: what is expected for the average clustering coefficient?

- regular graph?
- random graph?
- intermediary case?

Clustering in SW graphs

Quiz: what is expected for the average clustering coefficient?

- regular graph? neighbors are connected: 1
- random graph?
- intermediary case?

Clustering in SW graphs

Quiz: what is expected for the average clustering coefficient?

- regular graph? neighbors are connected: 1
- random graph? Erdős-Rényi case: low (like density)
- intermediary case?

Clustering in SW graphs

Quiz: what is expected for the average clustering coefficient?

- regular graph? neighbors are connected: 1
- random graph? Erdős-Rényi case: low (like density)
- intermediary case? somewhere inbetween

Paths length in SW graphs

Quiz: what is the expected for the path length?

- regular graph?
- random graph?
- intermediary case?

Paths length in SW graphs

Quiz: what is the expected for the path length?

- regular graph? some nodes are far away
- random graph?
- intermediary case?

Paths length in SW graphs

Quiz: what is the expected for the path length?

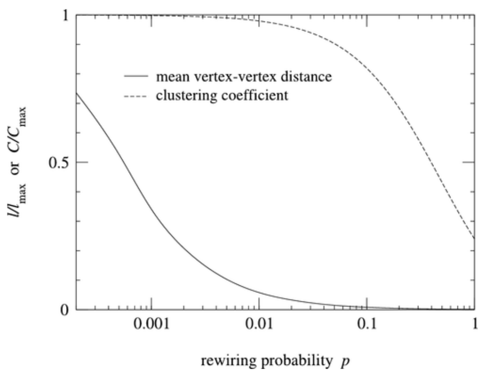
- regular graph? some nodes are far away
- random graph? Erdős-Rényi case $\sim log(n)$, short
- intermediary case?

Paths length in SW graphs

Quiz: what is the expected for the path length?

- regular graph? some nodes are far away
- random graph? Erdős-Rényi case $\sim log(n)$, short
- intermediary case? somewhere inbetween

Trade-off clustering/path length in SW graphs



https://mathinsight.org/applet/small_world_network

Small-World model implementation

Algorithm 1: WS generation

Parameters: p (rewiring proba); $G = (V, E)$: k-regular graph
foreach $u = (i, j)$ *in* E **do**
  $*$ $i' \leftarrow i; j' \leftarrow j$
  r_i random float $\in [0; 1]$
  if $r_i < p$ **then**
    draw a random node $n_i \in V \setminus \{i\}$
    $i' \leftarrow n_i$
  end
  r_j random float $\in [0; 1]$
  if $r_j < p$ **then**
    draw a random node $n_j \in V \setminus \{j\}$
    $j' \leftarrow n_j$
  end
  if no loop **and** no multi-edge **then** $u \leftarrow (i', j')$ **else** go to $*$
end

Summary of SW graphs properties

	real	small-world
density	low	?
connectedness	giant comp.	?
distances	low	?
degree distrib.	heterogeneous	?
clustering	high	?
communities	yes	?

Summary of SW graphs properties

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

Outline

Meaning of scale-free

fr: sans échelle ou invariant d'échelle

Reminder: degree distribution follows a power-law:

$$\mathcal{P}(k) = A \cdot k^{\alpha} + B$$

In practice for real networks:

$$\mathcal{P}(k) \propto k^{-\gamma} \text{ with } 2 \leq \gamma \leq 3$$

$\Rightarrow$ degree distrib. in log-scale is a line with slope $(-\gamma)$

Meaning of scale-free

fr: sans échelle ou invariant d'échelle

Reminder: degree distribution follows a power-law:

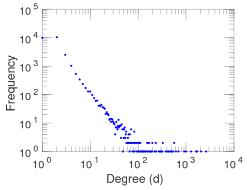
$$\mathcal{P}(k) = A \cdot k^{\alpha} + B$$

In practice for real networks:

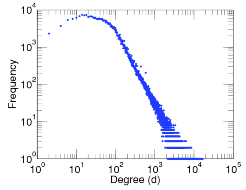
$$\mathcal{P}(k) \propto k^{-\gamma} \text{ with } 2 \leq \gamma \leq 3$$

$\Rightarrow$ degree distrib. in log-scale is a line with slope $(-\gamma)$

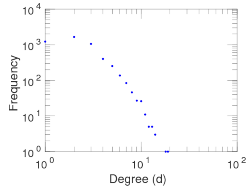
Internet AS



Actors



Powergrid



Scale-Free model

fr: modèle sans échelle

Barabási and 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 k_i

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

Ad hoc justification: generative process in agreement with the *"rich gets richer"* rule (or Merton's *"Matthew's effect"*)

Scale-Free model

fr: modèle sans échelle

Barabási and 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 k_i

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

Ad hoc justification: generative process in agreement with the *"rich gets richer"* rule (or Merton's *"Matthew's effect"*)

Scale-Free model

fr: modèle sans échelle

Barabási and 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 k_i

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

Ad hoc justification: generative process in agreement with the *"rich gets richer"* rule (or Merton's *"Matthew's effect"*)

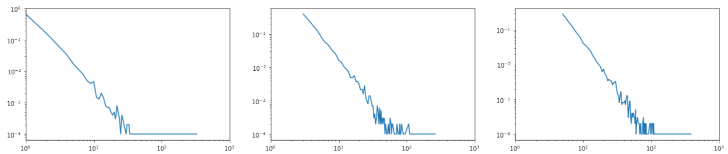
Barabási-Albert scale-free model

Algorithm 2: BA generation (n nodes, $(n-n_0)\alpha + m_0$ edges)

Parameters: n , G, α (degree arriving node)
with G connected graph with n_0 nodes and m_0 edges,
for i *from* (n_0+1) *to* n **do**
  add node i to G
  $num = 0$ // num: `number of links of` i
  while $num < \alpha$ **do**
    draw $j \in [\![0; i-1]\!]$ with probability $\mathcal{P}(j) = \frac{k_j}{\sum_{q=0}^{i-1} k_q}$
    if $(i,j) \notin G$ **then**
      add edge (i,j) in G
      $num{+}{+}$
    end
  end
end

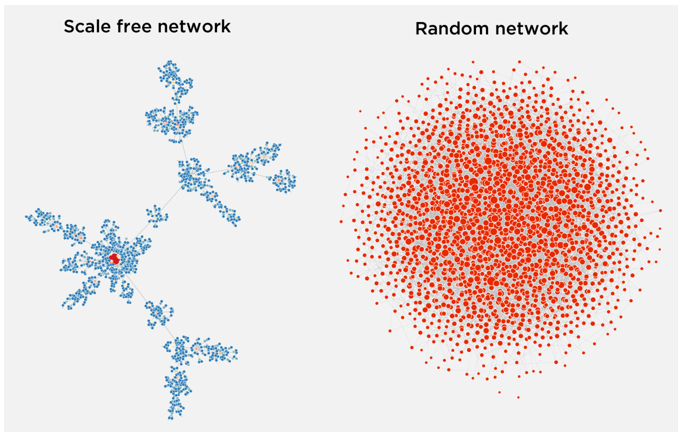
Degree distributions of BA graphs: influence of the parameters

For $\alpha = 1, 3, 5$:



Slope independent of m
Theoretical slope: $\gamma \simeq 3$

Visual comparison to ER graphs



credits image: V.Gauthier

Clustering of SW graphs

Quiz: what is the clustering coefficient in the $\alpha = 1$ case? other cases?

$\rightarrow$ clustering is very low, even null

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Clustering of SW graphs

Quiz: what is the clustering coefficient in the $\alpha = 1$ case? other cases?

$\rightarrow$ clustering is very low, even null

29/30

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Clustering of SW graphs

Quiz: spreading experiment: what do you expect comparing spreading on a real network and on a BA model?

$\rightarrow$ clustering (local density) tends to slow down spreading

29/30

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Clustering of SW graphs

Quiz: spreading experiment: what do you expect comparing spreading on a real network and on a BA model?

$\rightarrow$ clustering (local density) tends to slow down spreading

29/30

Properties of the *Scale-Free* model

	real	scale-free
density	low	?
connectedness	giant comp.	?
distances	low	?
degree distrib.	heterogeneous	?
clustering	high	?
communities	yes	?

Motivation
Graphs with given density: Erdős-Rényi graphs
Watts-Strogatz small-world model
Barabási-Albert scale-free model

Properties of the *Scale-Free* model

	real	scale-free
density	low	low
connectedness	giant comp.	giant comp.
distances	low	low
degree distrib.	heterogeneous	scale-free
clustering	high	low, even 0
communities	yes	no

30/30