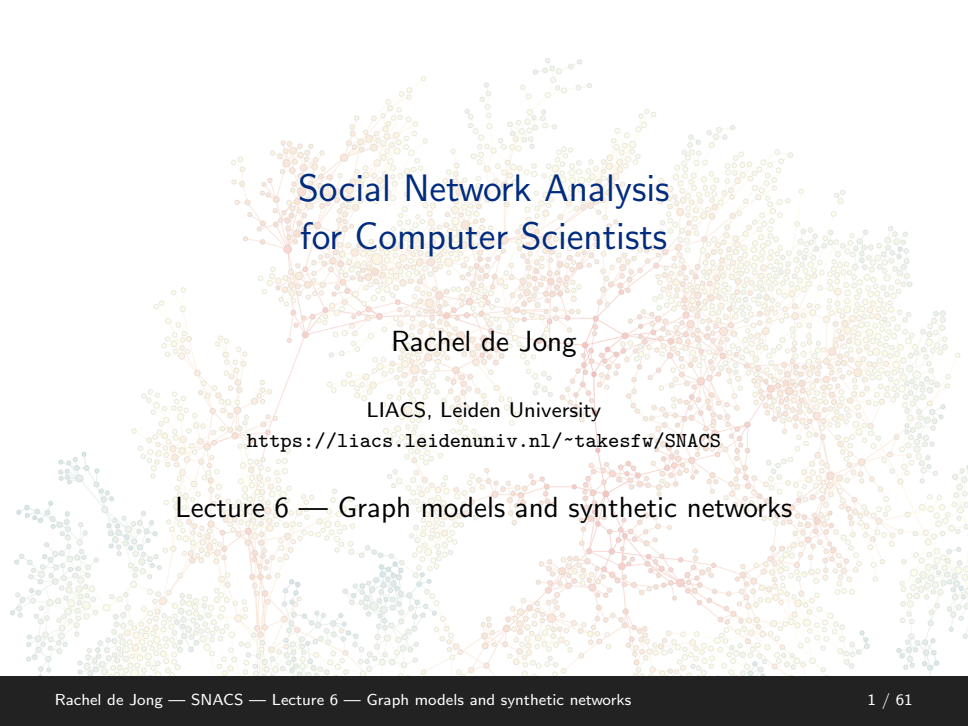




Social Network Analysis for Computer Scientists

Rachel de Jong

LIACS, Leiden University

<https://liacs.leidenuniv.nl/~takesfw/SNACS>

Lecture 6 — Graph models and synthetic networks

Today

- Recap
- Introduction
 - Why graph models?
- A brief history of network science, through graph models
 - The random graph (ER)
 - Explaining properties (WS, BA)
 - (More) realistic graph models (CM, SBM, LFR)
- Using graph models in experiments

Recap

Notation

Concept

- Network (graph)
- Nodes (objects, vertices, ...)
- Links (ties, relationships, ...)
 - Directed — $E \subseteq V \times V$ — “links”
 - Undirected — “edges”
- Number of nodes — $|V|$
- Number of edges — $|E|$
- Degree of node u
- Distance from node u to v

Symbol

$G = (V, E)$

V

E

n

m

$\deg(u)$

$d(u, v)$

Real-world networks

- 1 Sparse networks density
 - 2 Fat-tailed power-law degree distribution degree
 - 3 Giant component components
 - 4 Low pairwise node-to-node distances distance
 - 5 Many triangles clustering coefficient
- Many examples: social networks, communication networks, citation networks, collaboration networks, protein interaction networks, information networks, webgraphs, financial networks, ...

Advanced concepts

- Reciprocity
- Complete graph, complement of a graph
- Planar graph, lattice, regular graph
- Tree and dag
- Subgraphs
- Ego network
- Spanning trees
- Diameter
- Bridges
- Power law exponent

Centrality measures

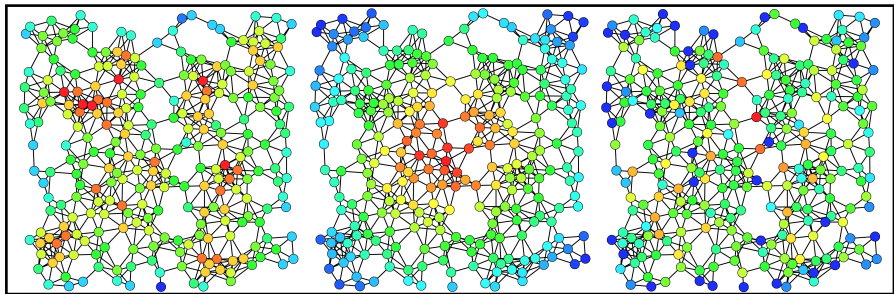


Figure: Degree, closeness and betweenness centrality.

Source: "Centrality" by Claudio Rocchini, Wikipedia File:Centrality.svg

Attribute assortativity

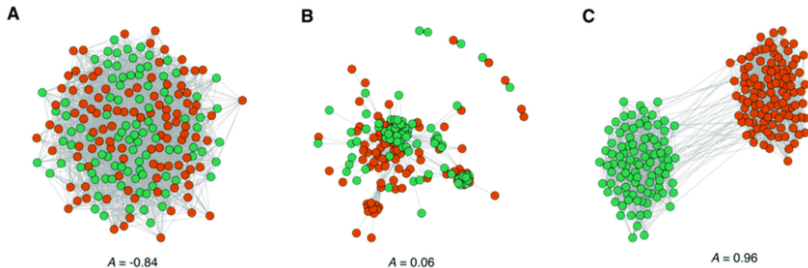


Figure: Attribute assortativity.

Image: Moya-García, A. et al. Identification of New Toxicity Mechanisms ... *Genes*, 13(7), 1292, 2022.

Projection

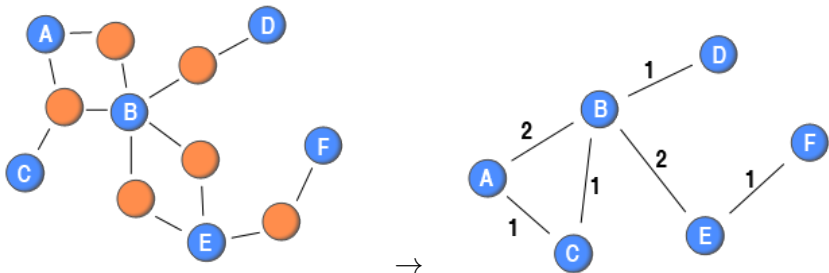


Image: <http://toreopsahl.com>

Community detection

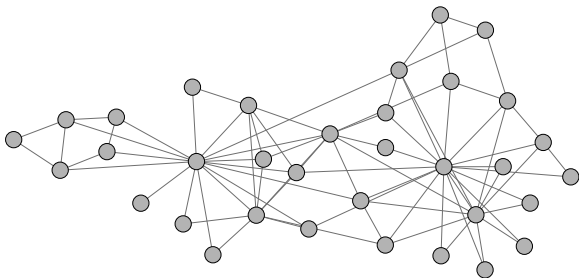


Figure: Communities: node subsets connected more strongly with each other.

- Partitioning a given graph into densely connected groups of nodes

Community detection

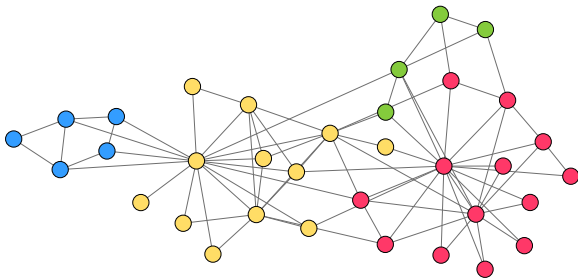


Figure: Communities: node subsets connected more strongly with each other.

- Divisive algorithms (Girvan-Newman), local search algorithms (Louvain) optimizing a quality function (e.g., modularity), label propagation and random walk based approaches.

Core/periphery structure

- Many large networks have a densely connected **core** and a **periphery** of nodes attaching to only a few neighbors
- **Core/periphery structure**: there is **one** dense area
- **Community structure**: there are **several**, with little periphery

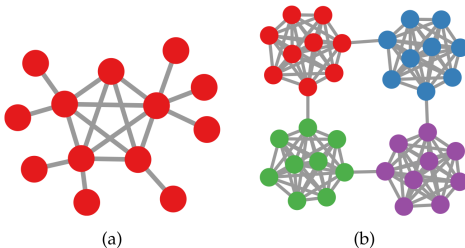
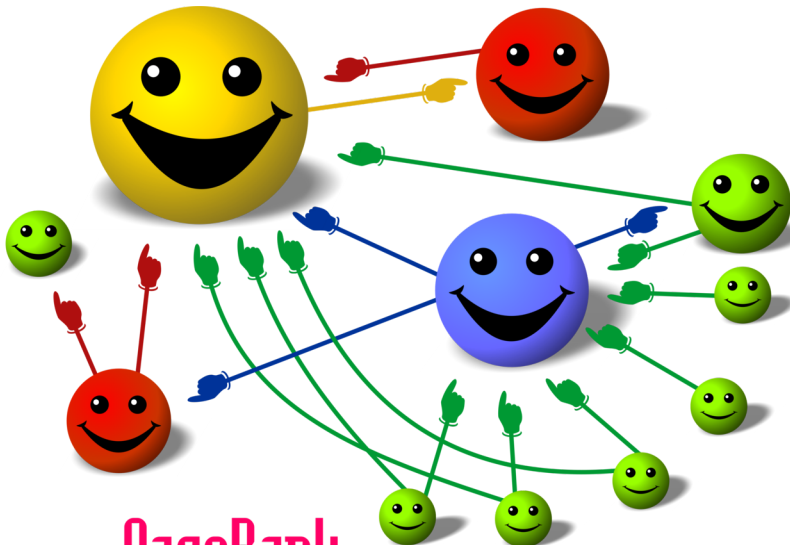
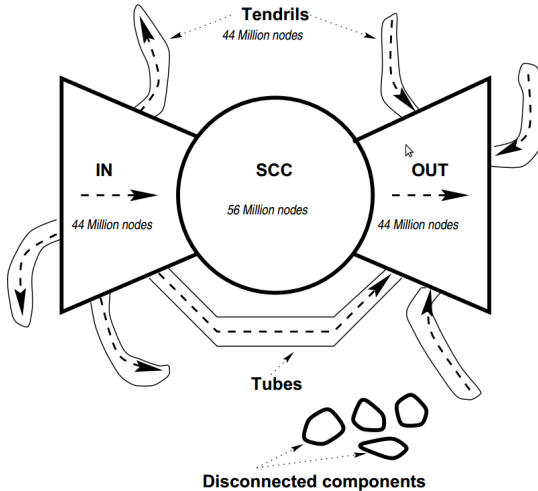


Figure: (a) Core/periphery structure and (b) Community structure.



PageRank

Webgraph in 1999



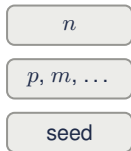
Broder et al., Graph structure in the web, Computer Networks 33(1): 309–320, 2000.

Graph models

Graph models and synthetic networks

- **Graph model:** A rule for generating networks with specific structural properties. Typically defined by a set of parameters and a generation mechanism.
- **Synthetic network:** a network generated by a graph model

Input parameters



Graph model

generation mechanism
ER, WS, BA, CM, ...



Synthetic network



Why graph models?

Why graph models?

- **Test your algorithm and analysis**

- How does the algorithm scale to networks with more nodes, edges, clustering...
- Will the analyses remain the same?

Why graph models?

- **Test your algorithm and analysis**

- How does the algorithm scale to networks with more nodes, edges, clustering...
- Will the analyses remain the same?

- **Explanatory:** why does this property arise?

- Apply rules, and see if a property emerges

Why graph models?

■ **Test your algorithm and analysis**

- How does the algorithm scale to networks with more nodes, edges, clustering...
- Will the analyses remain the same?

■ **Explanatory:** why does this property arise?

- Apply rules, and see if a property emerges

■ **Test statistical significance of a property**

- Is this property really strange?
- Does it not emerge in this model as well?

Why graph models?

- **Test your algorithm and analysis**

- How does the algorithm scale to networks with more nodes, edges, clustering...
- Will the analyses remain the same?

- **Explanatory:** why does this property arise?

- Apply rules, and see if a property emerges

- **Test statistical significance of a property**

- Is this property really strange?
- Does it not emerge in this model as well?

- Or... if you do not have network data

ON THE EVOLUTION OF RANDOM GRAPHS

by

P. ERDŐS and A. RÉNYI

*Dedicated to Professor P. Turán at
his 50th birthday.*

Introduction

Our aim is to study the probable structure of a random graph $\Gamma_{n,N}$ which has n given labelled vertices $P_1, P_2, \dots, P_n$ and N edges; we suppose that these N edges are chosen at random among the $\binom{n}{2}$ possible edges,

so that all $\binom{\binom{n}{2}}{N} = C_{n,N}$ possible choices are supposed to be equiprobable. Thus if $G_{n,N}$ denotes any one of the $C_{n,N}$ graphs formed from n given labelled points and having N edges, the probability that the random graph $\Gamma_{n,N}$ is identical with $G_{n,N}$ is $\frac{1}{C_{n,N}}$. If A is a property which a graph may or may not possess, we denote by $\mathbf{P}_{n,N}(A)$ the probability that the random graph $\Gamma_{n,N}$ possesses the property A , i. e. we put $\mathbf{P}_{n,N}(A) = \frac{A_{n,N}}{C_{n,N}}$ where $A_{n,N}$ denotes the number of those $G_{n,N}$ which have the property A .

Figure: Pál Erdős, Alfréd Rényi, "On the evolution of random graphs", 1960.

Random Graphs (ER)

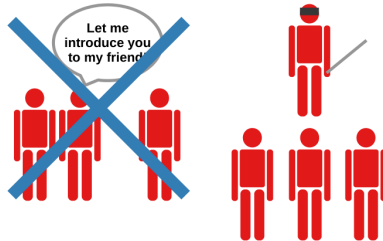


Figure: The creation of a random graph.

Random Graphs (ER)

Generating **random graphs**:

- $G_{n,p}$: **The Erdős-Rényi model (ER)**

- Input: n, p
- For every possible edge:
- Add with probability p
- Total number of edges may differ slightly

Random Graphs (ER)

Generating **random graphs**:

- $G_{n,p}$: **The Erdős-Rényi model (ER)**

- Input: n, p
- For every possible edge:
- Add with probability p
- Total number of edges may differ slightly

- $G_{n,m}$: Gilbert model

- Input: n, m
- Pick m edges at random
- Number of edges is set

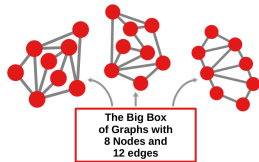


Figure: Three random graphs with 8 nodes and 12 edges.

Degree distributions in ER graphs

- Degrees follow a binomial distribution ($\approx$ Poisson distribution for large n)
- Average degree: $\bar{k} = p \cdot (n - 1)$

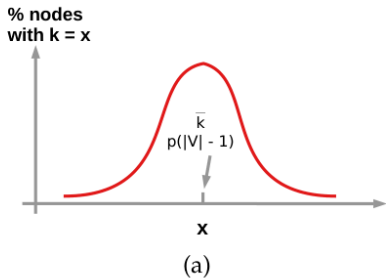


Figure: Degree distributions in random graphs.

Degree distributions in ER graphs

- Degrees follow a binomial distribution ($\approx$ Poisson distribution for large n)
- Average degree: $\bar{k} = p \cdot (n - 1)$
- Not the same as real-world networks.

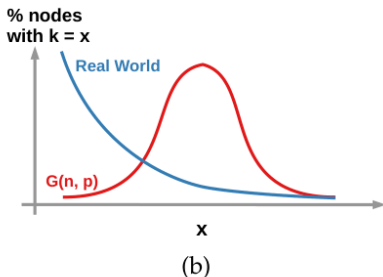
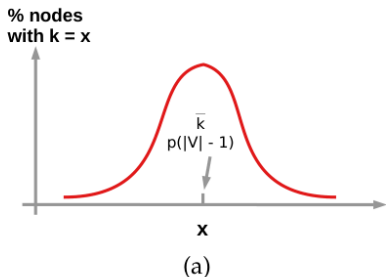
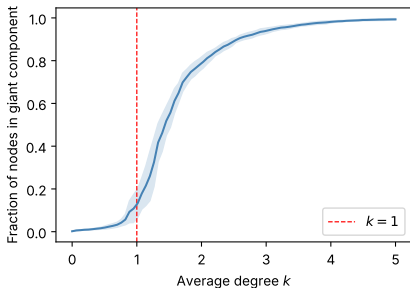


Figure: Degree distributions in random graphs (a) compared to real-world networks (b).

Components and distances of ER graphs

■ Average degree 1: formation of giant component

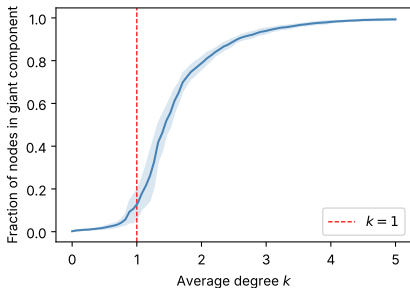


(a)

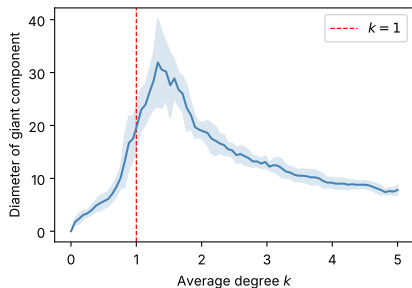
Figure: Giant component (a) and diameter (b) in random graphs ($n = 500$).

Components and distances of ER graphs

- Average degree 1: formation of giant component
- Diameter decreases when giant component contains most nodes



(a)

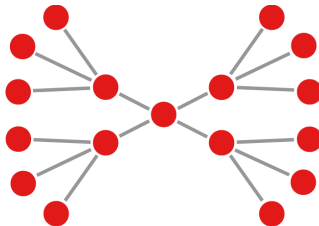


(b)

Figure: Giant component (a) and diameter (b) in random graphs ($n = 500$).

Distances in ER graphs

- Usually short distances
- For example, average degree 4
 - 4 nodes at distance 1
 - $4 + 4 \cdot 3 = 16$ nodes within distance 2
 - $4 + 4 \cdot 3 + (4 \cdot 3) \cdot 3 = 52$ nodes within distance 3



Recall: properties of real-world networks

- | | | |
|---|--|------------------------|
| 1 | Fat-tailed power-law degree distribution | degree |
| 2 | Giant component | components |
| 3 | Low pairwise node-to-node distances | distance |
| 4 | Many triangles | clustering coefficient |
| 5 | Often a clear community structure | modularity |

Which properties match with those of real-world networks?

■ Which properties are realistic in a ER / random graph?

- 1 Degree distribution?
- 2 Giant component?
- 3 Clustering coefficient?
- 4 Short distances?
- 5 Communities?

Which properties match with those of real-world networks?

■ Which properties are realistic in a ER / random graph?

- 1 Degree distribution: No
- 2 Giant component?
- 3 Clustering coefficient?
- 4 Short distances?
- 5 Communities?

Which properties match with those of real-world networks?

■ Which properties are realistic in a ER / random graph?

- 1 Degree distribution: No
- 2 Giant component: Yes
- 3 Clustering coefficient?
- 4 Short distances?
- 5 Communities?

Which properties match with those of real-world networks?

■ Which properties are realistic in a ER / random graph?

- 1 Degree distribution: No
- 2 Giant component: Yes
- 3 Clustering coefficient: No
- 4 Short distances?
- 5 Communities?

Which properties match with those of real-world networks?

■ Which properties are realistic in a ER / random graph?

- 1 Degree distribution: No
- 2 Giant component: Yes
- 3 Clustering coefficient: No
- 4 Short distances: Yes
- 5 Communities?

Which properties match with those of real-world networks?

■ Which properties are realistic in a ER / random graph?

- 1 Degree distribution: No
- 2 Giant component: Yes
- 3 Clustering coefficient: No
- 4 Short distances: Yes
- 5 Communities: No

Graph models and properties

	Degree distribution	Giant component	Clustering coefficient	Short distances	Communities
Erdős–Rényi (ER)		✓		✓	

✓ = yes, for most inputs * = depends on the input ** = possible with existing extensions

Late 1990s: when data became available...

[nature](#) > [letters](#) > [article](#)

Letter | Published: 04 June 1998

Collective dynamics of 'small-world' networks

[Duncan J. Watts](#) & [Steven H. Strogatz](#)

Nature **393**, 440–442 (1998) | [Cite this article](#)

[Save article](#)

227k Accesses | 37k Citations | 649 Altmetric | [Metrics](#)

Emergence of Scaling in Random Networks

ALBERT-LÁSZLÓ BARABÁSI AND RÉKA ALBERT [Authors Info & Affiliations](#)

SCIENCE • 15 Oct 1999 • Vol 286, Issue 5439 • pp. 509–512 • DOI: 10.1126/science.286.5439.509

37,623 27,680



- Real-world networks behave differently than random graphs!
- Finding and explaining universal patterns
 - The small world property
 - Scale free-ness
- Most cited network science papers to this day

Watts–Strogatz model

Also called the small world model

- 1 Put n nodes on a ring

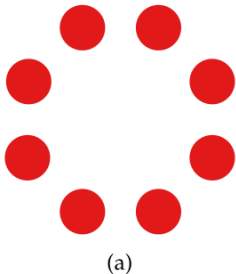


Figure: The generation process of a WS graph.

Watts–Strogatz model

Also called the small world model

- 1 Put n nodes on a ring
- 2 Connect each node to m nearest neighbors

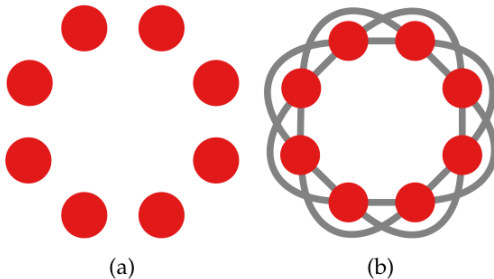


Figure: The generation process of a WS graph.

Watts–Strogatz model

Also called the small world model

- 1 Put n nodes on a ring
- 2 Connect each node to m nearest neighbors
- 3 Rewire edges with probability p_r

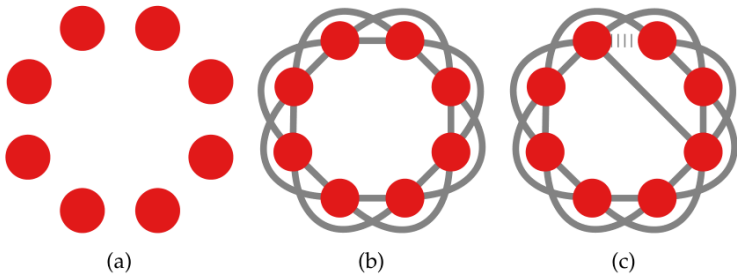
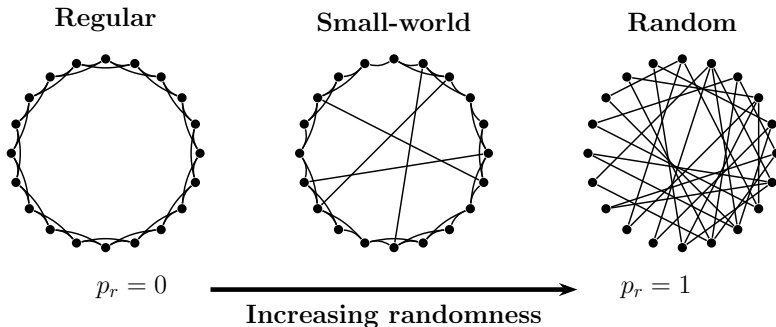


Figure: The generation process of a WS graph.

Watts–Strogatz model



Based on: Watts, D., Strogatz, S. Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998).

Which properties?

- Which properties are realistic in a Watts Strogatz graph?

- 1 Degree distribution?

- 2 Giant component?

- 3 Clustering coefficient?

- 4 Short distances?

- 5 Communities?

Which properties?

- Which properties are realistic in a Watts Strogatz graph?

- 1 Degree distribution: most nodes same degree... still no

- 2 Giant component?

- 3 Clustering coefficient?

- 4 Short distances?

- 5 Communities?

Which properties?

- Which properties are realistic in a Watts Strogatz graph?

- 1 Degree distribution: most nodes same degree... still no

- 2 Giant component: yes

- 3 Clustering coefficient?

- 4 Short distances?

- 5 Communities?

Which properties?

- Which properties are realistic in a Watts Strogatz graph?

- 1 Degree distribution: most nodes same degree... still no

- 2 Giant component: yes

- 3 Clustering coefficient: yes, but most connected to nearest neighbors.

- 4 Short distances?

- 5 Communities?

Which properties?

- Which properties are realistic in a Watts Strogatz graph?

- 1 Degree distribution: most nodes same degree... still no
- 2 Giant component: yes
- 3 Clustering coefficient: yes, but most connected to nearest neighbors.
- 4 Short distances: yes
- 5 Communities?

Which properties?

■ Which properties are realistic in a Watts Strogatz graph?

- 1 Degree distribution: most nodes same degree... still no
- 2 Giant component: yes
- 3 Clustering coefficient: yes, but most connected to nearest neighbors.
- 4 Short distances: yes
- 5 Communities: no

But...

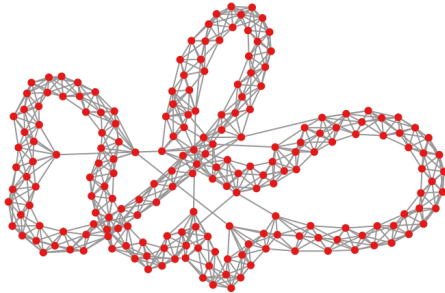


Figure: An example of a WS graph that looks a bit different from a real-world network.

Graph models and properties

	Degree distribution	Giant component	Clustering coefficient	Short distances	Communities
Erdős-Rényi (ER)		✓		✓	
Watts-Strogatz (WS)		✓	✓	✓	

✓ = yes, for most inputs * = depends on the input ** = possible with existing extensions

Preferential attachment

- Studying the web: most webpages have few links while only few have many
- Degree distribution is scale free
- This is unlike a random graph!
- There must be a different mechanism at play:

Preferential attachment

- Studying the web: most webpages have few links while only few have many
- Degree distribution is scale free
- This is unlike a random graph!
- There must be a different mechanism at play:
 - Cumulative advantage
 - Rich get richer
 - Think of social media, scientific papers...

BA model

- 1 Start with $m_0 \geq m$ nodes in a star / chain / clique
- 2 For each remaining node:

- 1 Connect to m nodes

- 2 $p(\{v, w\}) = \frac{\deg(w)}{\sum_{u \in V} \deg(u)}$

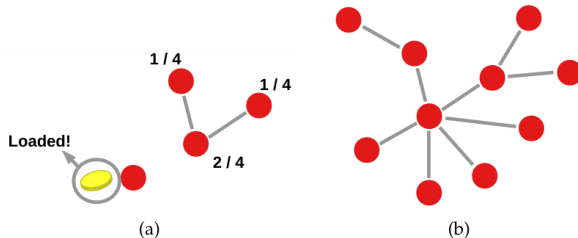


Figure: Choosing which edge to form (a), an example of a BA graph (b).

BA model

- Mechanism models that the graph keeps *growing*
- Different from temporal networks

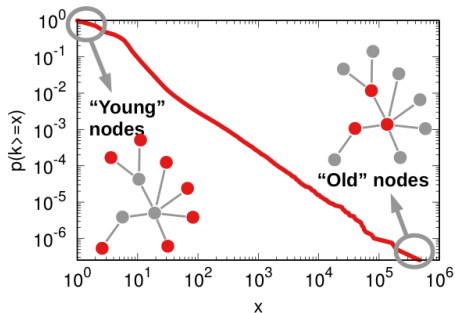


Figure: Cumulative advantage in BA models.

Which properties are realistic?

- Which properties are realistic in a Barabási–Albert graph?

- 1 Degree distribution?

- 2 Giant component?

- 3 Clustering coefficient?

- 4 Short distances?

- 5 Communities?

Which properties are realistic?

- Which properties are realistic in a Barabási–Albert graph?

- 1 Degree distribution: yes!
- 2 Giant component?
- 3 Clustering coefficient?
- 4 Short distances?
- 5 Communities?

Which properties are realistic?

- Which properties are realistic in a Barabási–Albert graph?

- 1 Degree distribution: yes!
- 2 Giant component: yes
- 3 Clustering coefficient?
- 4 Short distances?
- 5 Communities?

Which properties are realistic?

- Which properties are realistic in a Barabási–Albert graph?

- 1 Degree distribution: yes!
- 2 Giant component: yes
- 3 Clustering coefficient: better, but not there yet
- 4 Short distances?
- 5 Communities?

Which properties are realistic?

- Which properties are realistic in a Barabási–Albert graph?

- 1 Degree distribution: yes!
- 2 Giant component: yes
- 3 Clustering coefficient: better, but not there yet
- 4 Short distances: yes
- 5 Communities?

Which properties are realistic?

- Which properties are realistic in a Barabási–Albert graph?

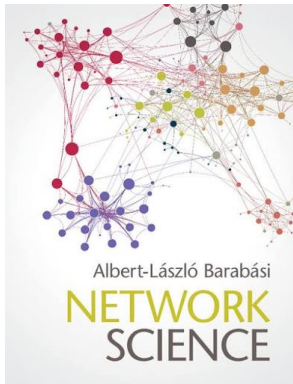
- 1 Degree distribution: yes!
- 2 Giant component: yes
- 3 Clustering coefficient: better, but not there yet
- 4 Short distances: yes
- 5 Communities: no

Graph models and properties

	Degree distribution	Giant component	Clustering coefficient	Short distances	Communities
Erdős–Rényi (ER)		✓		✓	
Watts–Strogatz (WS)		✓	✓	✓	
Barabási–Albert (BA)	✓	✓		✓	

✓ = yes, for most inputs * = depends on the input ** = possible with existing extensions

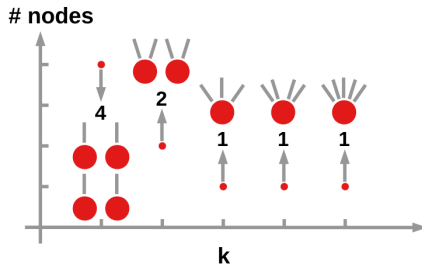
More about the BA model



- Fun read Chapter 0: A personal introduction describes the story and findings of how they arrived at this model.
- networksciencebook.com

The configuration model (CM)

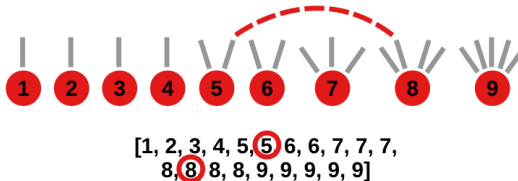
- Input: degree distribution
- Sum of degrees should be an even number
- Output: A graph with this degree distribution
- Create nodes with *stubs* that can form edges



The configuration model (CM)

■ The Molloy–Reed approach

- 1 Create n nodes with each $\deg(v)$ stubs stated in the degree distribution
- 2 Repeatedly connect two entries
- 3 Number of connections of entry equals its degree



The configuration model (CM)

- Molloy–reed approach:
- Sometimes (insignificant) number of unassigned edges
- Ok for testing algorithms
- May need exact distribution when comparing to real data
- Optionally add self-loops and multi-edges to unassigned nodes

Graph models and properties

	Degree distribution	Giant component	Clustering coefficient	Short distances	Communities
Erdős–Rényi (ER)		✓		✓	
Watts–Strogatz (WS)		✓	✓	✓	
Barabási–Albert (BA)	✓	✓		✓	
Configuration model (CM)	✓		**	✓	

✓ = yes, for most inputs * = depends on the input ** = possible with existing extensions

Stochastic block model (SBM)

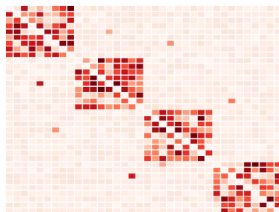
- Observation: Nodes in the same community are more tightly connected with each other, than nodes outside of the community

Stochastic block model (SBM)

- Observation: Nodes in the same community are more tightly connected with each other, than nodes outside of the community
- p_{in} : Probability of edges to form **within** communities
- p_{out} : Probability of edges to form **between** communities
- C : assignment of nodes into communities
- Usually $p_{in} > p_{out}$

Stochastic block model (SBM)

- p_{in} : Probability of edges to form **within** communities
- p_{out} : Probability of edges to form **between** communities
- C : assignment of nodes into communities
- Usually $p_{in} > p_{out}$



(a)

Figure: Matrix containing probability of two nodes to connect, darker colors indicate a higher probability (a).

Stochastic block model (SBM)

- p_{in} : Probability of edges to form **within** communities
- p_{out} : Probability of edges to form **between** communities
- C : assignment of nodes into communities
- Usually $p_{in} > p_{out}$

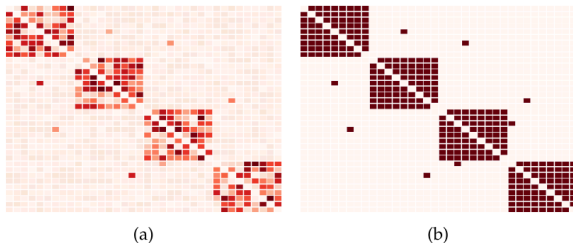


Figure: Matrix containing probability of two nodes to connect, darker colors indicate a higher probability (a), and a realization of the SBM using the matrix as input (b). Dark red cells indicate edges.

Stochastic block model (SBM) variants

- Degree-corrected SBM
 - Fix degree distribution, similar to configuration model (CM)
- Mixed-Membership SBM
 - Nodes can be in multiple communities at the same time

Graph models and properties

	Degree distribution	Giant component	Clustering coefficient	Short distances	Communities
Erdős–Rényi (ER)		✓		✓	
Watts–Strogatz (WS)		✓	✓	✓	
Barabási–Albert (BA)	✓	✓		✓	
Configuration model (CM)	✓		**	✓	
Stochastic block model (SBM)	**	*			✓

✓ = yes, for most inputs * = depends on the input ** = possible with existing extensions

LFR benchmark

- Test case for community discovery algorithms.
- Goal: produce networks with realistic properties.
- Hence, LFR needs many parameters:

LFR benchmark

- Test case for community discovery algorithms.
- Goal: produce networks with realistic properties.
- Hence, LFR needs many parameters:
 - α : power-law exponent of degree distribution

LFR benchmark

- Test case for community discovery algorithms.
- Goal: produce networks with realistic properties.
- Hence, LFR needs many parameters:
 - α : power-law exponent of degree distribution
 - β : exponent of the power law size distribution of the communities in the graph

LFR benchmark

- Test case for community discovery algorithms.
- Goal: produce networks with realistic properties.
- Hence, LFR needs many parameters:
 - α : power-law exponent of degree distribution
 - β : exponent of the power law size distribution of the communities in the graph
 - n : number of nodes

LFR benchmark

- Test case for community discovery algorithms.
- Goal: produce networks with realistic properties.
- Hence, LFR needs many parameters:
 - α : power-law exponent of degree distribution
 - β : exponent of the power law size distribution of the communities in the graph
 - n : number of nodes
 - k : average degree of the nodes (or k_{min} and k_{max})
 - s_{min} , s_{max} : community size

LFR benchmark

- Test case for community discovery algorithms.
- Goal: produce networks with realistic properties.
- Hence, LFR needs many parameters:
 - α : power-law exponent of degree distribution
 - β : exponent of the power law size distribution of the communities in the graph
 - n : number of nodes
 - k : average degree of the nodes (or k_{min} and k_{max})
 - s_{min}, s_{max} : community size
 - μ : mixing parameter, regulating the fraction of edges going outside their planted communities

LFR benchmark

- Test case for community discovery algorithms.
- Goal: produce networks with realistic properties.
- Hence, LFR needs many parameters:
 - α : power-law exponent of degree distribution
 - β : exponent of the power law size distribution of the communities in the graph
 - n : number of nodes
 - k : average degree of the nodes (or k_{min} and k_{max})
 - s_{min}, s_{max} : community size
 - μ : mixing parameter, regulating the fraction of edges going outside their planted communities
 - o_n : fraction of nodes overlapping between multiple communities
 - o_m : number of communities to which overlapping nodes belong

LFR benchmark

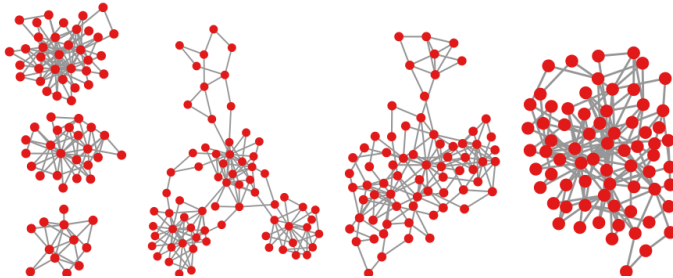


Figure: LFR models with $\mu = 0, 0.05, 0.2, 1$.

Graph models and properties

	Degree distribution	Giant component	Clustering coefficient	Short distances	Communities
Erdős–Rényi (ER)		✓		✓	
Watts–Strogatz (WS)		✓	✓	✓	
Barabási–Albert (BA)	✓	✓		✓	
Configuration model (CM)	✓		**	✓	
Stochastic block model (SBM)	**	*			✓
LFR benchmark	✓	✓	✓	✓	✓

✓ = yes, for most inputs * = depends on the input ** = possible with existing extensions

Graph models and parameters

- Graph models generate more realistic networks...
- ... but require more parameters
- Losing their explanatory nature
- Which one(s) to use depends on the goal

	Parameters
Erdős–Rényi (ER)	n, p
Watts–Strogatz (WS)	n, m, p_r
Barabási–Albert (BA)	n, m
Configuration model (CM)	degree distribution
Stochastic block model (SBM)	n, C, p_{in}, p_{out}
LFR benchmark	$\alpha, \beta, n, k, s_{min}, s_{max}, \mu, o_n, o_m$

Other works...

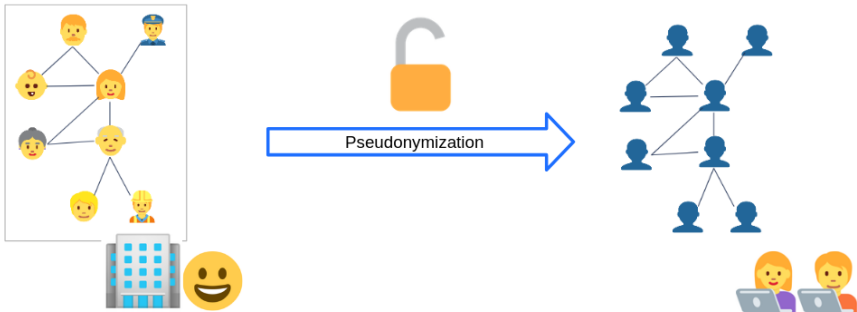
- Many other models exist: caveman model, random geometric graph, Kronecker graph, GN benchmark, ERGM and many more
- Graph neural networks
 - Goal: NN learns what a network should look like
- Digital twins: build a network based on a real-world network

Using graph models in experiments

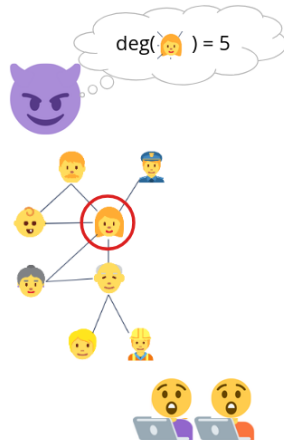
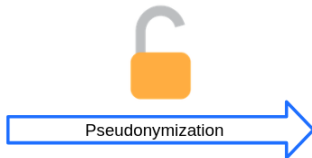
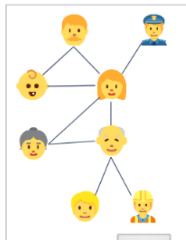
Using graph models in experiments

- Runtimes: Anonymity measurement
- Analysis: How does this property evolve

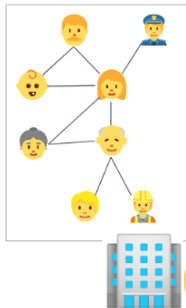
k -anonymity in networks



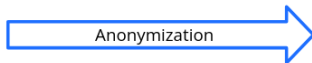
k -anonymity in networks



k -anonymity in networks

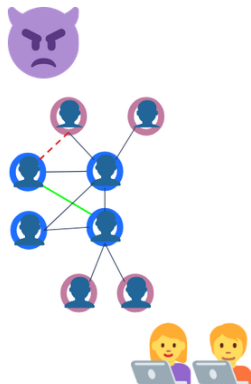


- **How to measure anonymity?**
- Preserving utility?
- Anonymization?



Deleting edges ---

Adding edges ---



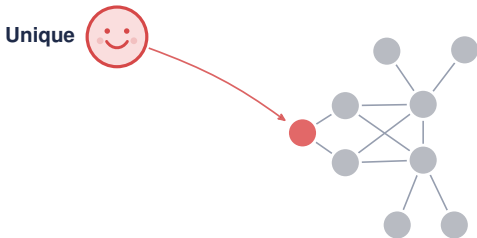
Measuring anonymity in networks

- **k -anonymity**: a node is k -anonymous if it is *equivalent* to at least $k - 1$ other nodes
- Equivalence if they have *isomorphic d -neighborhoods*



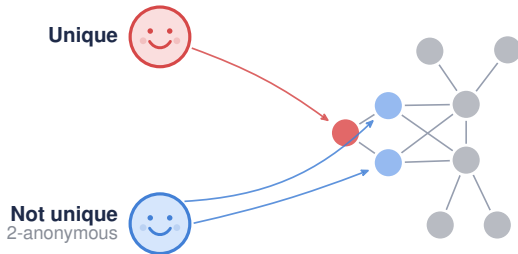
Measuring anonymity in networks

- **k -anonymity**: a node is k -anonymous if it is *equivalent* to at least $k - 1$ other nodes
- Equivalence if they have *isomorphic* d -neighborhoods



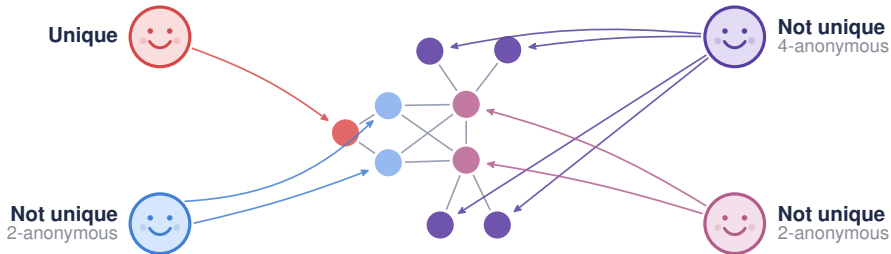
Measuring anonymity in networks

- **k -anonymity**: a node is k -anonymous if it is *equivalent* to at least $k - 1$ other nodes
- Equivalence if they have *isomorphic d -neighborhoods*



Measuring anonymity in networks

- **k -anonymity**: a node is k -anonymous if it is *equivalent* to at least $k - 1$ other nodes
- Equivalence if they have *isomorphic d -neighborhoods*



Measuring anonymity in networks

- Need to determine for each node pair if their d -neighborhoods are isomorphic

Measuring anonymity in networks

- Need to determine for each node pair if their d -neighborhoods are isomorphic
- **Isomorphism** is known to be computationally expensive!
- Solution: introduce computational approaches to speed-up computation.
- Iterative computation, use graph invariants

Measuring anonymity in networks

- Need to determine for each node pair if their d -neighborhoods are isomorphic
- **Isomorphism** is known to be computationally expensive!
- Solution: introduce computational approaches to speed-up computation.
- Iterative computation, use graph invariants
- Benchmark on real-world networks, and **graph models**.

Measuring anonymity in networks

- Need to determine for each node pair if their d -neighborhoods are isomorphic
- **Isomorphism** is known to be computationally expensive!
- Solution: introduce computational approaches to speed-up computation.
- Iterative computation, use graph invariants
- Benchmark on real-world networks, and **graph models**.
- What research questions can this benchmarking answer?
 - Can the algorithms handle dense graphs? Large graphs?
 - How does runtime change when density increases?

Measuring anonymity in networks

- Need to determine for each node pair if their d -neighborhoods are isomorphic
- **Isomorphism** is known to be computationally expensive!
- Solution: introduce computational approaches to speed-up computation.
- Iterative computation, use graph invariants
- Benchmark on real-world networks, and **graph models**.
- What research questions can this benchmarking answer?
 - Can the algorithms handle dense graphs? Large graphs?
 - How does runtime change when density increases?

Based on: de Jong, R.G., van der Loo, M.P.J., Takes, F.W. "Algorithms for efficiently computing structural anonymity in complex networks." *ACM JEA* 28 (2023): 1–22 and de Jong, R.G., van der Loo, M.P.J., Takes, F.W. "The effect of distant connections on node anonymity in complex networks." *Scientific Reports*, (2024) 14(1), 1156.

Measuring anonymity in networks

- ER, BA, WS $n = 1,000$, $m = \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$
- Heuristics (green) mostly the fastest.
- Surprising: takes longest for average degree 1! (sparse graph)

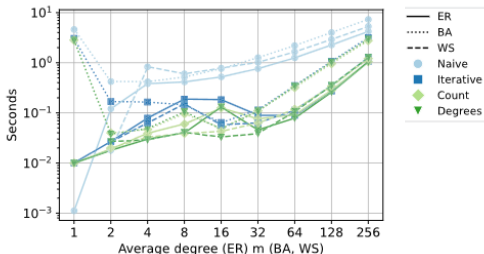
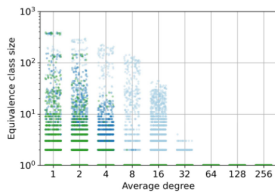


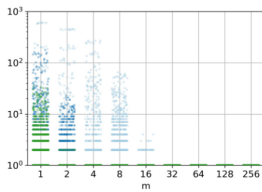
Figure: Runtime in ER, BA, WS models with $m = \{2^0, 2^1, \dots, 2^8\}$.

Measuring anonymity in networks

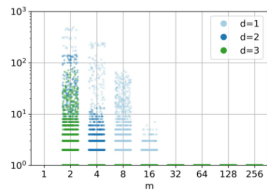
- Equivalence class size = k value of a node
- High degree $\rightarrow$ all nodes anonymous



(a) Erdős-Rényi (ER) model



(b) Barabási-Albert (BA) model



(c) Watts-Strogatz (WS) model

Figure: Anonymity in ER, BA, WS models with $m = \{2^0, 2^1, \dots, 2^8\}$.

Measuring anonymity in networks

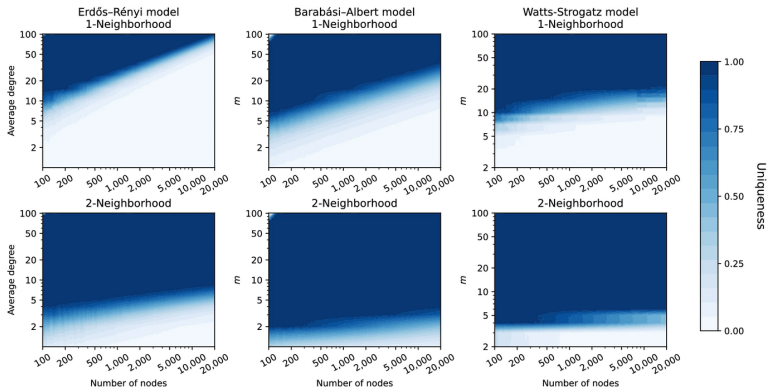


Figure: Unique d -neighborhoods in ER, BA, WS graphs. All models show a transition from all nodes non-unique to unique when density increases.

Takeaways

- Many graph models exist
- Useful to test your algorithm and analysis
- Provide a better understanding of a system and its evolution
- Test statistical significance of a property
- Model is never perfect
- Not all real-world properties are captured

Upcoming lab session and “Homework”

- Get started with Assignment 2.
- Remember you are allowed to work in teams of two!

Credits

- Several images are gratefully reused from Michele Coscia, The Atlas for the Aspiring Network Scientist, v3 of <https://arxiv.org/abs/2101.00863>