An introduction to network theory

Dr. Claudius Gräbner

July 6, 2017

Johannes Kepler University of Linz, Austria

Organizational stuff

- There is a more extensive script on the course HP
- Still preliminary, more extensive than these lectures
- Contains most relevant proofs and additional information
- Lectures focus on intuition and later on practical implementation
- Don't expect too much – network theory big field in its own right
  - Requires programming skill: you're on the right track;)

Motivation I

- Ontological justification:

  > *"[A system is] a complex object, **every part or component of which is connected with other** parts of the same object in such a **manner** that the whole possesses some features that its components lack – that is, **emergent properties**"*

- Networks provide a language to study an essential element of economies
- For more rigorous description and empirical investigation, a formal language is required

Motivation II

There are four different contribution to be made by network theory

1. Precise description of network structures, identification of empirical regularities
2. Ask questions about why relations are as they are
3. Recreate networks to design of models that are realistic in the sense that the interaction of objects is consistent with real interaction structure
4. Reconsider existing models with regard to their implicit assumptions on interaction structure
    - Help to quantify epistemological uncertainty w.r.t. adequacy of assumptions

Outline

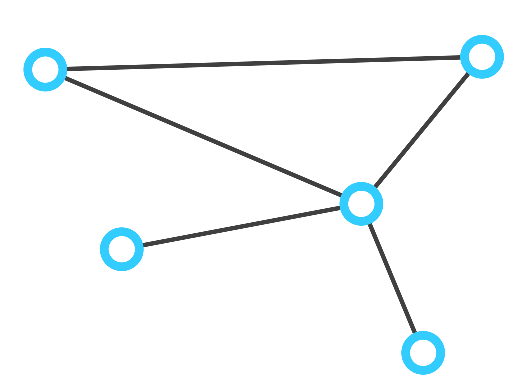
1. Lecture 1
    1.1 Motivation
    1.2 Basic vocabulary
    1.3 Descriptive measures for networks
    1.4 Ideal types of networks
2. Lecture 2
    2.1 Random graphs and Null models
3. Python lab
    3.1 Describing networks in Python
    3.2 Use networks for models once we introduced dynamics and ABM

Basic vocabulary

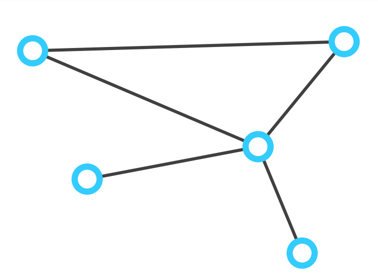
Basic vocabulary

Networks vs. graphs ?

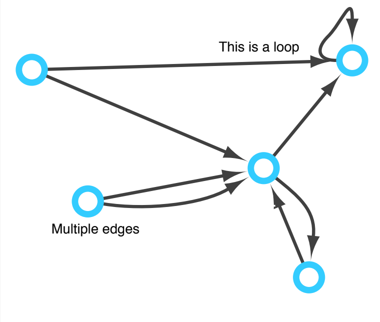
- Graph: $\mathcal{G}(V, E)$
- Vertices: $V(G) = \{v_1, ..., v_n\}$
- Edges: $E(G) = \{e_1, ..., e_n\} \subseteq V \times V$
- $e = \langle v_j, v_k \rangle$ with $v_j, v_k \in V$
- Vertices can be adjacent, edges incident



Types of networks: Simple vs. not simple



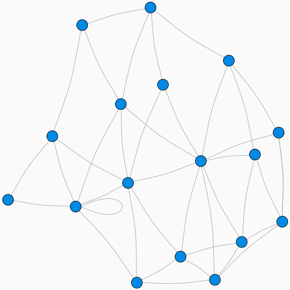
(a) Simple network



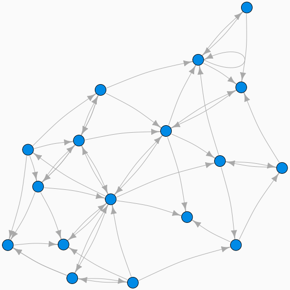
(b) Multiplex network

- Simple network: no loops, not multi-edges

Types of networks: Undirected vs. directed



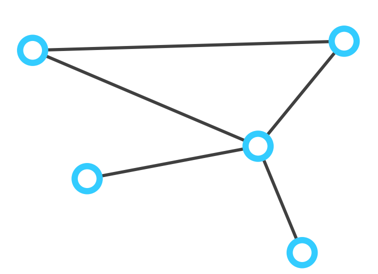
(c) Undirected



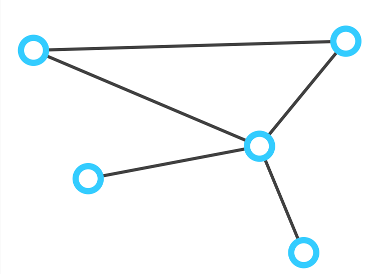
(d) Directed

- Undirected network: edges with $e = \langle v_j, v_k \rangle = e = \langle v_k, v_j \rangle$
- Directed network: arcs with $a_1 = \langle v_j, v_k \rangle \neq \langle v_k, v_j \rangle$

Types of networks: Weighted vs. unweighted



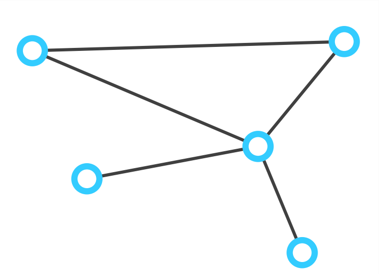
(e) Unweighted graph



(f) Weighted graph

- Every $e = \langle v_j, v_k \rangle$ may be associated with a weight w_{jk}

Final vocabs: Paths, distance, connectedness



(g)

- Walk: $[v_0, e_1, v_1, e_2, .., v_{k-1}, e_k, v_k]$, in which $e_i = \langle v_{i-1}, v_i \rangle$
- Distance is the length of shortest path: $d(v_i, v_j)$
- Connectedness of vertices and graphs

Examples

Examples for ‘networks’

Graphs are *always* only *representations* of real-world objects!!

Type of network	Graph type	Vertices	Edges
Innovation network	Simple	Firms	Joint R&D activity or patenting
Trade network			
Production network			
Friendship			
Employment		Employers, employees	
Production space			

Examples for ‘networks’

Graphs are *always* only *representations* of real-world objects!!

Type of network	Graph type	Vertices	Edges
Innovation network	Simple	Firms	Joint R&D activity or patenting
Trade network	Directed, weighted	Countries	Import/Export flows
Production network	Directed, weighted	Firms/sectors	Factor flows
Friendship	Directed	People	Friendship
Employment	Bipartite	Employers, employees	Work contract
Production space	Bipartite	Products, Countries	Revealed competitive advantage

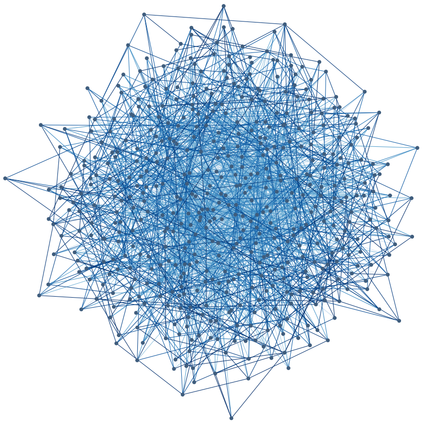
Representation

Why representation?

- Why not just use graphs? Isn't this what economists do?

Why representation?

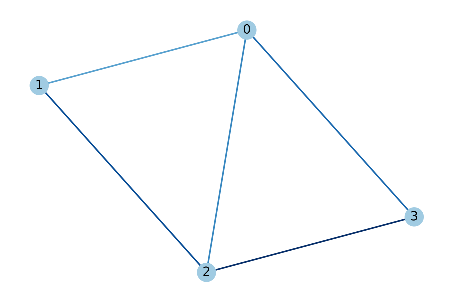
- Why not just use graphs? Isn't this what economics do?
- Mhm....



Types of representation

- Adjacency matrix
- $n \times n$ matrix $\boldsymbol{A}$
  - $a_{ij} = 1 \Leftrightarrow \langle v_i, v_j \rangle \in E(G)$ and zero otherwise

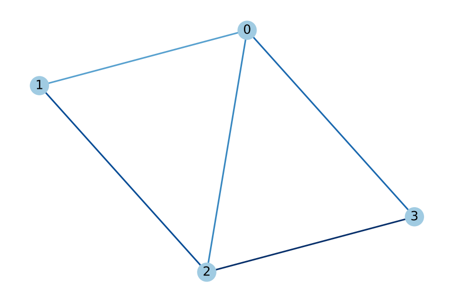
	0	1	2	3
0				
1				
2				
3				



Types of representation

- Adjacency matrix: which vertices are adjacent?
- $n \times n$ matrix $\boldsymbol{A}$
  - $a_{ij} = 1 \Leftrightarrow \langle v_i, v_j \rangle \in E(G)$ and zero otherwise

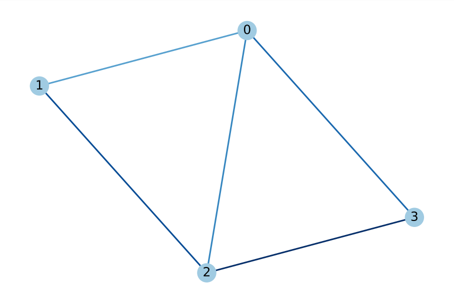
	0	1	2	3
0	0	1	1	1
1	1	0	1	0
2	1	1	0	1
3	1	0	1	0



Types of representation

- Adjacency list
- List of length n, contains tuples of connections

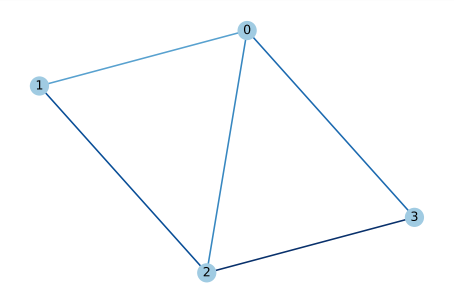
0
1
2
3



Types of representation

- Adjacency list
- List of length *n*, contains tuples of connections

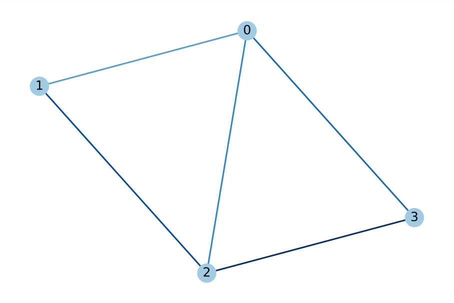
0	1	2	3
1	0	2	
2	0	1	3
3	0	2	



Types of representation

- Incidence matrix
- Which vertices are incident with the vertices?

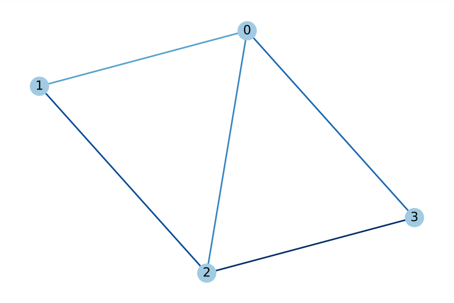
	1	2	3	4	5
0					
1					
2					
3					



Types of representation

- Incidence matrix
- Which vertices are incident with the vertices?

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

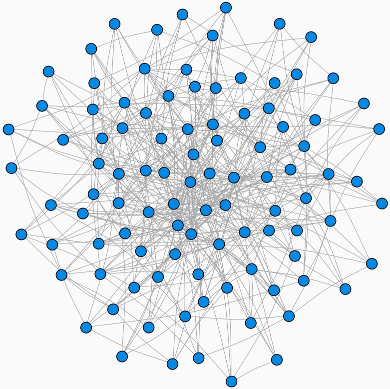


Descriptive measures

Descriptive measures

What would be interesting to know?

- Nb of vertices and edges
- Distances among vertices
- Paths and connectedness
- Density
- Grouping
- Importance of vertices

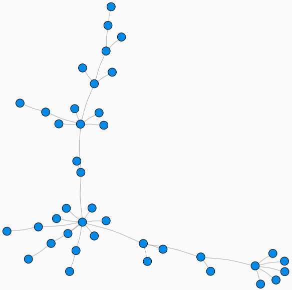


The diameter

The diameter of a graph $\mathcal{G}(V, E)$ is defined as the maximal shortest path within $\mathcal{G}$:

$$diam(G) = max\{d(v_i, v_j)|v_i, v_j \in V(G)\}.$$

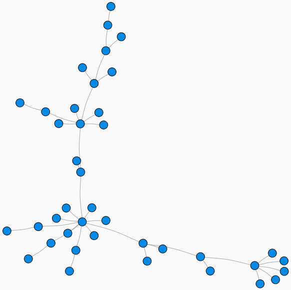
- General measure for size
- Only for connected graphs
- Sensitive to outliers



The average path length (vertex level)

The average length of the shortest paths from $v_i \in V(G)$ to all other vertices in G is given by $\bar{d}(v_i) = \frac{1}{n-1} \sum_{v_j \in V(G), i \neq j} d(v_i, v_j)$.

- Only for connected graphs/vertices
- General measure of size and proximity on the vertex level

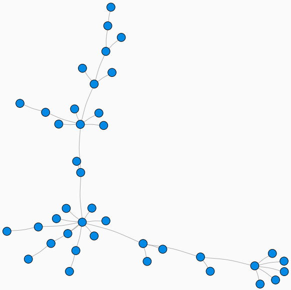


The average path length (graph level)

The average path length of $\mathcal{G}(V, E)$ is then defined as the average of all average shortest paths:

$$\bar{d}(G) = \frac{1}{n} \sum_{v_i \in V(G)} \bar{d}(v_i).$$

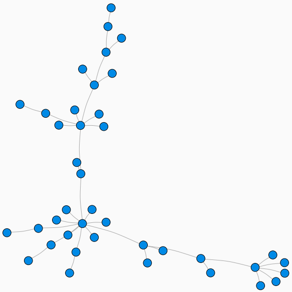
- Only for connected graphs/vertices
- General measure of size and proximity on the system level



The characteristic path length of a Graph

The characteristic path length of a connected graph $\mathcal{G}(V, E)$ is defined as the median of the set $\{\bar{d}(v_i)\}_{i \in N}$.

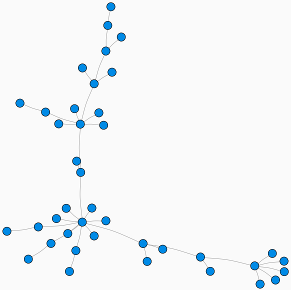
- Only for connected graphs/vertices
- General measure of size and proximity on the system level



The characteristic path length of a Graph

The density of an undirected graph $\mathcal{G}(V, E)$ with n vertices and m edges is given by $\rho(G) = \frac{m}{\binom{n}{2}}$, which corresponds to $\rho(G) = \frac{2m}{n(n-1)}$.

- Checks how 'complete' a graph is

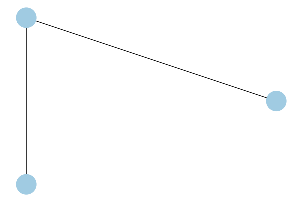


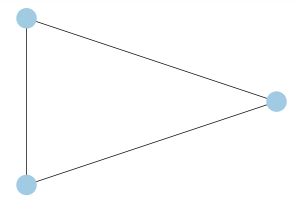
Transitivity

The transitivity of unweighted a graph $\mathcal{G}$ is defined as

$$\tau = \frac{3 \cdot \#\text{triangles}}{\#\text{triples}}.$$

- Checks the share of realized triangles
- Is the friend of your friend also your friend?
- The share of triples that are also triangles.
- Can also be expressed as: $\tau = \frac{\sum_{ijk} a_{ij} a_{jk} a_{kj}}{\sum_{ijk} a_{ij} a_{jk}}$





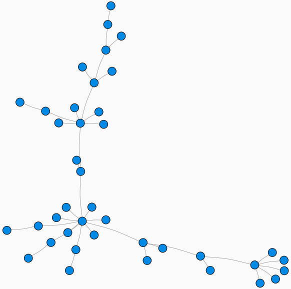
Vertex degree and degree sequence

For an undirected graph $\mathcal{G}$ with adjacency matrix $\boldsymbol{A}$, the degree of vertex v_i is denoted by $\delta(v_i)$ and is defined as:

$$\sum_{j=1}^{n} a_{ij}$$

i.e. the row sum of the adjacency matrix.

- From the vertex degree we can get the mean degree and the degree distribution
- Useful for building intuition

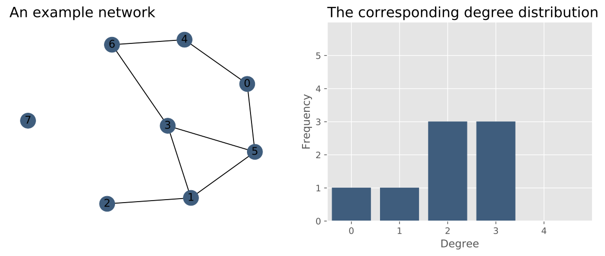


Vertex degree and degree sequence

For an undirected graph $\mathcal{G}$ with adjacency matrix $\boldsymbol{A}$, the degree of vertex v_i is denoted by $\delta(v_i)$ and is defined as:

$$\sum_{j=1}^{n} a_{ij}$$

i.e. the row sum of the adjacency matrix.

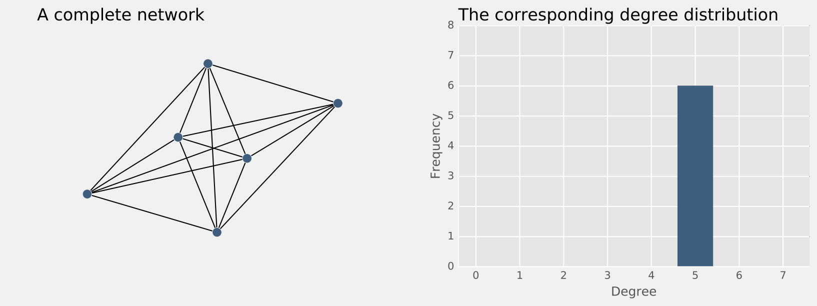


Ideal networks

Ideal network types

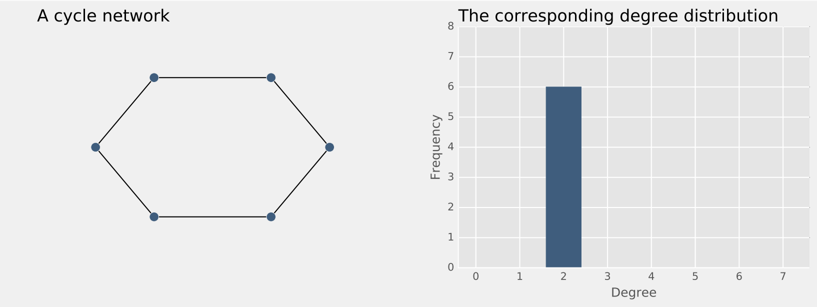
- Ideal types can be important to highlight aspects of real-worlds networks
- Cannot be found in reality as such
- Useful for building intuition

Complete network



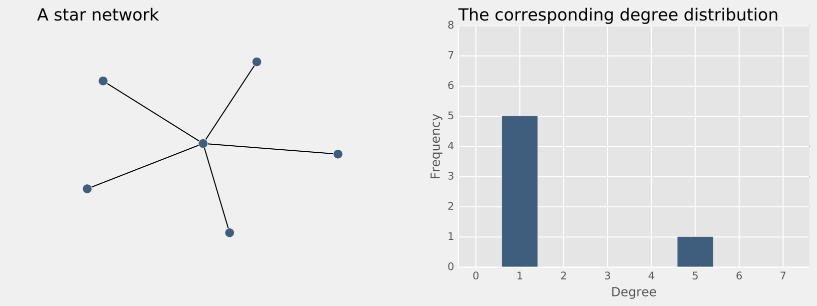
- Assumes that all agents being connected to each other.
- Consequently, all agents have the same degree and $\rho(G) = 1$
- Implicit assumption in many models

A star



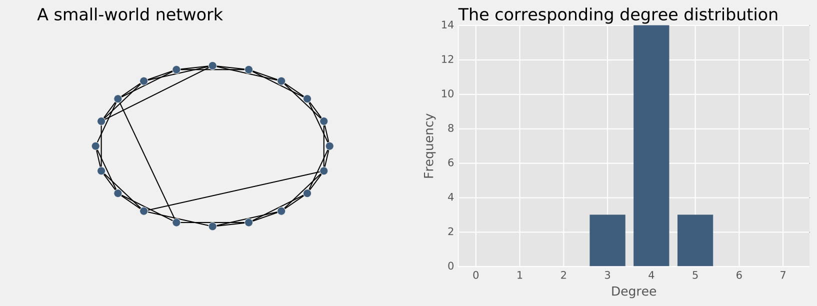
- Captures centralized organization

A complex star



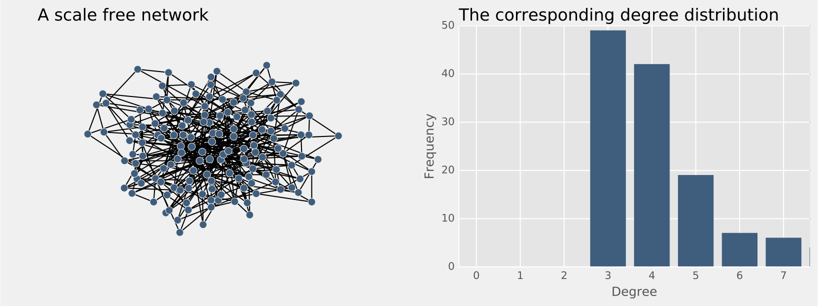
- Captures idea neighbourhood

Small worlds....



- Characterized by (1) a small average path length, and (2) a high degree of clustering
- Relates to Milgram's study on "six degrees of separation".

...and scale free networks



- Characterizes different networks all over the world, such as trade networks, friendships, ...
- Similarity across many disciplines
- Called scale-free networks, because it is one feature of a power law that it is self-similar (i.e. if you zoom in it looks just as before, i.e. the scale does not matter).

Network models

Content

- So far: vocabulary to describe networks
- But not sufficient:
    - If you analyse empirical networks, you need a reference point
    - If you want to consider networks in your model, you need to create one
    - If you want to understand why networks are as they are, you need to build hypotheses about the mechanisms
- To this end, we need models of networks

Content

Goal of analysis	Role of models	Kind of models used
Description of real world systems	Motivation for choice of measures	Previous theoretical models; mental models
Identify patterns	Provide reference points to identify “surprising features”	Purely random graph models, e.g. Erdős-Rényi
Re-create networks	Produce artificial networks, e.g. for ABM	Statistical random graph models, e.g. stochastic block models
Identify mechanisms	Implement theoretical mechanisms to produce artificial data; integrate networks into other models	Mechanismic network generator models, e.g. *preferential attachment*

Content

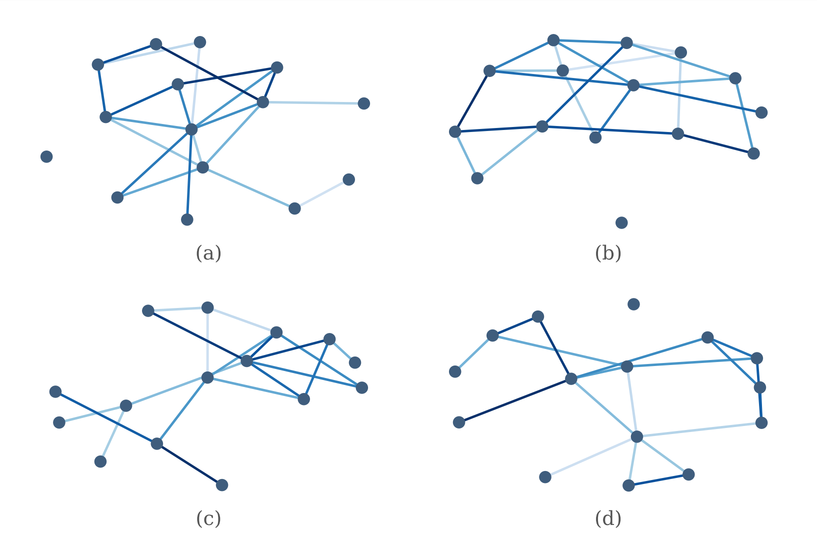
- Here we focus on two different model:
  1. The Erdős-Rényi graph: a random graph.
     - A null model for undirected networks against which to compare
     - Helps you to highlight structures in empirical networks
  2. The Barabási-Albert model
     - Implements the ‘preferential attachment’ mechanism
     - Produces scale-free networks

The Erdős-Rényi Model

The simplest random network: Erdős-Rényi

- Produces an undirected simple graph $G(n, p)$
- Two parameters: the number of vertices, n, and the probability that the edge $\langle v_i, v_j \rangle$ exists
- Probabilistic model: actual graph produced by this process is almost always different.

The simplest random network: Erdős-Rényi



Different instances for $n = 14$ and $p = 0.4$

Distribution of m

- Useful: we can proof that the ER mechanism produces graphs with particular properties
- The probability for an edge to exist is p for all edges (i.e. the edges are *iid*)
- $\binom{n}{2}$ possible edge; probability for an edge *not* to exist is $1 - p$
- $\mathcal{P}(m)$ as the probability to get a graph with exactly m edges:

$$\mathcal{P}(m) = \binom{\binom{n}{2}}{m} p^m \cdot (1-p)^{\binom{n}{2}-m}. \tag{1}$$

Expected degree for a given vertex

Fact (Expected degree of the ER model - vertex level)

The expected degree for any vertex v_i in an ER graph is given by:

$$\mathbb{E}(\delta(v_i) = k) = \sum_{k=1}^{n-1} k\mathbb{P}(\delta = k) = (n-1)p = c \qquad (2)$$

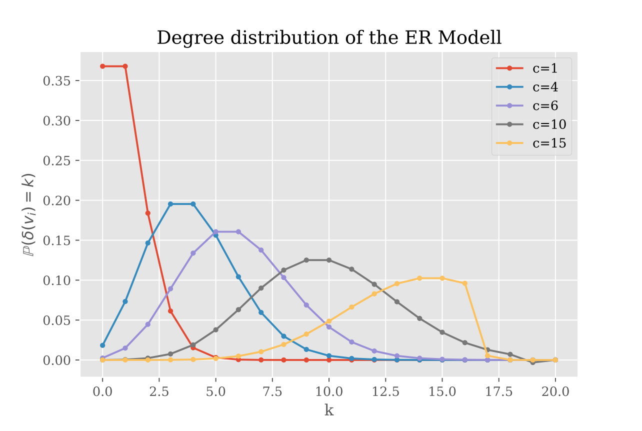
- The expected value for a discrete random variable is defined as a weighted average of all possible values for this variable.
- Expected degree as number of all possible relations (i.e. $n - 1$) times p:
    - number of possible neighbours times the probability that a link to another vertex exists

Degree distribution of ER graphs

Fact (Degree distribution of ER)

For larger ER with small n, the degree distribution is Poisson.

The Poisson distribution is not heavy-tailed:



Expected transitivity of ER graphs

Fact (Transitivity of the ER graph)

For graphs produced by the ER model, the following result regarding their transitivity does hold:

$$\rho^{ER} = \frac{\#triangles}{\#triples} = \frac{\binom{n}{3}p^3}{\binom{n}{3}p^2} = p = \frac{c}{n-1}. \quad (3)$$

- $\binom{n}{3}$ possibilities for triangles in a graph with n vertices.
- For each of this triangle to exist, we need three particular edges to exist;
- expected number of triangles is the number of possible triangles times their probability, i.e. $\binom{n}{3}p^3$.
- There are not many triangles in the ER graph

Lessons from the ER graph

- ER graphs tend to have degrees with a Poisson distribution...
  - ...real-world degree distributions are much more heavy-tailed!
- ER graphs tend to have few triangles....
  - ...social and econ networks tend to have *many more* triangles
- ER graphs tend to have short av path lenghths
  - ...as do real world networks
- ER graphs help us to spot surprising patterns!

The Barabási-Albert Model

History of the Barabási Albert Model

- Real-world networks tend to exhibit heavy-tailed degree distributions
- Herbert Simon has shown that the mechanism of *preferential attachment* can lead to heavy tailed distributions
- Derek de Solla Price showed in 1965 that citation networks have such distributions, and build a graph model based on *preferential attachment*
- Albert-Làszló Barabási and Réka Albert suggested the standard model for undirected networks in 1999 (BA)

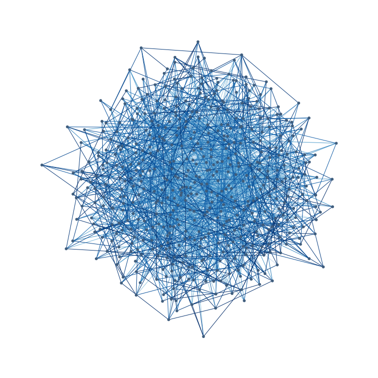
Functioning of the Barabási-Albert Model

- The model has two parameters: two parameters, n and m
- Start with network of n_0 vertices without edges
- Add vertices of degree $m < n_0$ subsequently until the network has n vertices
- Edges of the new vertices are wired to existing vertices probabilisitically according to:

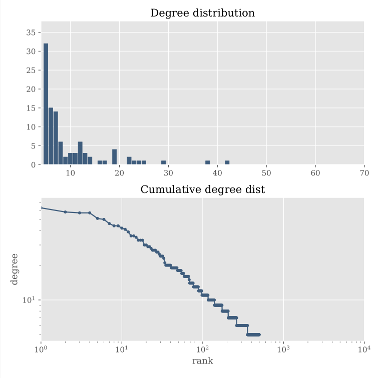
$$p^i(\delta(v_i)) = \frac{\delta(v_i)}{\sum_{v_j \in V(G)} \delta(v_j)} \tag{4}$$

- "Rich" vertices tend to grow "richer"

Functioning of the Barabási-Albert Model



(h) A BA network



(i) Degree distribution

Fact (Degree distribution of the BA model)

For large k, the degree distribution of the BA models follows a power law:

$$\mathbb{E}(\delta(v_i) = k) \sim k^{-3} \qquad (5)$$

Comparison of degree distributions

- In contrast to the heavy-tailed power law distribution, the Poisson distribution does not place probability mass on very high vertex degrees

Final thoughts on Barabási-Albert

- BA model suggests a certain mechanism accounting for power-law distributed degrees
- While *compatible* with the observation of power laws, no final evidence that it actually *is* this mechanism.
    - real world mechanisms are “concealed” and “have to be conjectured”
- Other forms of validation necessary: input, output validation
- But also deeper understanding of the implications: age of a vertex correlated with its degree
- Do not put too much faith in simple models such as the one by BA: they are only representations of reality, and they should be treated as such.
- It is important to seek (1) mechanismic explanation and (2) preserve an epistemological fallibilism.

Outlook

Goal of analysis	Role of models	Kind of models used
Description of real world systems	Motivation for choice of measures	Previous theoretical models; mental models
Identify patterns	Provide reference points to identify “surprising features”	Purely random graph models, e.g. Erdős-Rényi
Re-create networks	Produce artificial networks, e.g. for ABM	Statistical random graph models, e.g. stochastic block models
Identify mechanisms	Implement theoretical mechanisms to produce artificial data; integrate networks into other models	Mechanismic network generator models, e.g. *preferential attachment*

Keep in mind this is only a *rough* classification that may help in understanding the motivation for modelling!

How to implement graph models in NetworkX

Some Python code

- NetworkX allows you to generate graphs with a vast number of different models
- There is not really new syntax to learn
- To create an ER graph with $n = 14$ and $p = 0.2$, just type:

```
nx.fast_gnp_random_graph(14, 0.2)
```

- To create a BA graph with $n = 200$ and $m = 50$, just type:

```
nx.barabasi_albert_graph(200, 50,)
```

- Some notes on how to implement models in NetworkX, and on how to read in empirical data, will be distributed shortly