A short introduction to network theory

Claudius Gräbner[1,2,3]

[1]Institute for Comprehensive Analysis of the Economy, Johannes Kepler University
[2]Institute for Socioeconomics, University of Duisburg-Essen
[3]ZOE. Institute for future-fit economies

Seminar 'Agentenbasierte Modellierung in der Ökonomik' an der Cusanus Hochschule im WS 2019/20

Contents

Abstract

This script is preliminary and some parts are missing. Your feedback to Claudius is, of course, highly appreciated.

1 Motivation

This necessarily short introduction will introduce you to the basic vocabulary necessary to study networks and to get familiar with more advanced concepts on your own.

Given that we study complex systems by building representations of them, we need a proper language to describe the relations among its parts.

A verbal description of these relations is an important first step. But once we want to understand these systems in more detail, we need a more precise language. The language of choice to describe relations among components exactly is provided by *graph theory*, a branch of discrete mathematics. You can read a bit about the historical roots of graph theory in appendix A.

Before we move to the outline, it may be useful to highlight again the different purposes network theory can serve during the study of complex economies.

First, aside from describing structures that can be represented as networks[1] rigorously, network theory allows us to investigate these structures empirically and quantitatively. This helps us to identify striking regularities that would otherwise remain hidden.

Second, it helps us to design models that embed mechanisms and that help us to understand why the networks in reality are as they are.

Third, being able to recreate artificial networks that share properties of real world systems helps us to make more realistic models. For example, if we want to build an agent-based model to study the potential for industrial and trade policy, then we can integrate the core-periphery structure of real-world trade networks explicitly into our model. This is important because this particular core-periphery structure might have important implications on the conclusions of our model.

Finally, considering networks formally also allows us to reconsider the results of previous models in relation to the implicit assumed network structure. For example, the standard GET model assumes a well-mixed population, or, in other words, a complete network in which every agent may trade with every other agent. Albin and Foley (1992) built a simulation model which resembles almost all the standard assumptions of the Arrow-Debreu economy, but features a cycle network, which captures commonly found neighborhood structures. They found that as a consequence, the distributional implications of the equilibrium in an otherwise complete market are different. This relates to a deeper philosophical question about the source of knowledge and uncertainty in formal models. Cartwright (2007), for example, argues that whether the auxiliary assumptions of a model hold, is usually uncertain, and no matter how rigorous the derivation of the results, this uncertainty remains. Network theory helps us to highlight this uncertainty with regard to interaction structure. The example of the implicitly assumed network structure of GET is a nice example, and it represents a powerful critique of the monist conception of many neoclassical economists, who argue that rigor in the neoclassical sense is an essential quality criterion for economic models.

This script is organized as follows: first, section (2) introduces the formal apparatus to describe networks, i.e. the basic vocabulary and the most common network statistics. It also discusses some ideal types of networks, which can hardly be found in reality, but which help us highlighting some structural features and their implications

Section 5 then introduces models of networks. It starts with a discussion of the purpose of network models. Then it introduces the simple random graph model, which serves as an important reference point for

[1] I chose this ugly formulation to highlight again the difference between a model and the real world entity it represents. There are not networks in reality, only objects and systems than can be represented as networks. Whether this representation makes sense, depends on the adequacy of the assumptions made.

empirical investigations of networks. Then we discuss scale-free networks, which emerge as a result of a "the-rich-get-richer"-like mechanism called *preferential attachment*. The final section (6) gives an outlook and ideas for further studies. In the appendix you find an exposition of the historical roots of network theory, an example graph and a table with all descriptive measures discussed in section 2 and their corresponding Python commends, and finally some recommendations for graph visualization.

Again: this is a very preliminary script, which is not yet ready for distribution. But your feedback to Claudius is highly appreciated.

2 Foundations and network description

In this section, we first introduce basic vocabulary, then discuss how networks can be stored, then discuss common statistics to describe networks and finally look at some ideal types of networks.

2.1 Basic vocabulary

First, note that in these lectures we do not distinguish between a *network* and a *graph*. Usually, the term 'network' is used to refer to the phenomenon in reality, and 'graph' refers to the mathematical object with which we represent a certain network. But this distinction is not universally accepted and you should always make sure what authors mean by the two concepts.

A simple graph consists of two kind of objects: vertices (or nodes) and edges (or links).

The set of vertices – which can represent diverse stuff such as agents, computers, countries, or firms – will be denoted with V and the set of edges will be denoted by E. Edges may represent friendship between agents, an internet connection between computers, trade flows between countries, and joint innovation activities between firms. Let's state this more precisely:

Definition 1 *A graph $\mathcal{G}(V, E)$ is a tuple where $V(G) = \{v_1, ..., v_n\}$ denotes a set of vertices and $E(G) = \{e_1, ..., e_n\} \subseteq V \times V$ denotes a set of edges. For every edge $e_i \in E$ we have $e = \langle v_j, v_k \rangle$ with $v_j, v_k \in V$.*

Two vertices that are joined by an edge are called *adjacent*, the edge joining two vertices is called *incident* with the two vertices.

It is useful to distinguish some different forms of networks right from the start:

First, we speak of a *simple* graph if (1) there are no 'loops' in the graph, and (2) there are no multiple edges between two vertices.

Definition 2 *A graph $\mathcal{G}(V, E)$ is called simple if*
(i) $j \neq k \forall \langle v_j, v_k \rangle \in E(G)$ (no loops) and
(ii) $e_i = \langle v_j, v_k \rangle \wedge e_j = \langle v_j, v_k \rangle \Leftrightarrow e_i = e_j$ (no multiple edges).

Networks for which multiple edges between the same nodes are allowed are called *multigraphs*. They frequently come up when you consider logistic networks, where edges represent infrastructural connections between certain locations). See figure 2 for the distinction: in contrast to the simple graph on the left, the multi-graph on the right has both a loop and a multi-edge.

Second, we can distinguish *directed* and *undirected* networks. In an undirected network, edges are bi-directional, i.e. $e = \langle v_j, v_k \rangle = e = \langle v_k, v_j \rangle$. This means that whenever v_i is connected with v_j, v_j is also connected to v_i. Networks that represent joint innovation activities of firms are usually undirected because the edges indicate whether two firms engage in a joint R&D activity. On the other hand, production networks are usually directed networks: here an edge represents the flow of outputs from firm j to firm i, so the direction of the flow is important. In such cases edges are usually called *arcs* and $a_1 = \langle v_j, v_k \rangle \neq \langle v_k, v_j \rangle$. See figure 2 for an illustration: the graph on the left is undirected and has edges, the graph on the right is directed and has arcs.

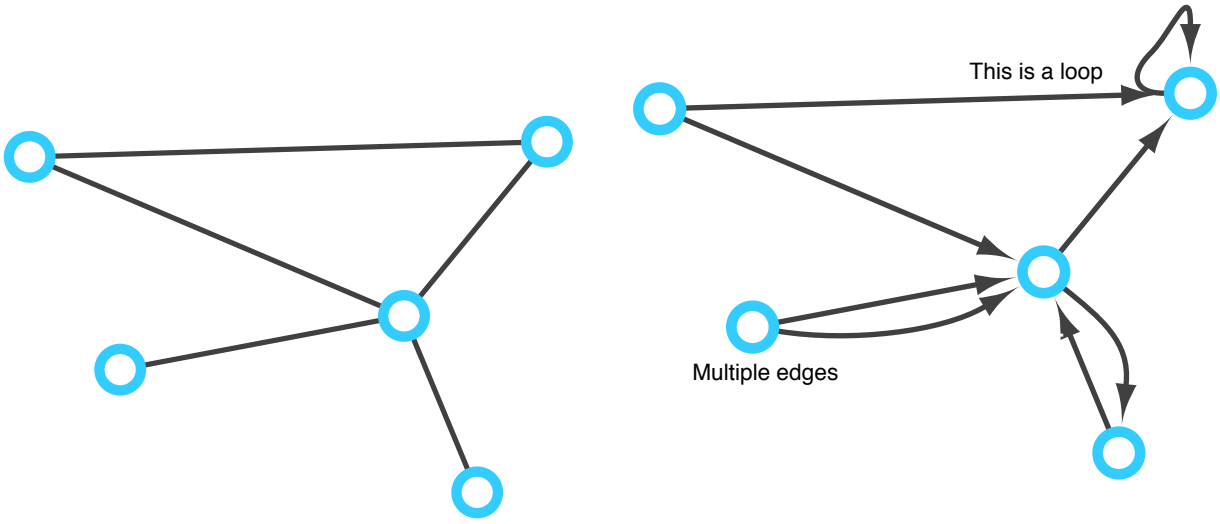


Figure 1: A simple (left) and a not-so-simple graph (right).

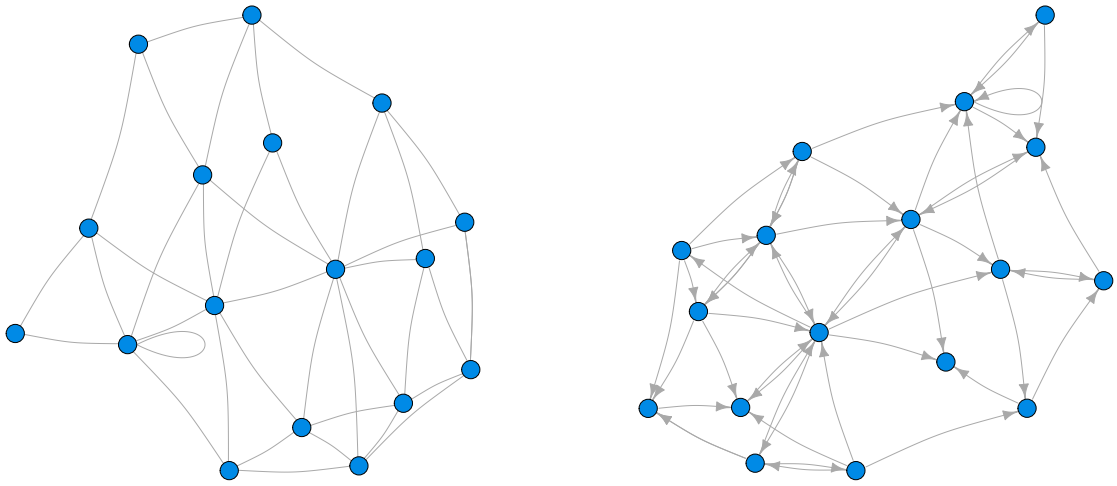


Figure 2: A undirected (left) and a directed graph (right).

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

Table 1: Examples for economic networks. The last example is an example for a *bipartite network*, where we have two types of vertices. More on these networks follows below.

Thirdly, we distinguish between *weighted* and *unweighted* networks. In unweighted networks edges can have the 'value' zero or one: either an edge between v_i and v_j exists, or it does not exist. In a friendship network, for example, where two vertices are connected by an edge if the persons represented by the vertices are friends, there is no weight for their relationship. Either they are friends or not. Trade networks, on the other hand, are usually weighted networks: here the edge $e = \langle v_j, v_k \rangle$ is associated with a weight w_{jk} (or a weight function) that provides information about the amount of exports from country i to country j.

Finally, we want to distinguish between k-partite networks. In a k-partite network there exist k different types of vertices, which are not connected among each other. This means that there are only edges between classes, but never within classes. The most common type of k-partite networks is the *bipartite* network, where $k = 2$. For example, in a firm-ownership network we have two types of nodes: legal persons and firms, and there is an edge between a firm and a person if the person holds shares of the firm. But since people cannot buy shares from other persons, and firms cannot buy shares from other firms (an assumption that may be subtle) we speak of a bipartite network. Just keep in mind that it is always possible to map a k-partite network into k simple networks via a one-mode projection, where two vertices in the projected networks are connected if they are connected to the same other vertices in the original graph.

Table 1 gives some examples for economic networks of different types, with the vertices and edges being well defined. There are, of course, other types of networks, such as multiplex networks, or dynamic networks, but we will not consider them in this course. When you do a network analysis, it is **absolutely essential** to be *very clear* about what is represented by the vertices and what is represented by the edges. Otherwise, you (a) will waste a lot of your time, (b) won't be able to interpret all the network measures that follow, (c) confuse your audience and screw your reputation, and other bad things will certainly happen as well.

When we introduce the basic vocabulary below we mostly stick to *simple graphs*. But this is not really a limitation because most of the measures can be readily generalized to more complex types of networks.

When we introduce the subsequent statistics and properties of graphs you will notice that some of them have a local scope in the sense that they describe individual vertices or groups of vertices, and other with a system-wide scope in the sense that they describe the network as whole. We will get back to this point at the end of the intro lecture.

The first measure we consider is the *neighborhood* of a vertex as the set of all those vertices in the same graph, which are adjacent to the vertex:

Definition 3 *In $\mathcal{G}(V, E)$, the neighborhood set of vertex i, $\mathcal{N}(v_i)$, is given by*

$$\mathcal{N}(v_i) := \{v_j \in V(G) | i \neq j, \exists e \in E(G) : e = \langle v_i, v_j \rangle\}.$$

The neighborhood refers to those vertices that are 'close' to a vertex in the sense that they are directly connected to each other via an edge. But it is also useful to talk about the relation between vertices that are not neighbors, but that are nevertheless within each others reach:

Definition 4 (Walks, paths, trails) *In a graph $\mathcal{G}(V, E)$ a sequence of edges and nodes $[v_0, e_1, v_1, e_2, .., v_{k-1}, e_k, v_k]$, in which $e_i = \langle v_{i-1}, v_i \rangle$, is called a walk from v_0 to v_k. The following special cases exist:*

If $v_0 = v_k$, (v_0, v_k) is called a closed walk.

If no edge appears twice, (v_0, v_k) is called a trail.

If neither an edge nor a vertex appears twice, (v_0, v_k) is called a path.

If no edge and no vertex appears twice, except v_0 and v_k, (v_0, v_k) is called a cycle.

It is thus clear that there might be vertices in a graph, which cannot be reached from each other. There can also be vertices without any edges (sometimes called *dangling node*). We call two vertices that can be reached by a walk *connected*. We may also introduce a property of the graph as a whole by considering the set of all vertices and call a *graph connected* if all of its vertices are connected.

Definition 5 (Connectedness) *In a graph $\mathcal{G}(V, E)$, two vertices $v_i, v_j \in V(G), i \neq j$ are connected if the path (v_i, v_j) exists in $\mathcal{G}$. If all vertices in $\mathcal{G}$ are connected, $\mathcal{G}$ is called a connected graph.*

Now that we have introduced the notion of connectedness, a logical next step is to think about a notion of *distance*: while vertices might be connected, it makes sense to differentiate between vertices that are close in the sense that they are in the same neighborhood, and vertices that are located in very different areas of the whole graph. We arrive at the simplest notion of distance by simply counting the edges that are necessary to get from one vertex to the other, and call the resulting measure *geodestic distance*.

Definition 6 (Geodestic distance) *In a graph $\mathcal{G}(V, E)$, the geodestic distance between two vertices $v_i, v_j \in V(G), i \neq j$ is the length of the shortest path (v_i, v_j) and is denoted as $d(v_i, v_j)$.*

If two vertices are not connected it is common to write $d(v_i, v_j) = \infty$, i.e. to assign an infite distance to them. This must be kept in mind when you use other statistics that built upon the distance and that may not be meaningful if two vertices are not connected.

We have so far clarified the most important vocabulary to talk about graphs. But consider you want to describe a graph to someone else, or store a graph on your computer, how do you actually represent the graph?

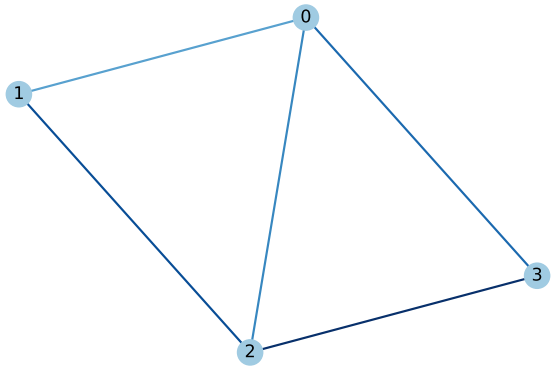
2.2 Representations of graphs

Her we consider the simple questions of how we can store information about the graphs we have considered so far. Of course, you could draw the graph and pass it on to someone else. But aside from the fact that representing graphs graphically is (1) not trivial and (2) inconvenient for very large graphs, it would require a lot of work to extract all the relevant information from a picture!

In contrast, the following methods are particularly common:

An **adjacency matrix** of a graph $\mathcal{G}(V, E)$ with n vertices is a $n \times n$ matrix $\boldsymbol{A}$. In case of an unweighted graph we have $a_{ij} = 1 \Leftrightarrow \langle v_i, v_j \rangle \in E(G)$ and zero otherwise. If we are dealing with a weighted graph and every edge has a weight ω, the elements of the adjacency matrix contain these weights, i.e. $a_{ij} = \begin{cases} \omega_{ij}, & \text{iff } \langle v_i, v_j \rangle \in E(G) \\ 0, & \text{otherwise} \end{cases}$ Note that if our graph does not have any self-loops, the main diagonal of $\boldsymbol{A}$ is necessarily zero, and if the graph is undirected $a_{ij} = a_{ji}$ and we could skip the upper-triangular part of the matrix because the matrix is symmetric.

If we condense the adjacency matrix, we come up with an **adjacency list**, i.e. a list of tuples where the n^{nt} element contains a tuple with the neighbors of v_n. In case of a directed graph, we see here the outgoing connections of every graph. In practice, you will encounter adjacency lists much more frequently than adjacency matrices because they require less memory. See figure 3 for an illustration.



(a) The graph.

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

(b) The adjacency matrix.

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

(c) The adjacency list.

	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

(d) The incidence matrix

v_i	v_j
0	1
0	2
0	3
1	2
2	3

(e) The edge list.

Figure 3: A graph and its different representations.

The **incidence matrix** of a graph $\mathcal{G}(V, E)$ with n vertices and m is a $n \times m$ matrix $\boldsymbol{M}$ in which $m_{ij} = 1 \Leftrightarrow \langle v_i, v_j \rangle \in E(G)$ and zero otherwise, i.e. m_{ij} indicates whether an edge is incident with a vertex. Similarly, if we are dealing with a weighted graph, we will use the sum of the weights instead.

Finally, an **edge list** is simply list of $E(G)$. Under which circumstances is an edge list sufficient to pin down a growth unambiguously? Correct, if there is no isolated vertex, i.e. a vertex without any neighbors. From this it follows that every graph considered so far can be unambiguously represented by its edge and vertex list. Tables 3d and 3e represent the incidence matrix and the edge list for the graph in figure 3a.

3 Describing graphs

In the following we will introduce a number of useful statistics to *describe* networks. This will be an important and indispensable step in *understanding* them.

3.1 Measures of size and density

We already encountered some 'simple' statistics: the number of vertices n, the number of edges m and the connectedness of a graph. They somehow measure the 'size' of a graph, but it is not immediately obvious what we mean by 'size' in the context of networks. In fact, this is one advantage of the precise vocabulary of graph theory: that it allows us to be more precise when talking about networks than mere words allow us to.

Let's consider some other measures that capture different aspects of the 'size' of a network.

We already defined the neighborhood of a vertex and introduced the idea of the geodestic distance between two vertices. These are measures on the level of individual vertices, but it is easy to extend it to a system-wide measure of size for the graph as a whole:

Definition 7 (Diameter) *The diameter of a graph $\mathcal{G}(V, E)$ is defined as the maximal shortest path within $\mathcal{G}$:*

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

Considering extreme measures, such as the diameter, can be confusing if the path lengths in the graph are very different, or if there is one large outlier. Thus, looking at the *average path length* or the *characteristic* (i.e. median) *path length* can be more illustrative. However, since $d(v_i, v_j)$ is infinite if they are not connected, the definitions of average and characteristic path length make sense only in the context of connected graphs.

Definition 8 (Average Path Length) *In a connected graph $\mathcal{G}(V, E)$, the average length of the shortest paths from $v_i \in V(G)$ to all other vertices in G is given by $\bar{d}(v_i) = \frac{1}{n-1}\sum_{v_j \in V(G), i \neq j} d(v_i, v_j)$. The average path length of $\mathcal{G}(V, E)$ is then defined as the average of all average shortest paths:*

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

Definition 9 (Characeristic Path Length) *The characteristic path length of a connected graph $\mathcal{G}(V, E)$ is defined as the median of the set $\{\bar{d}(v_i)\}_{i \in N}$.*

Another very basic property of a graph refers to the relation between the number of vertices and the number of edges: it would make much sense to consider some graphs as being 'tight' in the sense that there are a lot of edges compared to the vertices, and others as 'sparse' if there are only few edges and many vertices are unconnected. The notion of *density* formalizes this distinction:

Definition 10 (Density) *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 $\rho(G) = \frac{2m}{n(n-1)}$. For directed graphs we have $\rho(G) = \frac{a}{n(n-1)}$.*

Let's unpack this definition for the undirected case a bit: m is the actual number of edges in the graph. $\binom{n}{2}$ is the binomial coefficient, which gives us the number of possibilities to combine two distinct nodes. Thus, $\binom{n}{2}$ gives the maximum number of edges in the graph, provided we talk about simple graphs (i.e. graphs without loops and multiple edges), which is why this definition of density makes sense for simple graphs only. Because of this construction, $\rho(G)$ could also be seen as a measure for the 'completeness' of the graph in terms of relations: $\rho(G) = 1$ means that every possible connection in the graph actually exists. Such graphs are called *complete graphs*.

Excursus: The binomial coefficient

You may have forgotten about what the binomial coefficient is about, so let's recap, starting from the basics:

Remember that a *polynomial* is a mathematical expression containing constants and variables, which are related to each other via the operations of *addition*, *multiplication* and *exponentiation*. For the exponentiation, only integers greater or equal to zero are allowed. Thus, division and taking roots cannot be part of a polynomial. If we consider a polynomial with only one variable, x, it is called an *univariate* polynomial. Any univariate polynomial can be written as

$$\sum_{k=0}^{n} a_k x^k = a_n x^n + \cdots + a_2 x^2 + a_1 x + a_0,$$

where a_i are the constants forming the *coefficients* of the different terms of the polynomial. We say that a polynomial is of degree d where d is the highest degree of the term with the highest degree in the polynomial. The degree of a term is the exponent of the variable in this term (for *multivariate* polynomials it is the sum of their exponents).

Now, a *binomial* is a polynomial with two terms. An univariate binomial can generally be written as $a_1x^m + a_2x^n$, but for our application it is more illustrative to look a bivariate binomials:

$$a_1x^my^m + a_2x^ny^n$$

How does this relate to combinatorics? Consider the case with two variable, x and y. Let's write $(x+y)^n$ as $(x+y)(x+y)(x+y)\cdots(x+y)$ or, more specifically,

$$(x+y)^3 = (x+y)(x+y)(x+y).$$

If we expand this expression we get:

$$xxx + xxy + xyx + yxx + yyx + yxyxyy + yyy$$

The first and the last expression correspond to x^3 and y^3 respectively. The six terms in the middle can be summarized as $3x^2y + 3xy^2$. The coefficient '3' indicates that there are three ways to choose two xs or ys from the original set, which consisted of $n = 3$.

Thus, if we use the numbers 1,2 and 3 to denote the position of the xs in the original sets, there are three different ways to choose different two-element sets from the set $\{1, 2, 3\}$, namely: $\{1, 2\}$, $\{2, 3\}$, $\{1, 3\}$ (and there is –obviously– only one possibility to choose a three element set).

More general, the coefficient of $x^{n-2}y^2$ equals the binomial coefficient $\binom{n}{2}$. This is why we pronounce $\binom{n}{2}$ as "n choose k": there are $\binom{n}{2}$ ways to choose an unordered set with k elements from a set with n elements.

You can read more about this in every elementary algebra book.

When you study graphs, particularly later once you deal with models of networks, probability theory and combinatorics will become important. It is therefore always a good idea to review from time to time the foundations of these two fields.

3.2 Measures for clustering

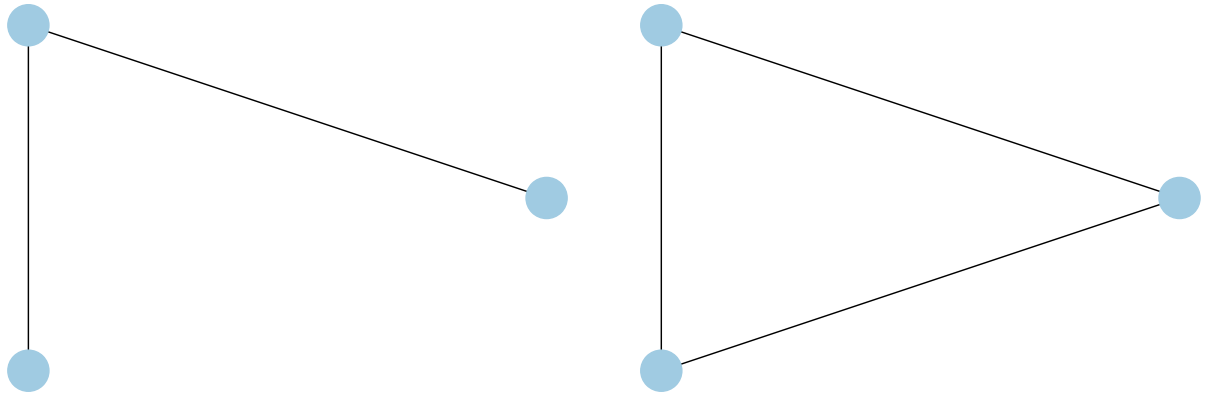
When we consider social networks, we often find that if two vertices v_i, v_j are connected to a third vertex, v_k, then the probability that v_i and v_j are also connected is quite high. This is intuitive once we consider friendships: if Claire is friend with Juni, and Juni is a friend of Tom, then its likely that Tom and Claire know each other; and since similar people tend to be friends with each other, it is also likely that Tim and Claire are friends.

To measure this, the local *clustering coefficient* $cc(v_i)$ of a node v_i is used. It shows how the neighbors of v_i are connected with each other. Formally, it is the ratio between the actual and potential number of edges in $\mathcal{N}(v_i)$.

Definition 11 (Local clustering coefficient) *The local clustering coefficient of a vertex v_i in an unweighted and undirected network $\mathcal{G}$ is defined as*

$$cc(v_i) = \frac{\sum_{ijk} a_{ij}a_{jk}a_{kj}}{\delta(v_i)(\delta(v_i) - 1)}.$$

Note that the enumerator in the definition measures number of pairs of neighbors of v_i that are connected, i.e. the nb of triples that – through the 'help' of v_i – become triangles. See figure 4 for the distinction between *triples* and *triangles*.



(a) A triple. (b) A triangle.

Figure 4: Triples and triangles.

To measure clustering on the aggregated level, we can use either the average clustering coefficient or a measure called *network transitivity*.[2]

Definition 12 (Global clustering coefficient) *The local clustering coefficient of a vertex v_i in an unweighted and undirected network $\mathcal{G}$ is defined as*

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

Network transitivity measures a very similar thing (for the subtle differences see Newman (2003, p. 184ff)): the clustering coefficient on the individual level measures, roughly speaking, the probability that the neighbors of a vertex are also neighbors. Network transivitity measures the ubiquity of triangles in a network, i.e. the share of triples, which are also triangles.

It is a striking phenomenon of *social* networks (be it economic or not) that they have high transitivity values when compared to natural and artificial networks. This indicates that there are some very specific mechanisms operating in social networks that deserve to be investigated. Thus, describing graphs using quantitative measures directly suggests new avenues for theorizing (and this theorizing may also include purely qualitative investigations). But the description can only be a first step because to understand why social networks are as they are, one needs *meachnismic* theories and models about the emergence of network structure, an important research area of complexity economists.

Definition 13 (Transitivity) *The transitivity of unweighted a graph $\mathcal{G}$ is defined as*

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

Why do we need the factor 3 in the enumerator? It is not strictly necessary, but it ensures that $\tau \in (0, 1)$ because every single triangle corresponds to three triples. Thus, τ can be interpreted as the share of triples that are also triangles.

Note that τ can also be expressed only in terms of the adjacency matrix $\boldsymbol{A}$ of the graph: $\tau = \frac{\sum_{ijk} a_{ij}a_{jk}a_{kj}}{\sum_{ijk} a_{ij}a_{jk}}$. This works because the sum in the enumerator is 1 if there is a triangle and zero otherwise, the sum in the denominator is 1 if there are two connections among the three vertices. This expression is sometimes considered more exact and more straightforward than the reference to the triangles and triads.

[2] There are different definitions, some referring to the individual and some on the aggregate level, some referring to the term 'clustering coefficient', others to transitivity.

3.3 Measures of centrality

3.3.1 Degrees and strength distribution

One of the most important statistics for graphs refer to the *degree* of its vertices, which is given by the number of edges that are incident to these vertices.

Definition 14 (Vertex degree, undirected) *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.

If we allowed for loops, such edges would be counted twice, but since we confine ourselves to simple graphs, this does not need to bother us for now. If we consider directed graphs, we need to distinguish between the *in* and *out* degree:

Definition 15 (Vertex degree, directed) *For a directed graph $\mathcal{G}$ with adjacency matrix $\boldsymbol{A}$, the in-degree of vertex v_i is denoted by $\delta^{in}(v_i)$ and is defined as:*

$$\sum_{j}^{n} a_{ji}$$

i.e. the column sum of the adjacency matrix. the out-degree is denoted by $\delta^{out}(v_i)$ and is defined as:

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

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

In the case of weighted graphs, the degree of a vertex is sometimes also called its 'strength'.

If m is the number of edges in our graph, it follows directly that $\sum_{v_i \in V(G)} \delta(v_i) = \frac{1}{2}m$.

An important descriptive statistic for a graph as a whole is the corresponding *degree sequence*, i.e. a list containing all degrees of the vertices in the graph.

From this list, it is straightforward to calculate the *mean degree* of a network, which contains both valuable information in itself, but also if compared with the degrees of the vertices: we can then say that a vertex has a particularly high or low degree compared to the rest of the network.

Excursus: The usefulness of mathematical modelling of networks

Here is a question that shows us the power of a precise language to talk about networks: if I gave you the following degree sequence, would you be able to construct a graph from it: $[1, 2, 3, 4, 5]$.

To answer the question, we could obviously try to construct such a graph. The problem with this approach is – aside from being time consuming – that we probably would never be sure whether it is actually impossible to construct such a graph, whether we forgot a possibility, or whether we are simply to stupid too succeed. Luckily, we can avoid such a disgraceful situation by using graph theory.

Here is a quite simple but not immediately obvious fact about networks:

Fact 1 *In all graphs the number of vertices with odd degree is necessarily even.*

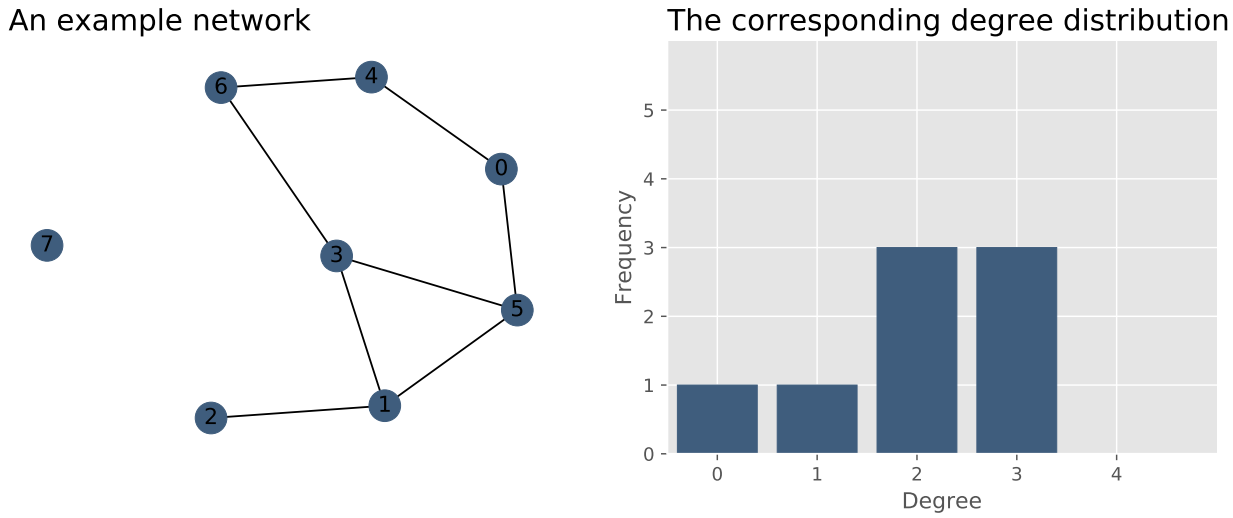


Figure 5: The degree distribution of a graph.

Proof 1 *First thing to note is that for every graph with n vertices and m edges, we necessarily have $\sum_i^n \delta(v_i) = 2m$, i.e. the sum of the vertex degrees is twice the number of edges (because every edge is a binary relation between vertices). Let us now divide $V(G)$ into two groups, those vertices with odd degree, $V_{odd}(G)$ and those vertices with even degree, $V_{even}(G)$. It is clear that $\sum_i^n \delta(v_i) = \sum_{v_i \in V_{odd}(G)} \delta(v_i) + \sum_{v_i \in V_{even}(G)} \delta(v_i)$. Since $2m$ is necessarily even (because an even number is a number that is divisible by two without remainder), $\sum_i^n \delta(v_i)$ must also be even. Because the sum of even numbers is also even, we know that $\sum_{v_i \in V_{even}(G)} \delta(v_i)$ is certainly even. And because the sum of an odd and an even number is uneven, and $\sum_i^n \delta(v_i)$ is even, $\sum_{v_i \in V_{odd}(G)} \delta(v_i)$ must be even as well. The only possibility for a sum of odd numbers to be even is that the sum consists of an even number of addends. This proofs the abovementioned fact.* □

The proof shows that it is a *logical necessity* that for all graphs the number of vertices with odd degree is necessarily even. Now, what about the graph given by the degree sequence $[1, 2, 3, 4, 5]$? Such a graph would have five vertices, of which three vertices had odd degrees. If follows from the preceding fact that such a graph cannot exist. Hence, the list $[1, 2, 3, 4, 5]$ is not the degree sequence of a graph. We say this list is not *graphic*.

From a degree sequence we can easily construct a *degree distribution* of a graph. A degree distribution counts the number of vertices with a given degree. A degree distribution contains a lot of information about a graph because it shows us how evenly the relations are distributed between the components of our network. Since in many instances the degree of a vertex has some meaningful interpretation (such as power, wealth, vulnerability,...), degree distributions are an important description of graphs. We will also see that there are some striking empirical regularities about degree distributions. We can visualize the degree sequence of a graph via a histogram, as indicated in figure 5. Another way to think of the degree distribution is this: it gives you the probability that a randomly chosen vertex has degree k.

Because degree distributions are sometimes hard to interpret is is better to look at the *cumulative* degree distribution. As with all cumulative distributions, the cumulative degree distributions gives you for every degree x the share of vertices in $\mathcal{G}$ with $\delta(v_i) \geq x$.

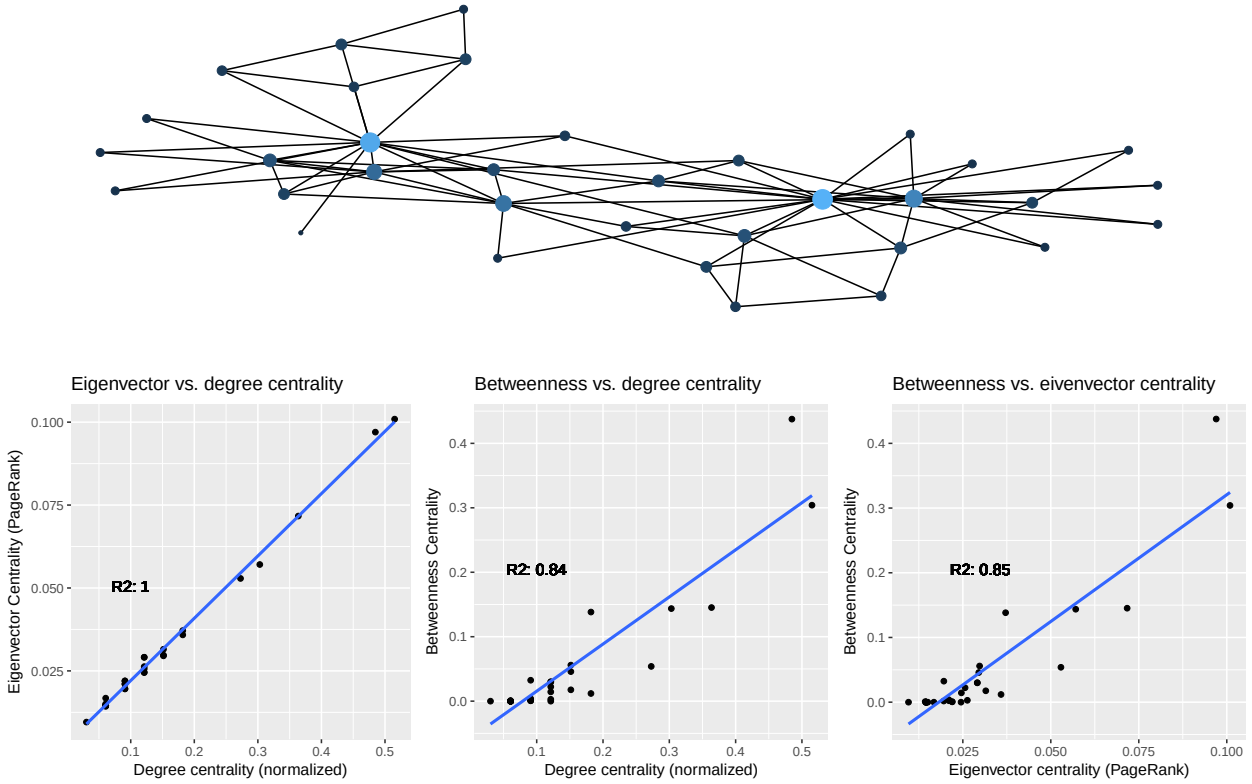


Figure 6: Different centrality measures for friendship network.

3.3.2 Eigenvector centrality

3.3.3 Betweenness centrality

3.3.4 Comparing centrality measures

Sometimes the different centrality measures we have encountered give very similar results, in other cases they might differ quite substantially. To better understand a network it is, thus, usually best to compute all of them, and then make a comparison, e.g. by computing the correlation of the measures.

In figure X we compare the various centrality measures for the so called 'Zachary Karate Club Network', which depicts friendships among members of a karate club. In the graph a vertex represents a member of the club, and an edge between two vertices indicates a friendship relationship between the two members. We see that in this case the different centrality measures are strongly correlated, and are particularly similar for the 'most important' vertices. But this is not necessarily so.

We have now summarized the most important vocabulary to describe networks. In the appendix, in figure 10 and table 4, you find a summary of all the measures introduced, applied to a larger example graph for which visual inspection would already have been impossible.[3]

The next step is to develop some ideal network types, which you may also use for models. Then we move on to network models.

[3]We have skipped one important class of descriptive statistics: measures of centrality. Centrality measures ask which vertices are particularly important in the network. But since there are different notions of 'importance', many different centrality measures exist. If you are interested in the importance of particular notes, you should look these measures up and think carefully about which one to use.

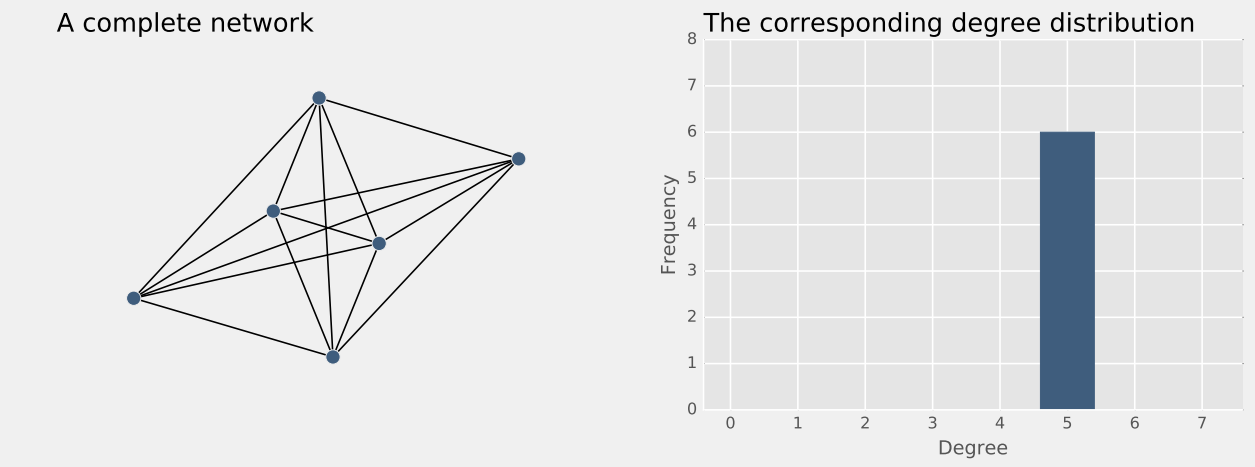
Examples:

- Friendship network for a simple graph
- Importance of GoT characters: Weighted network
- Trade network as a more complex example

4 Ideal network types

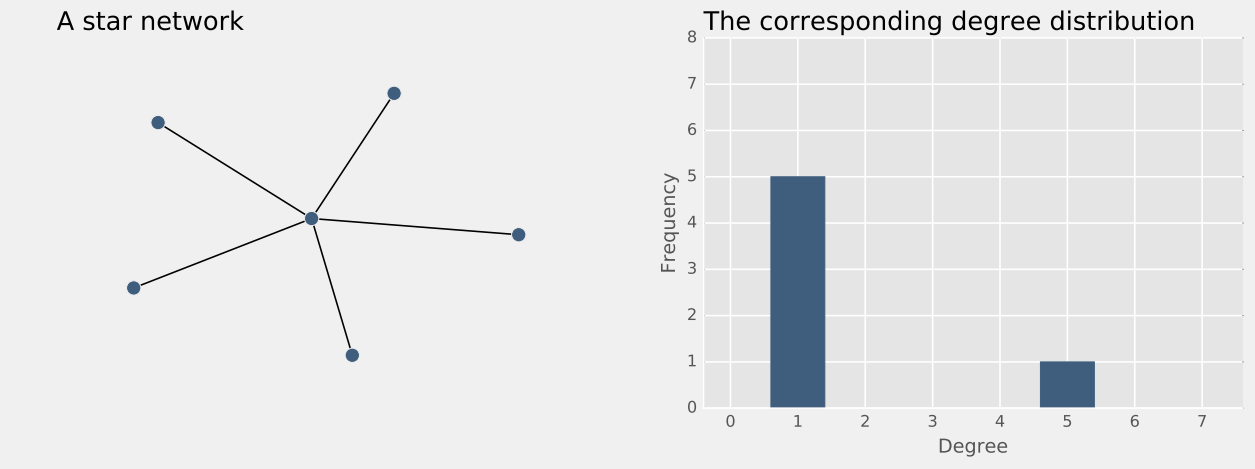
Based on the statistics and concepts introduced so far, we can consider a number of *ideal graph types*. Instances of such graphs are usually not found in the real world, but they capture some important elements of real-world networks and help us to structure our thoughts about real networks. They are also easy to implement in computational models such as ABM and are useful if you want to explore ideas about how the network structure affects the outcome of a certain model

The first ideal type is the complete network. This is a network with all agents being connected to each other. Consequently, all agents have the same degree and $\rho(G) = 1$ (see Definition 10). Theoretically, the complete network is much more important that you might have guessed: before the dawn of network theory, many economic models implicitly assumed that all agents are connected to each other. Many models still continue to do so. Now, imagine you would like to study to what extent the models depend on this – obviously highly unrealistic – assumption? First, you would need to translate the models into a formal language in which you can alter the network structure exogeneously. Most of the time ABM are the natural choice to do so, although there are examples where this can be done analytically. To make sure your model captures the basic mechanisms of the original model, you need to align them for a parametrization that best fits the original model. With regard to the network structure, this most means to use the complete network...



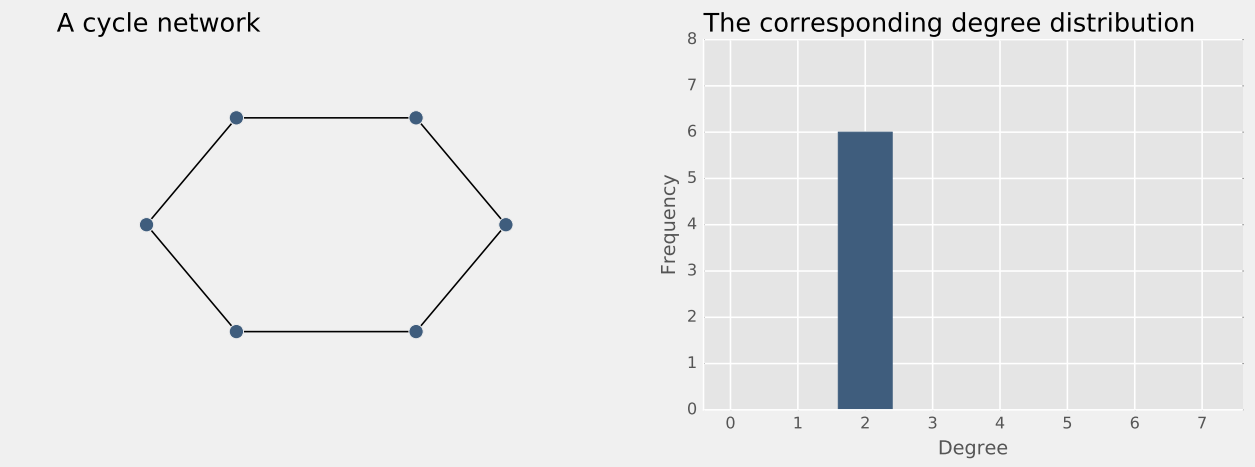
...or the star. In the star network we have one central agent that is connected to all other agents. Every other agent has only one degree, which originates from her connection to the central agent. The star network is important when you want to study situations in which one central authority organizes the interaction of the agents. The very standard general equilibrium setting is of this kind: you have the Walrasian auctioneer who calculates the equilibrium prices based on the demands of all agents.[4]

[4]This is a rather tedious task. Axtell (2005) analyzed the complexity of this decision problem and argues that decentralized trading is – simply because of the complexity argument – a better description of how markets work.

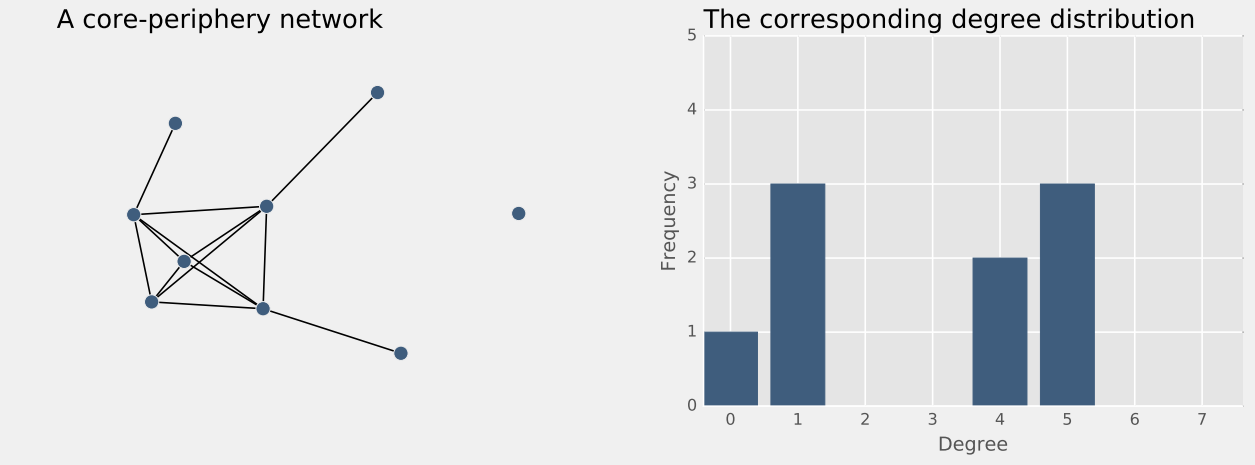


Another network type that has been used to investigate the results of GET in a simulation context (Albin & Foley, 1992) is a cycle network. The cycle network captures situations in which agents interact only locally with their neighbors.

It is also nice because one can often still derive analytical proofs about the proliferation of behaviour in a game theoretic context (see e.g. Bednar, Jones-Rooy, and Page (2015)).



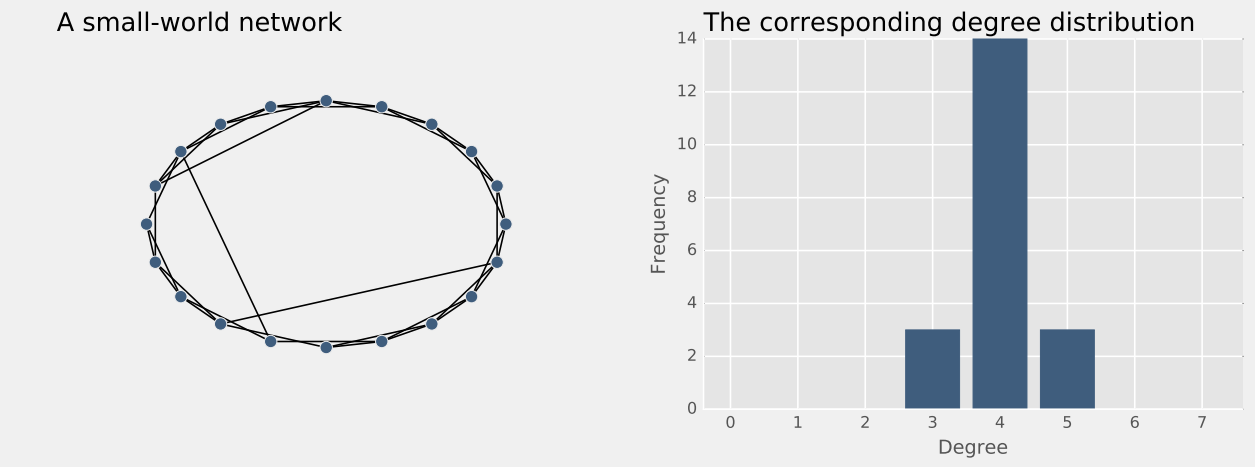
The core-periphery network is similar to the star network: it consists of a central cluster of vertices that is very well connected. The peripheral agents are connected only through the core of the network. Thus, for the vertices in the center, we find a high clustering coefficient, while for the peripheral nodes we do not. Many credit networks and, in particular, the world trade network can be considered core-periphery networks.



We will now move to a bit more complex ideal types, which, in their construction process, involve more randomness than the examples above.

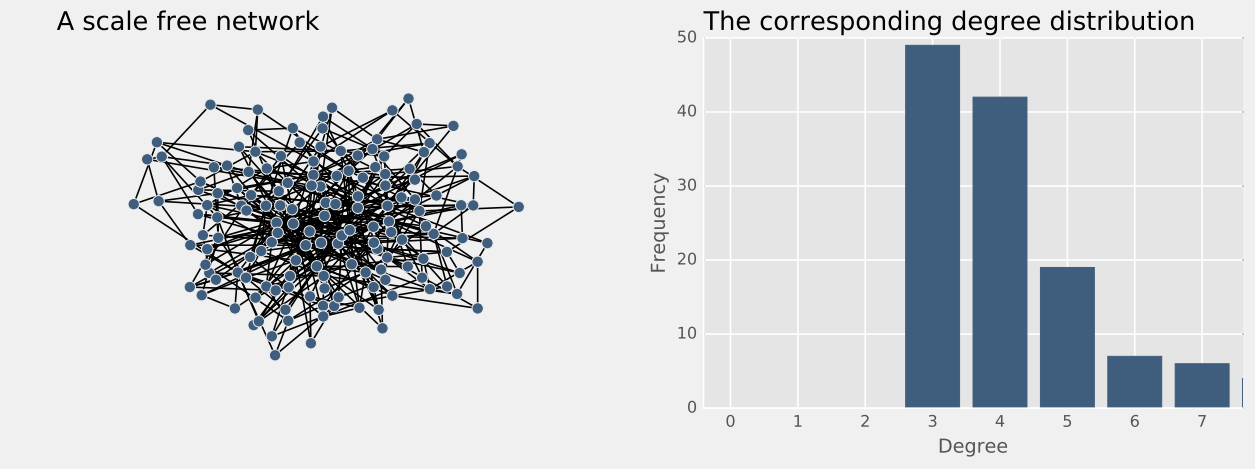
One of the most famous networks is the 'small world network', which was firstly presented by Watts and Strogatz (1998) and was inspired by the famous study of Stanley Milgram according to which all humans in the world know each other via "six degrees of separation". His statement was, of course, much less bold, but that is what the public discourse made out of it. You can read about this study basically everywhere in the web, preferable on wikipedia.

What motivated Watts and Strogatz (1998) was that many real world networks were characterized by (1) a small average path length, and (2) a high degree of clustering.[5] A network that combines these two characteristics is called 'small world network'. To construct such a network, Watts and Strogatz (1998) start with a regular cycle network, and the rewire some edges randomly with probability p. With $p = 1$ one arrives at a random network with short paths and low clustering, with $p = 0$ we have a cycle, which has long paths and high clustering. With only a very small p we observe a transition of the graph characteristics and get a network with high clustering, and short paths.



One striking feature of real world networks is that there are very few well-connected agents with high degrees, and many agents with low degrees. The degree distribution is heavy tailed and often follows a power law:

[5] We will see below that (1) is a common feature of simple random graphs, but that such simple random graphs are not clustered, which shows that there are some social mechanisms that make social networks different to random graphs.



Power laws come up in very many distinct situations and have a deep ontological relation to complex systems. Their ubiquity in social and natural systems is one of the great puzzles of complexity science, and complexity economics in particular.

One very simple – but often very convincing – mechanism to recreate networks with degree distributions that follow a power law is referred to preferential attachment: Start with an small graph and add vertices one by one. Connect new vertices to existing ones probabilistically. Hereby, the probability to connect the new vertex to an old one is an increasing function of the degree of the old vertex. This way, the "rich grow" richer. The most famous formalization was due to Barabási and Albert (1999), who generalized of model used by de Solla Price (1965) to model citation networks.[6] We explore this model in more detail in the second part of this introduction (see section 5.4).

These networks are also 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). We can find scale-free networks in many instances, such as banking networks, friendship networks, production networks, etc.[7]

The small-world and scale-free networks already relate to the second part of this introduction: models for networks. The aim of this second part is to go beyond the description of networks, and to study the mechanisms underlying their genesis and change.

5 Network models

5.1 Introduction

While the first part of the lecture provided you with the basic vocabulary to *describe* networks, this part introduces some basic models of networks. How do these parts relate? And what is the role of models, not only in network theory, but also in research more general? Table 2 provides a *rough* classification of the role of models, depending on the purpose of your research.

In the first case, you would like to describe the features of a system in reality, which you believe can be adequately be represented as a network. Thus, in the very strict sense, models already enter at this descriptive stage: the world trade system, for example, *is not a network*. But it may be useful to represent it and study it as a network. Now, suppose you have represented the world trade system via a network, i.e. you have retrieved that data, read it into your computer and now want to calculate some descriptive

[6] Price himself referred to the ideas of Simon (1955). Herbert Simon is clearly one of the most important figures in complexity science, and maybe science in general, in the 20th century.

[7] A caveat is in order: rigorously identifying power laws is difficult, we will have a section towards the end of this course in which we learn how to do this. See the standard receipt for identifying power laws, which is actually one of the most highly cited papers of our time!.

Goal of analysis	**Role of models**	**Kind of models used**
Description of real world systems as networks	Motivation for choice and development of measures	Previous theoretical models, which tell you where to look; mental models
Identify patterns	Provide reference points to identify "surprising features"	Purely random graph models, e.g. Erdös-Renyi
Re-create networks	Produce artificial networks, e.g. for ABM; suggest new exploratory questions	Statistical random graph models, e.g. stochastic block models
Identify mechanisms	Implement theoretical mechanisms to produce artificial data, which can be compared to empirical data; integrate networks into other models	Mechanismic network generator models, e.g. *preferential attachment*

Table 2: A rough classification of the different activities involved in network research. The boundaries between the categories are not strict, but highlight different core ideas.

statistics. If you only want to describe the network, you may have little idea about want to find in the data. For example, you take your data set about the trade flows between countries in the world and you just want to know how these trade relations look like. Of course, there will never be a situation in which you do not have *any* idea or expectation about what you are going to find. Your empirical view on the world always depends on your prior knowledge, your interests and your personal *Weltanschauung* (cf. Max Weber). This becomes evident when you compute the networks statistics we have discussed so far: after their computation you usually compare them against some benchmark and then consider them as either important and surprising or as unimportant and unsurprising. Thus, there is always at least a mental model involved in your analysis, which tells you what measures to look at. Nevertheless, it seems fair to say that at this stage of the analysis, models play a comparatively minor role, and they mostly come from theories that have motivated you to have a look at the data.

In a next step, you may want to identify certain patterns in the data, such as the exact structure of the world trade flows. To this end, you need to have a more precise reference point. In the context of networks such reference points are provided by *random graphs*, i.e. graphs where vertices and edges exist on purely random grounds.[8] Models of random graphs help you in looking at empirical data and to identify the aspects of your network that are worth mentioning. For example, the world trade network differs from a completely random network in important respects such as its clustering and its degree distribution. It does not differ from the random network with respect of average path lengths. This suggests that (1) the degree of clustering and the trade intensity of countries is of prime interest when studying this networks, and (2) something peculiar is going on with respect to these measures. Thus, in this second step, we use mainly random graph models that provide a useful reference point for empirical networks.

Once you have identified certain patterns in the system you investigate, you may want to consider these patterns in models that refer to this system. For example, if you want to study technological change in open economies, you may want these economies to be connected to each other in a way consistent with the patterns of trade you have identified before. Or, if you want to study the emergence of economic trust among people, you may want the interaction structure to be similar to that of real social networks. In any case, you need to re-create the networks for your model. To do so, it is not so important whether the algorithm that produces this network can be intuitively interpreted, but that it produces a networks that captures the aspects of real-world economies you are interested in. The small-world model of Watts and

[8]The term is a bit ambiguous since many algorithms that create networks according to some clear mechanisms also involve randomness. As always, the distinction is not 100 per cent bullet-proofed, but it highlights the basic idea of relying either on mechanisms or on pure chance.

Strogatz (1998) is an example for a simple model (it is absolutely clear that social network do emerge from neighbourhood structures from which some links are randomly rewired). So called block models frequently used in empirical research are another prominent example. These models are also used to generate new questions: if you are not able to fit your statistical model to the data, then this also tells you something about the direction for future study of these networks.

This leads us to the final class of questions. If you have identified certain patterns in a network, you may want to understand *why* these patterns exist. To this end, you need to derive a theory on why this happens. Taking up the example of the world trade network, which shows a clear core-periphery structure, you could argue, for example, that poor countries tend to be located in the periphery of the world trade network because they do not have the capabilities to produce products that are demanded by households and firms in the rich countries. You could also argue that some countries have been marginalized and denied access to beneficial free trade agreements. Thus, these countries remained poor and are now located in the periphery of the world trade network.

These are two – overly simplistic – hypothesis about the genesis of the world trade network. In the best case you have formalized your hypotheses to a degree such that you can derive its quantitative implications, which you can then test against real data. To this end, you will need a model of a network, and this model should have a plausible structure. This means that the mechanisms in the model should have some reasonable counterpart in the real world: they should represent real mechanisms. Therefore, such models could be termed *mechanism-based models of networks* and they allow for some (more or less clear) *mechanismic* interpretation in the real world (cf. the ontology of Bunge). That means that the mechanisms creating the mathematical graph could also be one of the reasons for why the mathematical graph shares important features with real-world networks. An important example for this is the preferential attachment model, which will feature prominent in this section and which we have mentioned in the context of scale-free networks. It was originally designed to explain why citation networks emerge as scale-free networks, and the hypothesis was that once a paper is cited by several other papers, the probability that others will find it in a reference list and cite it as well, grows. This is exactly the mathematical mechanism in the model, and it is likely that this mechanism operates in the real world as well.

In the rest of this part of the lecture we first use the descriptive tools gathered in the previous section to get an idea of how real-world networks look like.

Then we discuss the most basic random network one can use as a reference point for one's analysis.

Then we study the model of preferential attachment as an example for a mechanismic model of network creation.

5.2 The benchmark: stylized facts about real-world networks

The section must necessarily remain short. But I will provide you with an extensive reading list about empirical studies of economic networks, which you can use to get a better idea about the empirical structure of economic networks.

Here we will focus on three particularly important aspects of real world economies: (1) the degree distribution, (2) the degree of clustering, and (3) the size in the sense of the average path lengths. Of course, the precise results depends on the system you study, particularly how you define the vertices and edges (which is *the most important step in any network study*). But there are still some rough regularities that are actually quite surprising. These regularities are summarized in table 3. From this presentation a first question immediately arises: what do we mean by 'high' clustering and 'small' size? This already implies a certain reference point. This reference point will be the random graph discussed in the next section; after this we compare the statistics of the random graph with the one of some selected empirical networks.

	Empirical networks
Degree distribution	heavy-tailed, often power-law
Transitivity	High
Path lengths	Small

Table 3: Some stylized facts about social networks.

5.3 The simplest random network: Erdos-Renyi

The most common reference point for basically all network models is the Erdös-Renyi model. It produces an undirected simple graph, i.e. a graph without multiple edges and without self-loops (see Definition 2).

In fact, there are two versions, which are both referred to as 'the random graph', or the ER model: one was developed by Erdös and Rényi (1959). The model produces a graph $G(n, m)$, and has two parameters: the number of vertices, n, and the number of edges, m. Thus, one picks, with uniform probability, one of the possible graphs with n vertices and m edges.

The other version was suggested by Gilbert (1959). The model also produces an undirected graph $G(n, p)$, and has two parameters: the number of vertices, n, and the probability that the edge $\langle v_i, v_j \rangle$ exists. Thus, in contrast to the previous model, the number of edges here is a random variable.

In the following we will use the model of Gilbert (1959), which is more common in the actual literature, mostly because it is a bit easier to work with.

The great thing about the model is that for many properties of interest we can get analytical results. But before we move on, a short note on terminology is required. Since the ER model is a probabilistic model, the actual graph produced by this process is almost always different. Figure 7 shows four different graphs that were all created via the ER model with $n = 15$ and $p = 0.2$. If you calculated some descriptive statistics, they would obviously differ from each other. Thus, when one studies the properties of the ER model, one makes statistical statements about the ensemble of all graphs that can be produced by this model (which, in fact, are all graphs with n vertices). In the following I will use the notation $\mathcal{G}(n, p)$ to refer to this ensemble of graphs. Instead, I will use $G(n, p)$ if I refer to one particular graph.

5.3.1 The degree distribution of the ER model

As argued above, there are many different $G(n, p)$ that can be produced by the ER model. But we can make statement about the probability that the ER model produces a graph with particular characteristics, and this is very useful. In particular, we can show analytically that the ER model produces graphs whose degree distributions follow a Poisson distribution, which is not a heavy-tailed distribution. That means that we can say with certainty that the heavy-tailed distributions we find empirically must be caused by different mechanisms.

The probability for an edge to exist is p for all edges (i.e. the edges are *iid*). We will denote the probability to get a graph with exactly m edges as $\mathcal{P}(m)$. Given that there $\binom{n}{2}$ possible edges, and the probability for an edge *not* to exist is $1 - p$, it follows that distribution of $\mathcal{P}(m)$ over $\mathcal{G}(n, p)$ is given by

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

The left half of the equation refers to the probability of the required edges to exist, and the right part to the probabilities that the other edges do not exist.

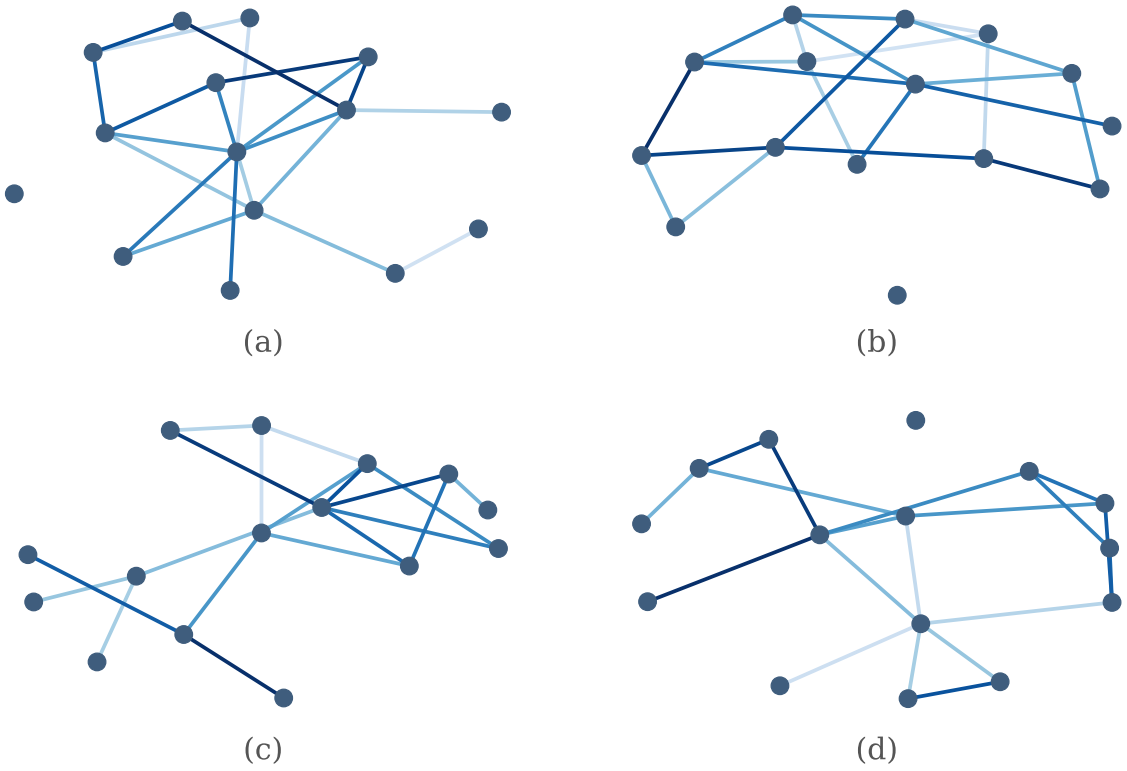


Figure 7: Some graphs that are produced by the ER model with $n = 15$ and $p = 0.2$.

Since the degrees of an ER graph are a random variable, we can ask about its expectation value. The average degree for a vertex in an ER graph corresponds to its expected value $\mathbb{E}(\delta(v_i) = k)$ and can be derived analytically.

Fact 2 (Expected degree of the ER model - vertex level) *The expected degree for any vertex v_i in an ER graph is given by:*

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

Proof 2 *The formulation $\mathbb{E}(\delta(v_i) = k) = \sum_{k=1}^{n-1} k\mathbb{P}(\delta = k)$ comes from the basic definition of the expected value for a discrete random variable as a weighted average of all possible values for this variable. From this it also follows that the probability that a vertex v_i has degree k corresponds to the probability of being joint with k other vertices (which corresponds to the possible connection times the probability for these edges to exist $\binom{n-1}{k}p^k$) and and not being joint with $(n-1-k)$ other vertices (which is given by $(1-p)^{n-1-k}$). We therefore have:*

$$\begin{aligned}
\mathbb{E}(\delta(v_i) = k) &= \sum_{k=1}^{n-1} k\mathbb{P}(\delta = k) \\
&= \sum_{k=1}^{n-1} k\binom{n-1}{k}p^k(1-p)^{n-1-k} \\
&= \sum_{k=1}^{n-1} k\frac{(n-1)!}{k!(n-1-k)!}p^k(1-p)^{n-1-k} \\
&= \sum_{k=1}^{n-1} k\frac{n-1}{k}\frac{(n-2)!}{(k-1)!(n-1-k)!}pp^{k-1}(1-p)^{n-1-k} \\
&= pk\frac{n-1}{k}\sum_{k=1}^{n-1}\frac{(n-2)!}{(k-1)!(n-1-k)!}p^{k-1}(1-p)^{n-1-k} \\
&= p(n-1)\sum_{k=1}^{n-1}\frac{(n-2)!}{(k-1)!(n-1-k)!}p^{k-1}(1-p)^{n-1-k}
\end{aligned}$$

It is now simpler to proceed with $l = k-1$:

$$\begin{aligned}
\mathbb{E}(\delta(v_i) = k) &= p(n-1)\sum_{k=1}^{n-1}\frac{(n-2)!}{(k-1)!(n-1-k)!}p^{k-1}(1-p)^{n-1-k} \\
&= p(n-1)\sum_{l=0}^{n-2}\frac{(n-2)!}{l!(n-l-2)!}p^l(1-p)^{n-l-2} \\
&= p(n-1)\sum_{l=0}^{n-2}\binom{n-2}{l}p^l(1-p)^{n-l-2}
\end{aligned}$$

Now fix $m = n-2$ to simplify the last line on the right to $p(n-1)\sum_{l=0}^{m}\binom{m}{l}p^l(1-p)^{m-l}$, in which the sum yield 1 and we remain with $\mathbb{E}(\delta(v_i) = k) = p(n-1)$. □

This means the expected degree is the number of all possible relations (i.e. $n-1$) times p, an intuitive result because c is the number of possible neighbours times the probability that a link to another vertex exists.

But this is not yet the final result we wish to acquire. Most of the time when dealing with ER graphs we are interested in large graphs (with large n) and small to moderate values of p. In this case, we can make more precise statements on the resulting degree distribution.

Fact 3 (Degree distribution of ER) *For larger ER with small n, the degree distribution is Poisson.*

Proof 3 *First note that an equivalent way of expressing equation 2 is $\mathbb{E}(\delta(v_i) = k) = \binom{n-1}{k} p^k \cdot (1-p)^{n-1-k}$. For small p we can simplify the right term by first taking logs and then expanding it as a Taylor series:*

$$\ln[(1-p)^{n-1-k}] = n - 1 - k\ln(1-p)$$

and from equation 2 we know that $\frac{c}{n-1} = p$ such that

$$\begin{aligned}\ln[(1-p)^{n-1-k}] &= n - 1 - k\ln(1 - \frac{c}{n-1}) \\ &\approx -(n-1-k)\frac{c}{n-1} \approx -c\end{aligned}$$

It follows that $\mathbb{E}(\delta(v_i) = k) \approx \binom{n-1}{k} p^k e^{-c}$. Now consider the term $\binom{n-1}{k}$:

$$\binom{n-1}{k} = \frac{(n-1)!}{(n-1-k)!k!} \approx \frac{(n-1)^k}{k!}$$

which we can plug in to get the desired approximation.

$$\begin{aligned}\mathbb{E}(\delta(v_i) = k) &\approx \frac{(n-1)^k}{k!} p^k e^{-c} \\ \mathbb{E}(\delta(v_i) = k) &\approx \frac{(n-1)^k}{k!} (\frac{c}{n-1})^k e^{-c} \\ \mathbb{E}(\delta(v_i) = k) &\approx \frac{c^k (n-1)^k}{(n-1)^k k!} e^{-c} \\ \mathbb{E}(\delta(v_i) = k) &\approx \frac{c^k}{k!} e^{-c}\end{aligned}$$

As you can see from the plots of the Poisson distribution in figure 8, this distribution is not heavy-tailed.

5.3.2 Transitivity the ER model

From the definition of the ER model it follows that the edges are identically and independently distributed (*iid*). In fact, this is one reason why the $\mathcal{G}(n,p)$ model is a bit easier to handle than the $\mathcal{G}(n,m)$ model. Therefore, it is easy to calculate network transitivity (see definition 13).

Fact 4 (Transitivity of the ER graph) *For graphs produced by the ER model, the following result regarding their transitivity does hold:*

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

The formula is built following the same logic we described during the definition of transitivity: there are $\binom{n}{3}$ possibilities for triangles in a graph with n vertices. For each of this triangle to exist, we need three particular edges to exist; since edges are *iid* and every edge exists with probability p, the expected number of triangles is the number of possible triangles times their probability, i.e. $\binom{n}{3}p^3$. The argument for triples and the denominator is analogous, the rest follows from the previous discussion.

Fact 4 implies that ER graphs tend to have few triangles, particular for large n. This is not what we observe in real world social networks, for which we know that transitivity is high. The mechanisms underlying the generation of relationships in social networks thus likely differ from pure randomness.

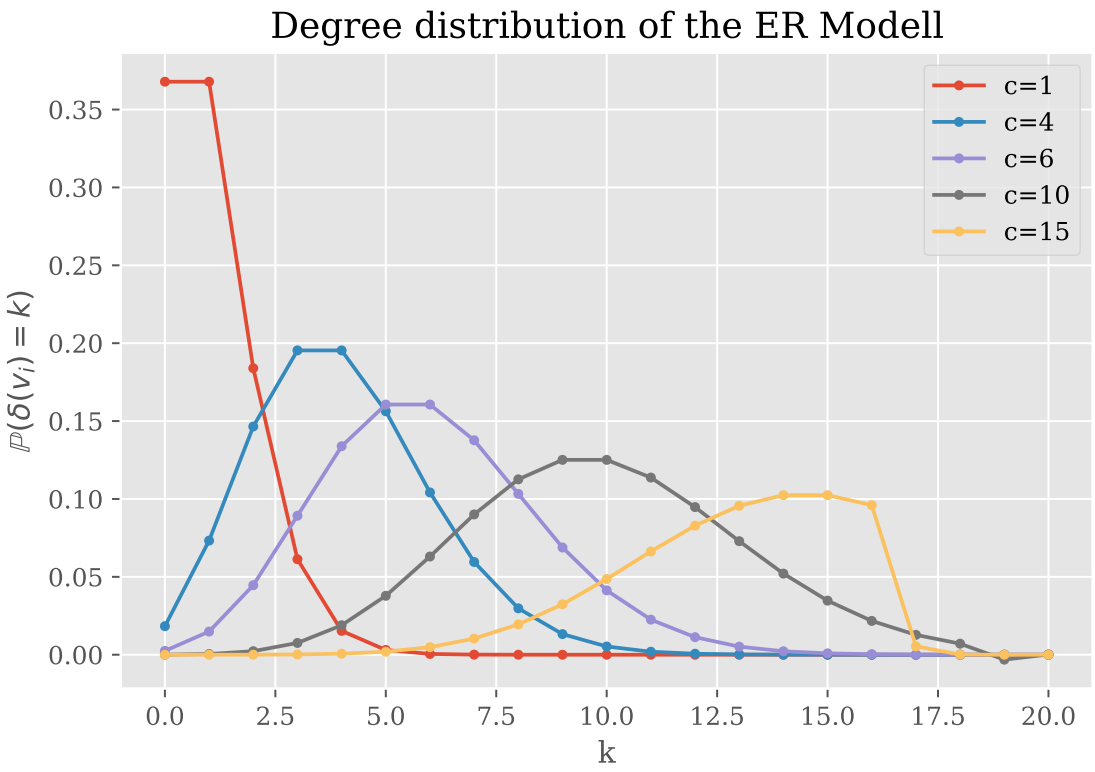


Figure 8: The Poisson distribution is not a heavy tailed distribution.

5.3.3 What do we learn from ER?

The ER model is an important reference point for network analysis. We only considered two particular properties of the ER graph: its degree distribution and its clustering. Both do not fit with what we know about real world networks. Does this mean the model is useless? No, because it helped us to be clear by what we mean by "high transitivity" or "heavy tailed" distributions.

Also, and we did not cover this here, we know that the average path lengths in ER graphs is small. This is also something we observe in many economic networks, but given that path lengths are small even in random graphs indicates that this feature of real world interaction structures is less surprising.

5.4 Preferential attachment and scale-free networks

As indicated earlier, heavy-tailed degree distributions, particularly power-law distributions, are a striking empirical regularity found in many social and natural networks. One of the most famous mechanisms that has been shown to generate power-law distributions is *preferential attachment*, formerly known as *cumulative advantage*. Herbert Simon showed showed (Simon, 1955) that this process generates power-law distribution when he was interested in the distribution of wealth, which is clearly power-law distributed. The basic idea of preferential attachment is that those who already have much, tend to get even more in the future. This reasoning can be applied to networks when we say that those vertices that already have a high degree tend to increase their degree over time more dramatically than those vertices, which have low degree.

The first application of this idea was due to de Solla Price (1965), who built a model of directed graphs and applied it to citation networks. The distribution of citations is clearly power law distributed, and Price conjectured that this is due to a mechanism he termed *cumulative advantage:* papers that are highly cited appear in many reference lists, and are thus more likely to get cited in the future. In contrast to pure random graph models, his model was stochastic, but it formalized a clear mechanism. It thus counts as a mechanismic model, the last category discussed in table 2.

The most widely known model, which generates undirected networks is due to Barabási and Albert (1999). This model, which has two parameters, n and m, generates a network as follows:

Start with network of few (n_0) vertices without edges, and add vertices of degree $m < n_0$ subsequently until the network has n vertices. In every step, the edges of the new vertices are wired to existing vertices probabilisitically, and the probability that a vertex v_i gets chosen, p^i, depends *linearly* of $\delta(v_i)$:

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

This process produces a network such as this shown in figure 9. The degree distribution of these networks follows a power-law and the model is thus consistent with empirical observations in many different areas of research.

Fact 5 (Degree distribution of the BA model) *For large k, the degree distribution of the BA models follows a power law:*

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

Proof 4 *Using differential (or 'master') equations, see Bollobás, Riordan, Spencer, and Tusnády (2001).*

The specification 'for large k' is necessary because random variables are power-law distributed only after a certain cut-off value. We will discuss this in another session, but see this receipt for details.

There is one thing that I want to highlight at the end of this section: First, the BA model suggests a certain mechanism that accounts for the power-distributed degrees found in many networks. This is great because it satisfies the epistemological demands we have formulated previously. But while the *preferential attachment* mechanisms is *compatible* with the observation of power laws, this is in itself no evidence that it actually

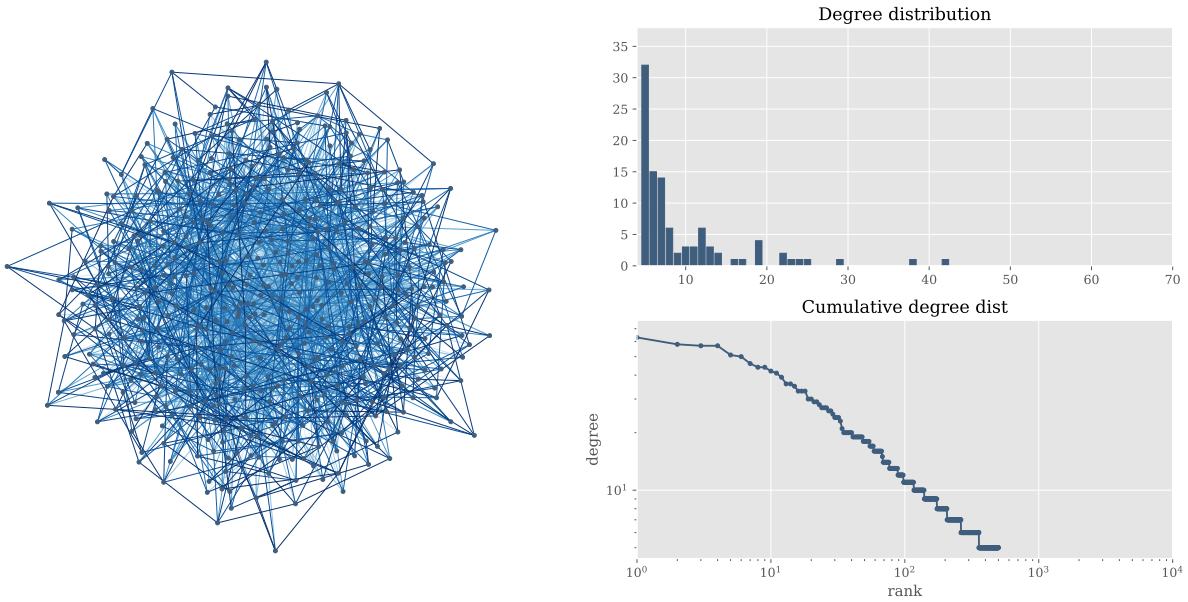


Figure 9: A graph generated by the BA model, and its degree and cumulative degree distribution.

is this mechanism. As it was highlighted by Bunge (2004, p. 186), real world mechanisms are "concealed" and "have to be conjectured", which is why there cannot be absolute certainty about which mechanism is operating. Fortunately, we can study the plausibility of models embodying a particular mechanism further using the various validation approaches discussed in the session on meta-theory. For example, with regard of BA's preferential attachment model, Krapivsky and Redner (2001) have shown that another implication of the model is that the age of a vertex correlated with its degree. The applicability of the mechanism as discussed by BA can then be further investigated by testing, whether this correlation also holds for the system under investigation. For example, Adamic and Huberman (2000) argue that because of this correlation, the model does not work for web-pages, because we cannot observe such a correlation there. Thus, you should never put too much faith in models, particularly not so simple models such as the one by BA, of which, many variants exist. After all, they are only representations of reality, and they should be treated as such.

The second point I want to make is related: while the preferential attachment mechanism produces power-law distributed degrees, it is not the only mechanism that does so. Therefore, once you have found a mechanism consistent with your observations, this should not be a reason to look for further, maybe even more plausible, mechanisms.[9] In the case of innovation networks, Powell, White, Koput, and Owen Smith (2005) or Fritsch and Kudic (2016) make a similar arguments: they argue that the preferential attachment mechanism is not able to explain the emergence of innovation networks , although it can be calibrated to fit the empirical degree distribution. Rather, they argue that mechanisms such as *triadic closure* might be more plausible and more consistent with what is going on the real world. This result is important, because the practical implications, e.g. if one wishes to avoid power law distributions, are different. It also serves as a great motivation for doing ever more and better research: the process of good scientific investigation is open-ended, and there is always potential for improvement.

This illustrates that if one investigates complex systems, it is important to seek (1) mechanismic explanation and (2) preserve an epistemological fallibilism.

[9] Also, mechanisms can often be further specified. As indicated in equation 5.4, the BA model features *linear* preferential attachment. For many situations of technology choice, non-linear processes have been shown to be a better representation of reality.

6 Outlook

You are now familiar with the basic concepts of network theory. After the python labs, you will also be able to generate artificial networks.

You could then immediately consider the structure of such networks in your models, at least if the kind of models you use does allow for this. Later in the course we learn about agent-based modeling, a methodology that is well-suited to integrate knowledge about network structure into a model.

We have spent less time on the empirical analysis of networks. We also spent less time on concrete models of networks and their analytical treatment. These are interesting questions, but they also require more time. However, with the knowledge of this short course and the references provided below, you are very well equipped to explore these avenues on your own!

A Historical roots of graph theory

To be added shortly

B Example graph and overview over basic graph measures

This example graph summarizes the formal graph measures introduced before.

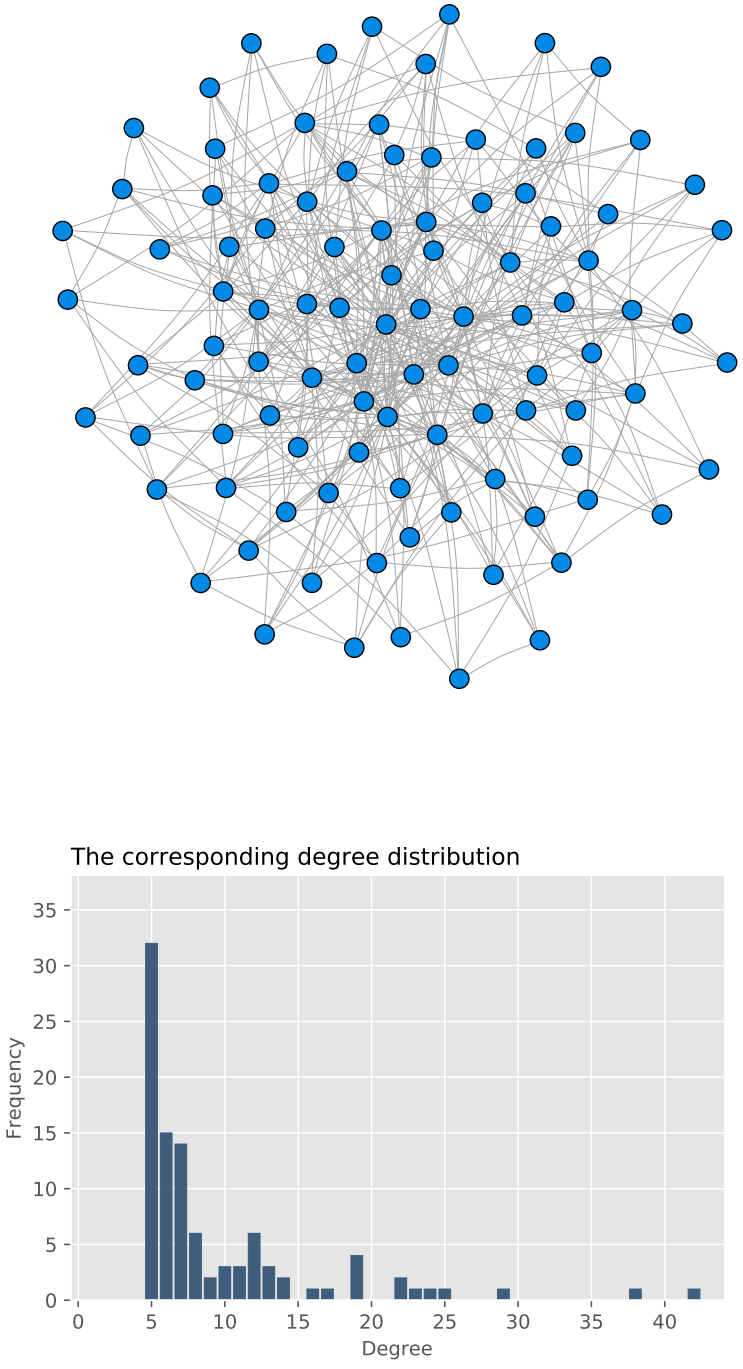


Figure 10: An example graph. The measures we discussed so far are summarized in table 4

Measure	Level	Graph type	Definition	Def nb	NetworkX command	Value
Nb of vertices	System	D, U	$n = \#V(G)$	1	`g.number_of_nodes()`	100
Nb of edges	System	U	$m = \frac{1}{2}\sum_{ij} \boldsymbol{A}_{ij}$	1	`g.number_of_edges()`	475
Nb of arcs	System	D	$a = \sum_{ij} \boldsymbol{A}_{ij}$	1	`g.number_of_edges()`	NA
Neighborhood	Vertex	U	$\mathcal{N}(v_i) := \{v_j \in V(G) \vert i \neq j, \exists e \in E(G) : e = \langle v_i, v_j \rangle\}$	3	`nx.all_neighbors(g, v)`	NA
Connectedness	Vertices	U	$d(v_i, v_j) \neq \infty$	5	`nx.has_path(g,` v_i`,`v_j`)`	NA
Connectedness	System	U	$d(v_i, v_j) \neq \infty \ \forall \ v_i, v_j \in V(G)$	5	`nx.is_connected(g)`	True
Distance	Vertices	U, D	length shortest path (v_i, v_j)	6	`nx.shortest_path(g,` v_i`,` v_j`)`	NA
Distance	Vertices	U, D	length shortest path (v_i, v_j)	6	`nx.shortest_path(g,` v_i`,` v_j`)`	NA
Diameter	System	U, D	$diam(G) = \max\{(d(v_i, v_j)\}_{i,j \in V(G)}$	7	`nx.shortest_path(g,` v_i`,` v_j`)`	4
Av path length	System	U, D	$\bar{d}(G) = \frac{1}{n}\sum_{v_i \in V(G)} \bar{d}(v_i).$	8	`nx.average_shortest_path_length(g)`	2.2
Char path length	System	U, D	$CPL(G) = median\{(d(v_i, v_j)\}_{i,j \in V(G)}$	9	`XXXXX`	NA
Density	System	D	$\rho(G) = \frac{a}{n(n-1)}$	10	`nx.density(g)`	NA
Density	System	U	$\rho(G) = \frac{2m}{n(n-1)}$	10	`nx.density(g)`	0.096
Transitivity	System	U	$\tau = \frac{\sum_{ijk} a_{ij}a_{jk}a_{kj}}{\sum_{ijk} a_{ij}a_{jk}}$	13	`nx.transitivity(g)`	0.16
Clustering	Vertex	U	$cc(v_i) = \frac{\sum_{ijk} a_{ij}a_{jk}a_{kj}}{\delta(v_i)(\delta(v_i)-1)}$	11	`nx.clustering(g,` v_i`)`	NA
Clustering	System	U, D	$CC(G) = \frac{1}{n}\sum_{v_i \in V(G)} cc(v_i)$	12	`nx.average_clustering(g)`	0.218
Degree	Vertex	U	$\delta(v_i) = \sum_{j=1}^{n} a_{ij}$	14	`g.degree(`v_i`)`	NA
In-degree	Vertex	D	$\delta^{in}(v_i) = \sum_{j}^{n} a_{ji}$	15	`g.in_degree(`v_i`)`	NA
Out-degree	Vertex	D	$\delta^{out}(v_i) = \sum_{j}^{n} a_{ij}$	15	`g.out_degree(`v_i`)`	NA
Av degree	System	U	$\bar{\delta}(G) = \frac{2m}{n}$	NA	`np.mean(list(g.degree().values()))`	9.5
Av in/out-degree	System	D	$\bar{\delta}(G) = \frac{2m}{n}$	NA	`np.mean(list(g.out_degree().values()))`	NA

Table 4: The network measures discussed so far for the example graph in figure 10 It is assumed that the graph in connected and unweighted and $a_{ij} \in \boldsymbol{A} = 1$ if $\langle v_i, v_j \rangle$ exists and zero otherwise. The NetworkX code assumes the graph is named `g`.

References

Adamic, L. A., & Huberman, B. A. (2000). Power-Law Distribution of the World Wide Web. *Science*.

Albin, P., & Foley, D. K. (1992). Decentralized, dispersed exchange without an auctioneer. *Journal of Economic Behavior and Organization*.

Axtell, R. (2005). The Complexity of Exchange. *The Economic Journal*.

Barabási, A. L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*.

Bednar, J., Jones-Rooy, A., & Page, S. E. (2015). Choosing a future based on the past: Institutions, behavior, and path dependence. *European Journal of Political Economy*.

Bollobás, B., Riordan, O., Spencer, J., & Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*.

Bunge, M. (2004). How Does It Work? The Search for Explanatory Mechanisms. *Philosophy of the Social Sciences*.

Cartwright, N. (2007). *Hunting Causes and Using Them*. Cambridge University Press.

de Solla Price, D. J. (1965). Networks of Scientific Papers. *Science*.

Erdös, P., & Rényi, A. (1959). On Random Graphs I. *Publicationes Mathematicae*.

Fritsch, M., & Kudic, M. (2016). Preferential Attachment and Pattern Formation in R&D Networks - Plausible explanation or just a widespread myth? *Jena Economic Research Papers*.

Gilbert, E. N. (1959). Random Graphs. *The Annals of Mathematical Statistics*.

Krapivsky, P. L., & Redner, S. (2001). Organization of growing random networks. *Physical Review E*.

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review*.

Powell, W. W., White, D. R., Koput, K. W., & Owen Smith, J. (2005). Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American Journal of Sociology*.

Simon, H. A. (1955). On a Class of Skew Distribution Functions. *Biometrika*.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*.