Chapter 1

Methodology

1.1 Complex networks

The complex system has several properties. It consists of many components such as units or individuals who interact with each other. The properties of the complex system can not be predicted from the behaviour of one individual. In such systems, without any central force, collective behaviour can emerge. In societies, people's interactions lead to civilisation, economy, formation of social groups or even traffic on the roads. In the animal, populations are present at different levels of the organisation, as in ants and bees colonies or the school of the fishies showing flocking patterns. [**?**]

The research in complex systems focuses on the structure of the interactions between units. Knowing how branches of the system are connected, we can determine the emergence of the collective behaviour of the system. We can construct networks with neurons and synapses, representing their connections. The structure of the brain network and its properties are fundamental for brain functioning, and neurons in the same brain area are closely connected. Similarly, we can represent the communication between people. The structure of these interactions gives us insights, for example, how information propagates through the system. The presence of people with many connections can lead to faster information flow.

Despite the differences between complex systems, they can be studied using complex networks; with sets of nodes and edges. Elements in the system are nodes, while interactions between them are given as edges. This approximation allows us to treat equally social (graph of actors), biological (network of proteins) or even technological systems (internet, traffic). In recent years, complex network theory has application in different fields, and the availability of big data incurs its development.

The complex network theory originates from the graph theory in mathematics. These days, the graph and network are used as equivalent terms. The first mathematical problem solved using graph theory was *Konigsberg* problem of seven bridges. The city *Konigsberg* had seven bridges connecting the city's parts across the river and the island in the middle. The question was, is it possible to find a walk that crosses all seven bridges only once. Representing the problem as a graph, Euler managed to simplify the problem; the parts of the land are represented as nodes while bridges between them are links. Crossing each bridge only once is possible if each part of the land has an even number of connections. Thus, it was not possible in this case, as each piece of land had an odd number of bridge connections, see Fig. 1.1.

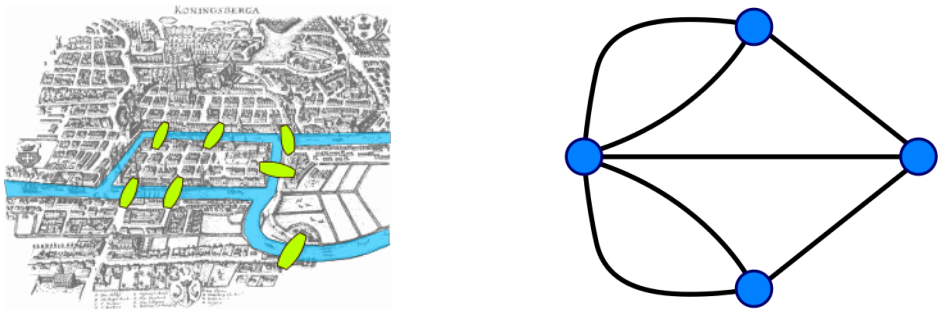

Figure 1.1: The Kronigsber problem of seven bridges.

1.2 Types of networks

The graph or network G is defined as $G = (V, E)$, where V is a set of nodes (vertices), and E is a set of edges. The edge is pair of nodes $e_{ij} = (i, j)$, and $i, j \in E$. The graph structure has undirected edges meaning that edges are symmetric: (i, j) implies (j, i). Edges are also unweighted, meaning that all edges are equally important. The specific properties of the nodes are also neglected in this representation. The **adjacency matrix** $A = N \times N$ has value 1 if there is connection between two nodes, otherwise it is 0 **[?]**

For example, if we consider unweighted, undirected network, with only 3 nodes and two connections, adjacency matrix is:

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{1.1}$$

The self loops usually are not considered, meaning that $A_{ii} = 0$.

Other equivalent way of representing graph is as edge list. Instead of having adjacency matrix, graph is described with the list of links that are in graph. The graph is then pair of N, g, where N is set of nodes, and g is collection of links, listed as subset of N of size 2. So the network is written as $g = \{\{1, 2\}, \{2, 3\}\}$.

Sometimes it is essential to include specific properties of the system in the network representation, which can help create more realistic models. The additional properties can be added on the edge, node or network level.

In a **directed network**, edges have broken symmetry. The interaction from node i to node j does not need to have the same property as the interaction from node j to node i. A typical example is WWW, where webpages are nodes, and hyperlinks are directed edges. In biological networks, gene regulation and neural activation can be given as directed networks.

The frequency of interaction between nodes is emphasised if edges are associated with different scalar values; networks are **weighted**. The edges may be signed, representing activation in the biological system or trust and distrust in the social system. In general, edges can be associated with any categorical variable. If attribute describes the time when an interaction between nodes happened; network is called **temporal**. Finally, **multigraph** allows the presence of multiply edges between two nodes. For the network between cities, edges may be different driving paths between them. In neuron cells, multiply synapses are represented as distinct edges.

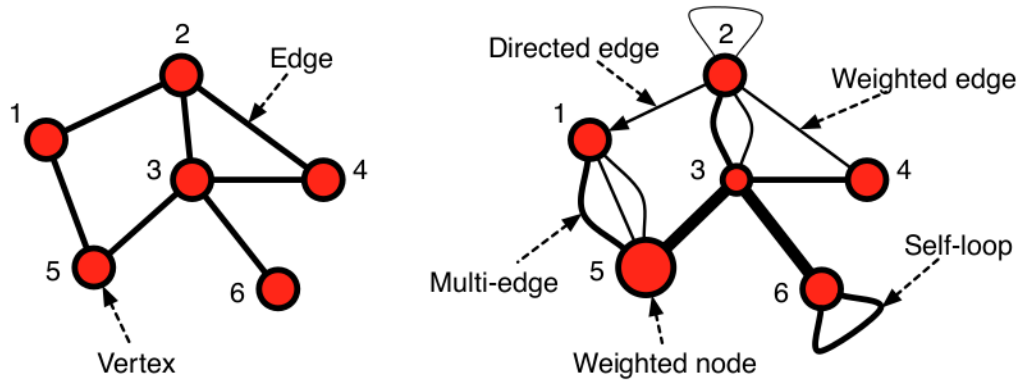

Figure 1.2: Graph types

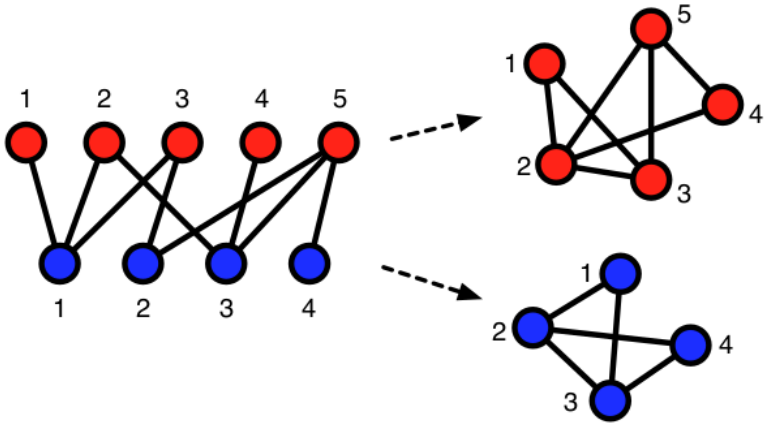

Figure 1.3: Graph types

A **bipartite network** has two partitions, U and V. The nodes in the same partition are not connected while links exist only between nodes of a different kind. In general, we can define k-bipartite graph. The set of nodes V has k distinct classes of nodes. When $k = 2$ the network is bipartite. Bipartite networks represent the membership of people or items in groups. For example, we can define the network of actors as a bipartite graph. In one partition are actors and in other movies. There are no edges between actors or movies, but the actor is connected to the film if it plays in that movie. Another example is a recommender network, such as a network of people and items they like.

The equivalent representation of bipartite network is incidence matrix B. If n is number of people and g number of groups, this matrix is gxn, having elements B_{ij} 1 if persoon i belongs to group j.

Even bipartite networks give realistic representation of the system, there is often need to analyze the single type of nodes. From a bipartite network, we can generate two projections. The first one connects nodes partition V if they point to node u. Similarly, we can project the network on U partition, connecting u nodes. The one mode projection between actors and movies onto actors is undirected network of actor collaborations. Actors are connected if they appear in the same movie. We can also create one-mode projection onto movies, where two movies are connected if they share the same actor.

The projections are useful in some manner, but they also lose some important information, for example how many groups nodes share in common. This information can be propagated adding the weight to the edges, equal to the number of common groups.

The product B_{ki} and B_{kj} is 1 if i and j belong to the same group k. Thus the total number of groups to which nodes i and j belong is $P_{ij} = \sum_{k=1}^{g} B_{ki}B_{kj} = \sum_{k=1}^{g} B_{ik}^{T}B_{kj}$. The matrix P is matrix of one-mode projection. The diagonal elements are non-zero, and represent the number of groups node i belongs to. To derive the weighted adjacency matrix, the diagonal elements are set to 0. The adjacency matrix of unweighted projection, each non-zero element needs to be replaced with 1.

The important consequence of the one-mode projection is the construction of the cliques; subgraph in which every pair of nodes is connected. Every node that is being projected is represented as a clique of size, because all pairs of its neighbours are exactly distance two away from each other. All actors in the movie will be joined in a clique in the one-mode actor projection.

Another consequence of the one-mode projection is that one mode projection may originate from different bipartite networks. Meaning that projection is not one-to-one; no bijection. It is surjective operation, meaning hat any projected network P has at least one bipartite network such that the projection of B results in P.

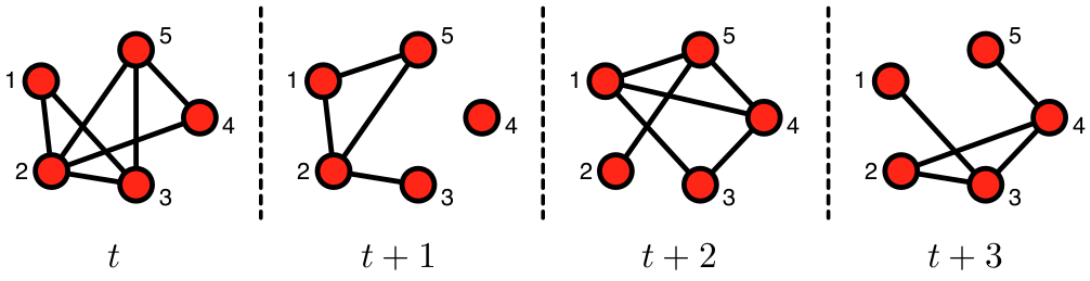

Figure 1.4: Graph types

In **temporal networks** nodes and edges evolve. Many real networks are not static, networks grow over time, and edges and nodes may emerge or disappear. Also, some edges may be active at regular intervals, reflecting the circadian rhythms of the nodes. In citation networks, new nodes (paper) join the network, creating links with cited papers.

Let consider the temporal network with N nodes, over time interval t_{max}. The event representation consists in viewing the temporal networks as a collection of time-stamped edges. In this representation, each edge (i, j) is defined as$(i, j, t, \Delta t)$, where t is the time of the event and Δt its duration. In a temporal network, the same link can appear multiple times, and the duration of the event may vary. An example of an event temporal network may be phone-calls networks, where a call between two persons i and j started at time t and ended at $t + \Delta t$.

A temporal network can be represented as sequence of networks $G = G(1), ...G(t_{max})$, where t_{max} is number of networks. While in the event-based representation, time can be both discrete and continuous, in the snapshot representation, time is only discrete. The temporal network is seen as a structure that evolves in time, and at each timestamp, we can analyse the system's macroscopic properties. We need to specify time windows for coarse-grain event-based representation of the temporal information in the data. If we use uniformly

time window of length w then the events occurring in $0 \leq t < w$ enter snapshot G(1), those occurring in $w \leq t < 2w$ enter G(2) and so forth. We can not recover the original data points from the snapshot representation, and more information loss is present with a larger time window. If the time window is $w = t_{max}$, there is only one snapshot, temporal data are no more available, and the network is static.

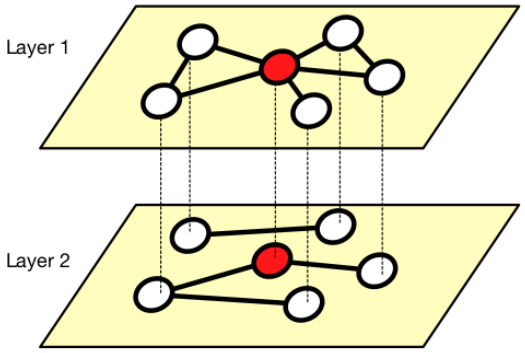

Figure 1.5: Graph types

A network in which edges are marked by which "layer" they exist in is called a multiplex or multilayer network. These networks are used to represent a system in which there are multiple types of interactions, and we store the connectivity of each kind in a different "layer" of the multiplex network. A temporal network is a special kind of multiplex network, where these layers form a temporal (ordered) sequence. Crucially, there can dynamics on each vertex that govern which layer some kind of interaction occurs on, so multiplex networks are not merely a special kind of graph in which different colors or layer numbers annotate edges.

Spatial networks are a special kind of node-annotated network, in which the annotations represent the node's location in some d-dimensional space. This graph property is most common in transportation networks, e.g., as road and city networks, airport transportation networks, oil and gas distribution networks, shipping networks, etc., but can also appear in social networks. Planar graphs are a special case of spatial networks in which the nodes are embedded on a 2-dimensional surface and edges do not cross.

Hypergraphs are another type of network, in which edges denote the interaction of more than two vertices, e.g, E in V × V × V . Scientific collaboration graphs can be represented as a hypergraph, in which each "edge" is the set of coauthors on a scientific article. However, collaboration networks are more commonly represented as bipartite graphs, in which scientists and papers form two sets of vertices, and scientist-nodes are connected to all the paper-nodes on which they are authors.

1.3 The structure of complex networks

1.3.1 Paths and cycles

As network is connected structure, the nodes may be influenced also by distant nodes, and the information might spread through the links of a network. Analyzing the paths of the network is important task.

A path of the network between two nodes i and j is a sequence of links, $i_1 i_2, .., i_k$ such $i_k i_{k+1} \in g$ that $i_1 = i$ and $i_k = j$, each node in the sequence is distinct. The path is walk, if

nodes in the sequence are not distinct, so walk can visit one node more than once. The cycle is walk that starts and ends at the same node, while other nodes are distinct.

Note that, if we use adjacency matrix representation where $A_{ii} = 0$, then A^2 tells us how many walks of length 2 exist between any two nodes.

Network is connected if every two nodes in the network are connected by some path. So the network is connected id for every node $i \in N$ and node $j \in N$ exists a path between them.

A component of the network are distinct maximal connected subgraphs of a network. In the example there are 4 components. Note that in this definition of the component isolated node is also component. In directed network, set of nodes that are reachable from each other is a strongly connected component, while a set of nodes where either i is reachable from j or j is reachable from i is weakly connected component.

There are a few particular network structures that are commonly referred to. A tree is a connected network that has no cycles. A forest is a network such that each component is a tree. Thus any network that has no cycles is a forest, as in the example pictured in Figure 2.1.6. A particularly prominent forest network is a star. A star is a network such that there exists some node i such that every link in the network involves node i. In this case i is referred to as the center of the star. There are a few facts about trees that are easy to derive (see Exercise 2.2) and worth mentioning. A connected network is a tree if and only if it has n 1 links. A tree has at least two leaves, where leaves are nodes that have exactly one link. In a tree, there is a unique path between any two nodes. The complete network is one where all possible links are present, so one where where g ij = 1 for all i 6 = j.

A circle (also known as a cycle-graph) is a network that has a single cycle and such that each node in the network has exactly two neighbors. In the case of directed networks, there can be many di¤erent stars involving the same set of nodes and having the same center, depending on which directed links are present between any two linked nodes. On occasion, it will be useful to distinguish between these.

Eulerian tours and Hamiltonian cycles

A walk is said to be closed if it starts and ends at the same node. It is clear that in order to have a closed walk that involves every link of a network exactly once it must be that each node in the network has an even degree. 39 This follows since each time a node is "entered" by one link on the walk it must be "exited"by a di¤erent link, and each time the node is visited, it must be by a link that has not appeared previously on the walk. Euler's [?] simple but remarkable theorem is that this condition is necessary and su¢ cient for there to exist such a closed walk.

A connected network g has a closed walk that involves each link ex- actly once if and only if the degree of each node is even.

One can ask a related question for nodes rather than links: when is it possible to …nd a closed walk that involves each node in the network exactly once? Such a closed walk must be a cycle, and is referred to as a Hamilton Cycle or a Hamiltonian. A related question is whether there exists a "Hamilton path"that hits each node exactly once. Clearly a network that has a Hamilton cycle has a Hamilton path, while the converse is not true (consider a line). Discovering whether or not a network has a Hamilton cycle is a much more chal- lenging question than whether it has a Euler tour; and this has been an active area of research in graph theory for some time. It has direct applications to the "traveling salesman problem,"where a salesman must visit each city on a trip exactly once, cities are nodes on a network, and the path must follow the links. The seminal theorem on Hamilton cycles is due to Dirac [?]. Stronger theorems have since been developed, as we shall shortly see, but it is worth stating on its own, as it has an intuitive proof that helps one see the paths to proving some of the later.

1.3.2 Community structure

1.3.3 Core-periphery structure

1.4 The network analysis

1.4.1 Degree distribution

The simplest network measure is **node degree**, k. The degree of node i gives the number of nodes attached to node i, $k_i = \sum_j A_{ij}$.

The density of the network is average degree divided by $N - 1$, where N is number of nodes. It is relative fraction of nodes in the network.

In the case of regular networks, such as grids, each node has an equal degree, meaning that nodes in the network have similar roles. In the general case, the networks have more complicated structure. If degree sequence is skewed, we are able to identify nodes with high degree, hubs. Removing hubs may partition a connected network into several components. Finally, if we are able to test isomorphism between two graphs, the starting point would be to compare their degree sequences are the same. If they are not, then they can not be isomorphic.

To calculate the degree distribution we can consider the fraction of k degree nodes N_k, $p(k) = N_k/N$. It is the probability , $P(k)$, that randomly chosen node has degree k. Similarly, we can order nodes according to their degree and plot the node degree.

If the nodes of the graph are statistically independent, the degree distribution completely determines the properties of a network. If degree distribution is present $P(k)$, and the total number of nodes is fixed to N. We can construct a graph with random connections. Label N vertices. To node j of the graph ascribe degrees k_j, taken from the distribution $P(k)$. Connect at random ends of pairs of distinct quills belonging to distinct vertices.

- The Poisson distribution. The degree distribution in random network, where all nodes have the same connecting probability, follows Poisson distribution $P(k) = \frac{(Np)^k e^{-Np}}{k!}$, where k is the mean degree distribution.
- Exponential distribution. $P(k) = e^{-k/k}$. This is degree distribution of the growing random graph. Even for infinite networks all moments of distributions are finite, and have natural scale of the order of average degree.
- In real networks degree distribution follows a power law. $P(k) = k^{-\gamma}$, where γ is exponent of the distribution. In this distribution there is no natural scale, so they are called scale-free networks. In infinite networks all higher moments diverge. If the average degree of scale-free networks is finite, than γ exponent should be greater than 2. Therefore, real networks have a scale-free structure with the emergence of the hubs [**?**]. In finite size networks, fat-tailed degree distributions have natural cut-offs.

When plotting the degree distribution, it is common to use scaling of the axis. As explained, we rank the nodes according their degree and plot the node degree of each kth node. In the first example, we used linear scale for both axes. As many nodes have low degree, it is more useful to use logarithmic scale. On the logarithmic scale the distance between two points is proportional to the logarithm of distance between two points, meaning that distance on logarithm scale is same between points 10 and 100, and 100 and 1000. Now it is more easily notices that data-points follow straight line, meaning that degree distribution is some kind of exponential function.

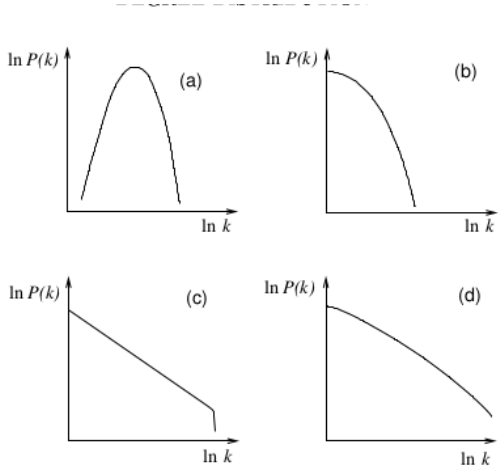

Figure 1.6: Distributions, ovde hocemo 3 distribucije+ linearna vs. logarithm scala, takodje pokazati kako izgleda frequency order

1.4.2 Degree correlations

Correlation is defined through a correlation coefficient r. If x and y are two stochastic variables, for which we have a series of observation pairs $(x_1, y_1), (x_2, y_2), ...(x_n, y_n)$. The correlation coefficient $r(x, y)$ between x and y is defined as:

$$r(x,y) = \frac{\frac{1}{n}\sum_{i=1}^{n}((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1.2}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, is the average over variable x.

Taking the definition of correlation coefficient we can define it for vertex degrees. For simple graph G with vertex set $V(G) = \{v_1, ..v_n\}$, $\boldsymbol{A}[i,j] = 1$ if there is a link between nodes v_i and v_j. If G is a simple graph with adjacency matrix $\boldsymbol{A}$ and degree sequence $\boldsymbol{d} = [d_1, ..., d_n]$

$$r_{deg}(G) = \frac{\sum_{i=1}^{n}\sum_{i=1+1}^{n}((d_i - \bar{d})(d_i - \bar{d})\boldsymbol{A}[i,j])}{\sum_{i=1}^{n}(d_i - \bar{d})^2} \tag{1.3}$$

Using adjacency matrix, allow us to calculate the correlations between neighboring nodes. If two nodes are not connected $A[i,j] = 0$, the degree correlation between them does not have contribution to the r.

The **degree-degree correlations** in the network are measured by **assortativity**. If correlations are positive, networks are assortative; there is a tendency that connections exist between similar degree nodes. The negative correlations indicate that large degree nodes have preference to connect nodes with small degree; dissasortative networks. The average first neighbor degree k_{nn} can be calculated as $k_{nn} = \sum_{k'} k' P(k'|k)$. The P is conditional probability that an edge of degree k points to node with degree k. The norm is $\sum_{k'} P(k'|k) = 1$, and detailed balance conditions **[?]**, $kP(k'|k)P(k) = k'P(k|k')P(k')$ **[?]**. If the node degrees are uncorrelated, k_{nn} does not depend on the degree, otherwise increasing/decreasing function indicates on positive/negative correlations in the network.

The Newman defined the assortativity index r in slightly different way:

$$r = \sum_{kl} kl(e_{kl} - q_l q_k)/\sigma_q^2 \tag{1.4}$$

where e_{kl} is the probability that randomly selected link connect nodes with degrees k and l, q_k is probability that randomly choosen node is connected to node k and equals $q_k = kp_k/\langle k \rangle$, while σ_q is variance of the distribution q_k.

1.4.3 Distances

Between two nodes in the network, we can define different paths, but the most important one are the shortest paths. The distance between two nodes $d(i, j)$ is defined as the length of shortest path between two nodes. In the case of weighted networks, is defined such that the path has minimal weight, and the length of such path does not have to be minimal. Distances on the network, again can give us insight how to networks are similar, and to give indication of relative importance of the node in the network.

The eccentricity of the vertex, gives us how far the farthest vertex is positioned in the network. The radius is minimum over all eccentricity values, while the diameter defines the largest distance between nodes in the network. These definitions apply to directed and undirected graphs.

The diameter gives us useful information, it may not be powerful enough to discriminate among graphs. An important metric is to consider the distribution of path lengths. The average distance between nodes is useful for network description.

If G is connected graph with vertex set V and $\bar{d}(u)$ is average length of the shortest paths from node u, to any other node v in network G.

$$\bar{d}(u) = \frac{1}{|V| - 1} \sum_{v \in V, v \notin u} d(u, v) \tag{1.5}$$

From there we can define the average path length as mean value over $\bar{d}(u)$.

$$\bar{d}(G) = \frac{1}{|V|} \sum_{u \in V} \bar{d}(u) \tag{1.6}$$

while the characteristic path length of G is median over all $\bar{d}(u)$.

1.4.4 D-measure

For each node i we can define the distribution of the shortest paths between node i and all others nodes in the network, $P_i = \{p_i(j)\}$, where $p_i(j)$ is percent of nodes at distance j from node i. The connectivity patterns can efficiently describe difference between two networks. To specify how much G and $G^{'}$ are similar we use D-measure **[?]**

$$D(G, G^{'}) = \omega \left| \sqrt{\frac{J(P_1, ..P_N)}{log(d)}} - \sqrt{\frac{J(P_1^{'}, ..P_N^{'})}{log(d^{'})}} \right| + (1 - \omega) \sqrt{\frac{J(\mu_G, \mu_{G^{'}})}{log 2}} \tag{1.7}$$

D-measure calculates Jensen-Shannon divergence between N shortest path distributions,

$$J(P_1, .., P_N)) = \sum_{i,j} p_i(j) log(\frac{p_i(j)}{\mu_j}) \tag{1.8}$$

where $\mu_j = (\sum_{i=1}^{N} p_i(j))/N$ is mean shortest path distribution.

The first term in equation 1.7 compares local differences between two networks, and Jensen-Shannon divergence between N shortest path distributions $J(P_1, ..., P_N)$ is normed with network diameter $d(G)$. The second part determines global differences, computing $J(\mu_G, \mu_{G'})$ between mean shortest path distributions. The D-measure ranges from 0 to 1. The lower D-measure is, networks are more similar and for D-measure $D = 0$, structures are isomorphic.

1.4.5 Clustering coefficient

The **clustering coefficient** is a measure describing the neighbourhood's structure. In networks exist tendency to form triangles or clusters. This is common in friendship networks where two friends of one person have a high probability of being friends. The clustering can be measured by computing the number of links between neighbours of one node,

$$c_i = 2e_i / (k_i(k_i - 1)) \tag{1.9}$$

Averaging it over all network nodes, we can calculate the mean clustering coefficient. It ranges from $\langle c \rangle = 0$ where connections between neighbouring nodes do not exist, network has the structure of three. On the other hand, $\langle c \rangle = 1$ indicates a fully connected network.

Newman proposed the alternative definition for the clustering coefficient based on the number of triples and triangles in a graph. A triangle at node v is complete subgraph with 3 nodes, including v. A triple on the node v is a subgraph of exactly three nodes and two edges, where v is incident with two edges. The network transitivity is defined as the ratio of number of triangles in the network over the number of triples. The network transitivity is seen as global clustering, as it considers the whole network.

1.4.6 Real-world networks

Real-world networks share similar properties. The mean distance between nodes is smaller than the number of nodes in the network $l << N$, called small-world phenomena. This cause the fast spread of information or even diseases in the complex systems. In small-world networks number of vertices grow exponentially with distance; thus l increase as $log(n)$ or slower. Logarithmic scaling can be proved from various network models; also, it is observed in real-world complex systems. The clustering coefficient in real-world networks is usually high. Real-world networks have one important feature; power-law degree distribution; such networks are called scale-free networks.

1.5 Network models

1.5.1 Random network model

The random graph model was introduced by mathematicians Paul Erdős and Alfred Rényi in 1959. In this model, connections between nodes are chosen randomly, and every link has the same probability of existing. The graph is characterized only by a number of the nodes N and the linking probability p, so Erdős-Rényi graph is written as $G(n, p)$.

The creation of ER random network consists of the following steps:

- we start with N isolated nodes

- between each $N(N-1)/2$ pair of nodes we create link with probability p; sampling random number $r \in (0,1)$, we create link if $r \leq p$

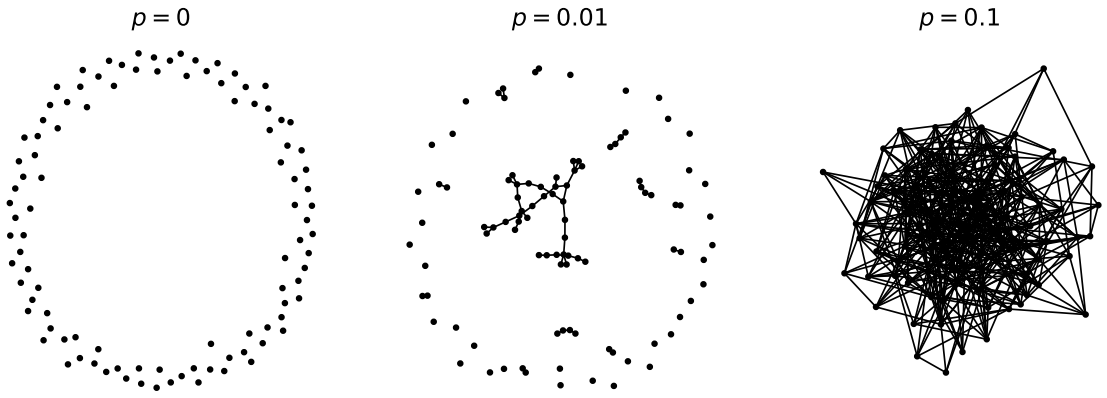

Figure 1.7: ER graph with $N = 100$ nodes and different linking probabilities p.

We should note that this process is stochastic. The networks $G(N,p)$ with the same parameters do not need to have the same structure; i.e. they differ in the number of links. Therefore, the single random graph is only one graph from all the possible realizations in the statistical ensemble.

Two simple quantities that could be estimated are the average number of links and the average degree. For complete graph with N nodes, number of edges is $N(N-1)/2$. As the probability of drawing every edge is p, the **average number of links** is simply given as

$$\langle L \rangle = \frac{N(N-1)}{2} p \tag{1.10}$$

From there, we conclude that the network's density is equal to probability p. The **average degree** is approximated as: $\langle k \rangle = 2\langle L \rangle / N$, leading to:

$$\langle k \rangle = (N-1)p \tag{1.11}$$

The **degree distribution** of ER random graph follows the binomial distribution.

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{1.12}$$

The probability that the node has degree k is given with the second term p^k, while the probability that other N-1-k links are not created is given with the third part of the equation. Finally, there are $\binom{N-1}{k}$ combinations for one node, to have k links from $N-1$ possible links.

The binomial distribution describes very well small networks. For larger networks, we find that they are sparse and that the average degree is much smaller than a number of nodes $\langle k \rangle << N$. In this limit, binomial distribution becomes the Poisson, which now depends only on one parameter $\langle k \rangle$

$$p(k) = \frac{1}{k!} e^{-\langle k \rangle} \langle k \rangle^k \tag{1.13}$$

The random graph has a very small **average path length**, it is given as $\langle l \rangle = \frac{lnN}{ln(pN)}$ that is characteristic of many large networks. The clustering coefficient is proportional to linking probability, $\langle C \rangle = p$, so in large random networks, we find a small clustering coefficient,

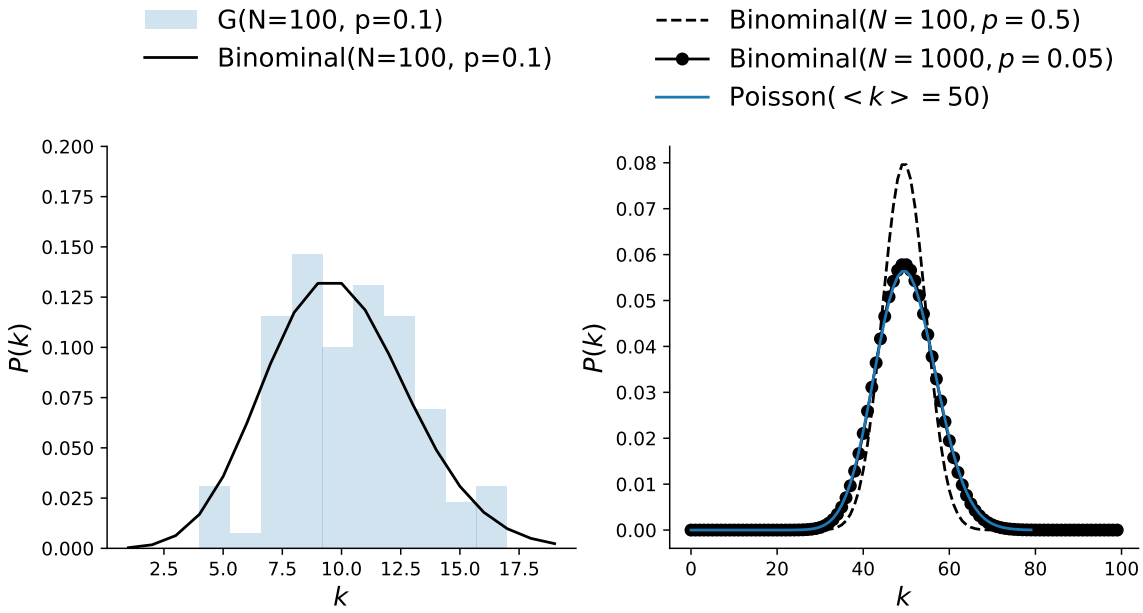

Figure 1.8: Degree distribution of ER graph. Degree distribution of small networks follow binominal. Larger networks are better approximated with Poison distribution, and degree distribution for fixed average degree $< k >$ becomes independent of the network size.

contrary to real-world networks.

The figure 1.7 shows how the network becomes more connected by increasing the linking probability p. When $p = 0$, all nodes are disconnected. In the other limit, $p = 1$, the network is fully connected. Between those two probabilities exists critical probability, where the giant component appears. The giant component is a sub-graph, which size is proportional to the network size. In other words, the network does not have disconnected components. Such change in the network is a phase transition in network connectivity and is related to percolation theory.

The phase transition occurs when average degree is $\langle k \rangle = 1$, which gives us: $p_c = \frac{1}{N-1}$, meaning that all nodes have degree larger than 1. When the $\langle k \rangle < 1$, the network is in the sub-critical regime where all components are small. In the critical regime, the size of the giant component is proportional to the $N^{2/3}$. In the supercritical regime, $\langle k \rangle > 1$, the probability of a giant component appearing is 1.

1.5.2 Small-world networks

Inspired by the idea that real-world networks are highly clustered and the average distance is small, Watts and Strogatz proposed the "small-world" model. The model starts from the regular lattice, and with rewiring links, the network starts to resemble small-world property. The procedure is the following:

- At the beginning, nodes are placed on the ring lattice, and each node is connected to $k/2$ first neighbours on the left and the right side. Initially, the clustering coefficient is high, $c = 3/4$.

- For each link in the network, with probability p, we choose a random node to rewire the link. This makes long-distance nodes connect, decreasing the network's average path length.

The model interpolates between the regular graph when the probability is $p = 0$ and the random graph with $p = 1$ when all links are randomly rewired. Short distances and high clustering are present in the network for the critical probabilities.

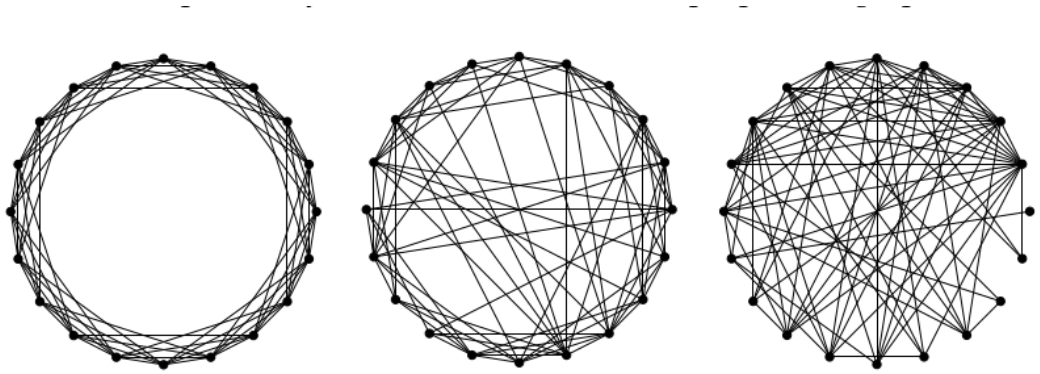

Figure 1.9: Watts and Strogatz graph model creation

Even though the small-world network model lacks the power-law degree distribution found in the real-world networks, it is an important model that motivated the research on random graphs.

1.5.3 Barabási-Albert model

The ER random graph model and WS small-world model are static models, where the number of nodes is fixed. It is one of the reasons why they can not fully explain the properties of real systems. The size of real systems does not remain constant; real networks grow. For the network, the growth means that at each time step, new nodes are added to the network. The simplest model that produces the scale-free networks is Barabasi-Albert model.

- The model starts from the small number, n_0 randomly connected nodes, with m_0 links.
- At each time step, new node with m links joins to the network. New node creates links with the nodes already present in the network, following the linking rules; i this case rules of preferential attachment.

The preferential attachment is important ingredient for generating system with scale-free properties. In the real-system the linking between nodes is not random process, there exists the preference toward specific types of nodes. For example the popular web-pages can easily get more visits or it is common that already popular papers will get more citations. This effect is also called rich-get-richer or preferential attachment.

The simplest formulation of the preferential attachment model is that new nodes tend to connect with high degree nodes. The linking probability Π is then proportional to node degree k:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{1.14}$$

As at each time step one node arrive, we can estimate the number of nodes at the time step t, $N(t) = n_0 + t$, with links $L(t) = m_0 + mt$.

First we can calculate the evolution of network degree in time.

$$\frac{dk_i}{dt} = m\Pi(k_i) = m\frac{k_i}{\sum_j k_j} = m\frac{k_i}{m_0 + 2mt} \tag{1.15}$$

Note that new node, that arrived at time point t_i has degree m, as it links to m old nodes. Solving the equation we get that at $t > t_i$, has degree that grows as square root of time, also it shows that younger nodes easily acquire larger degree.

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\frac{1}{2}} \tag{1.16}$$

Degree distribution follows power-law, and for large k is aproximated with $P(k) = k^{-\gamma}$, such that $\gamma = 3$. More precisely, the degree distribution has form:

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \tag{1.17}$$

For large k it is exactly power-law. It is also independent of the time and size of the system, meaning the emergence of stationary scale-free state. Distributions do not depend on the N. If we vary m the slope of distributions is the same, but they are parallel. After re-scaling $p(k)/m^2$, they fall on the same line.

As network grows nodes with larger degree becomes bigger, so we end up with few nodes with many links, called hubs. The **network diameter**, represents the maximum distance in network, $d \sim \frac{lnN}{lnlnN}$. The diameter grows slower than lnN, making the distances in BA model smaller than in random graph. The difference is found for large N. Knowing that BA network has hubs, that shorten the path between less connected nodes. Also, if hubs are removed from the network, network easily partition in several components, loosing its properties.

The **clustering coefficient** of the BA model follows $C \sim \frac{lnN^2}{N}$. It is different from clustering found in random networks, and BA networks are in general more clustered.

The combination of the growth and preferential attachment linking is crucial for getting scale free networks. For example, eliminating the preferential attachment; in growing network with random linking, degree distribution is stationary, but it follows exponential. In contrast, the absence of growth leads to the non-stationary degree distribution. When number of nodes is fixed, while the network grows only in number of links, such that randomly chosen node i connects to node j according to probability Π. At the beginning, the degree distribution follows the power-law, same as in BA model. As more links are added to the network, the distribution changes it's shape, first the peak appears, while at the end network becomes complete graph, where all nodes have the same degree.

1.5.4 Nonlinear preferential attachment model

In the nonlinear preferential attachment model linking probability also depends on the node degree. The dependence is not linear and has the following a form:

$$\Pi(k_i) = {k_i}^{\beta} \tag{1.18}$$

The probability that newly added node attaches to node i depends on the existing i-th node degree k_i, and the parameter β. When $\beta = 1$, the model is BA model, where degree distribution follows the power-law. When $\beta = 0$, linking probability becomes uniform; i.e. it corresponds to random network model, and degree distribution is Poisson; there is exponential decay.

For $\beta > 1$, the effects of preferential attachment are increased, leading to emergence of super-hubs. The hub-and-spoke network appear in this regime, where almost all nodes are connected to few high-degree nodes.

On the other hand, if $\beta < 1$, the model is in so called sub-linear preferential attachment regime. The linking probability is not random so degree distribution does not follow Poisson; but also the preference toward high degree nodes is too week for having the pure power-law. Instead degree distribution converge to stretched exponential.

1.5.5 Ageing model

To understand how aging can impact the network structure we look into probability dependent on two parameters, nodes degree k and age of node i at the time point t $\tau_i = (t - t_i)$, where t_i is the time when node i is added to the network.

$$\Pi_i(t) \sim k_i \tau_i^{\alpha} \tag{1.19}$$

The parameter α controls the linking probability dependence on the nodes' age if $\alpha = 0$, the ageing of nodes is disregarded.

If $\alpha > 0$ is positive, the older nodes are more likely to create connections. In this regime, the preferential attachment stays present, and the high-degree and older nodes are preferred. For very high α, each node is connected to the oldest node in the network. The scale-free properties are present; the power-law exponent γ deviates from $\gamma = 3$. It is found that γ ranges between 2 and 3.

When α is negative, ageing overcomes the role of preferential attachment, and scale-free properties are lost. For significant negative α network becomes a chain; the youngest nodes are those who get connected.

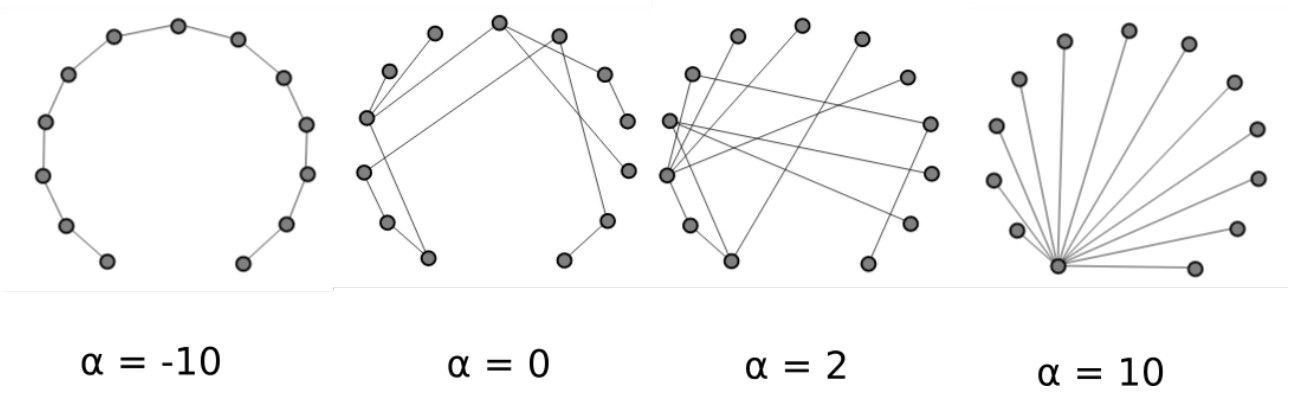

Figure 1.10: Aging model

In the general ageing model, the non-linearity on the node degree is introduced, so this model has two tunable parameters α and β. The probability that a link is created between the new node and the existing node is defined as

$$\Pi_i(t) \sim k_i(t)^{\beta} \tau_i^{\alpha} \tag{1.20}$$

As before, depending on model parameters network evolves to different structures [**?**].

- For example if we fix $\beta = 1$ and $\alpha = 0$ generated networks are scale-free; degree distribution is $P(k) \sim k^{-\gamma}$ with $\gamma = 3$.

- In the case of nonlinear preferential attachment $\beta \neq 1$ and $\alpha = 0$ scale-free properties disappear.
- Scale-free property can be produced along the critical line $\beta(\alpha^*)$ in the $\alpha - \beta$ phase diagram, see Figure 1.11.
- For $\alpha > \alpha^*$ networks have **gel-like small world** behavior.
- For $\alpha < \alpha^*$ and near critical line $\beta(\alpha^*)$ degree distribution has **stretched exponential** shape

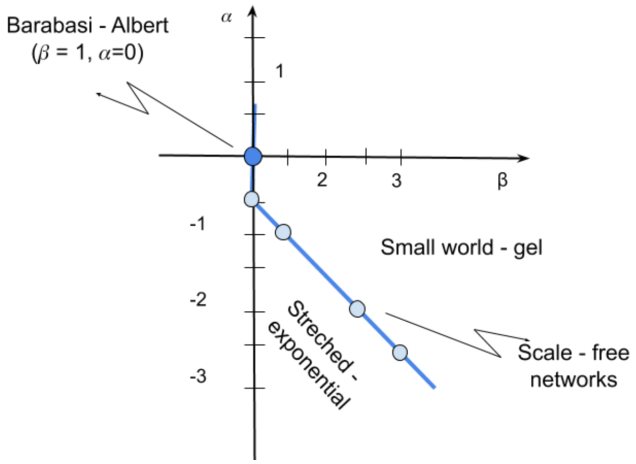

Figure 1.11: Phase diagram of aging network model

1.6 The properties of probability distributions

1.6.1 Poisson distribution

Poisson distribution is exponentially bounded distribution. The property of these distributions is that they decay exponentialy or faster for high x. The largest expected value grows as $logx$, meaning that outliers representing the high x values are rate. In the network science the most common distribution is Poisson distribution, while outside the network science the most common distribution is Gaussian distribution.

On the other hand, fat tailed distributions, long tailed or heavy tailed distributions refer to distributions whose decay for large x is slower than exponential. In this distributions events with large x value are rare but possible.

1.6.2 Power law

The power-law distribution is defined as

$$p(k) = Ck^{-\gamma} \tag{1.21}$$

where parameter γ is an exponent of the power-law distribution while the C is the normalizing constant.

The distribution can take both discrete and continuous values and it is defined for positive values $k > 0$ so there is lower bound to the power-law function k_{min} and it . For the discrete case $C = 1/\zeta(\gamma, k_{min})$, while in the continuous case $C = (\gamma - 1)k_{min}^{\gamma-1}$.

The likelihood function for continuous data set is defined as

$$l(k) = \prod p(k) = \prod \frac{\gamma - 1}{k_{min}} (\frac{k_i}{k_{min}})^{-\gamma} \tag{1.22}$$

Minimizing the loglikelihood function, exponent is calculated as $\gamma = 1 + n[\sum ln \frac{k_i}{k_{min}}]^{-1}$.

For discrete distribution the log likelihood is $log(l) = ln \prod \frac{k_i^{-\gamma}}{\zeta(\gamma, k_{min})}$. For this equation does not exist the analytical solution, but the equation can be numericaly optimized.

The power-law distribution is called scale-free distribution. This is distribution that is the same on all scales.

For example if we take power law distribution:

$$\frac{p(x)}{p(2x)} = \frac{Ax^{-\alpha}}{A(2x)^{-\alpha}} = 2^{\alpha}$$

This ratio is constant and does not depend on the x. If we take any other distribution, we'll find that this criteria is not satisfied.

In the general form, scale free function is defined as:

$$p(bx) = g(b)p(x)$$

Solving this equation we get $p(x) = p(1)x^{-\alpha}$, where $\alpha = -p(1)/p^{'}(1)$, concluding that if function is self-similar, it has to be a power-law.

1.6.3 Lognormal distribution

The variable x has the lognormal distribution if the random variable $y = ln(x)$ is distributed as normal distribution.

$$f(y) = \frac{1}{2\pi\sigma} e^{-(y-\mu)^2/2\sigma^2} \tag{1.23}$$

where μ is mean, and σ is standard deviation. The density distribution of the lognormal distribution is defined as

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(log(x)-\mu)^2/2\sigma^2} \tag{1.24}$$

The lognormal distribution has finite mean $e^{\mu+1/2\sigma^2}$, and the variance $e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$. **[?]**.

Despite the finite moments, the lognormal distribution can be similar to power-law distribution. If the variance of lognormal distribution is large then the probability function on log-log plot appears linear for large range of values.

The normal distribution has property that the sum of two independent normal random variables is normal random variable with mean $\mu_1 + \mu_2$, and variance $\sigma_1^2 + \sigma_2^2$. It follows that two log-normally distributed random variables also have a lognormal distribution. **[?]**.

1.6.4 Power law with exponential cutoff

The density function has the form

$$p(x) = Cx^{-\gamma} e^{-\lambda x} \tag{1.25}$$

where $x > 0$ and $\gamma > 0$. This function combines the power-law term and exponential term responsible for exponentially bounded tail. Taking the logarithm $ln(p(x)) = lnC - \gamma lnx - \lambda x$, when $x << 1/\lambda$ the second term dominates, so distribution follows te power-law, with exponent γ. Otherwise, the λx term domintes, resulting in exponential cutoff for high x.

1.6.5 Streched exponential

the stretched exponential

$$p(x) = cx^{\beta-1}e^{-(\lambda x)^{\beta}} \tag{1.26}$$

the parameter β is streching exponent determining the properties of the function $p(x)$. For $\beta = 1$, the function is exponential. For $\beta < 1$ it is hard to distinguish the distribution from power-law. For $\beta > 1$ we have a compressed exponential function, so x vary in the narrow range.

Whenever we encounter fat-tailed distributions, there is discussions asking which distribution offers the best fit to the data. In many systems empirical data is not sufficient to distinguish between distributions.

1.6.6 Plotting the fat-tailed distributions

With plotting the distribution of the power-law data on the double logarithmic scale we should obtain the straight line, as $ln(p(k)) = \gamma ln(k) + c$. The tail of the distribution is noisy. As the size of the bins is constant, the density of the bins for large distribution values becomes large. To avoid the fluctuations in the tail, we can use logarithmic binning. The noise is reduced by dividing the x axis into n bins $b_n = c^n$, so the following bin is wider than before. For the base c we can choose any value $c > 1$. Similarly, the binning can take the following form $b_n = k_0 \exp(cn)$, where k_0 is the minimum data point, while the c is the arbitrary base. All data points between values $[b_n, b_{n+1})$ are represented with one point $p(k_n) = N_n/b_n$, where N_n is number of nodes found in the bin b_n and $k_n = \sum_i k_i/N_n$ is average degree of the nodes in the bin b_n. By this, averaging over bins in the tail, which now have more sampled points, reduces the statistical errors.

Instead of plotting the probability distribution (i.e. creating histogram), it is possible to calculate the cumulative distribution. $P(k) = \int_x^{\infty} p(x')dx' = x^{-(\gamma-1)}$. It is also the power-law distribution with exponent $\gamma - 1$. Note that for cumulative distribution it is not necessary to use log-binning.

1.6.7 The goodness-of-fit

Minimizing the loglikelihood of function for given data, allow us to analitically or even numerically fit distributions and estimate fit parameters. Still, it does not tell us how good the fit is.

The figure shows the distributions of three small datasets, drawn from power-law with $\gamma = 2.5$, lognormal $\mu = 0.3, \sigma = 2$ and exponential with $\lambda = 0.125$. The distributions look as straight line on the log-log plot, we could try to fit them to the power law distribution and obviosly, some model parameters could be estimated. It is not straightforward to say weather particular data really follows given distribution. Even if data follow powe-law, their observed distributions are not likely to exactly follow power-lawl there are some deviations because of the random nature of sampling procedure.

The basic approach is to sample many synthetic data sets from a tru power-law and measure how they fluctuate from power-law form and compare results on empirical data. If empirical data are much further from power-law than syntetic one, the power-law is not plausible fit to the data.

A standard aproach to answering this kind of questions is to use goodness-of-fit test which generate the p-value that quantifies the plausibility of the hypothesis. Measuring the

distance between distributions of empirical data and the model. p-value is defined to be fraction of synthetic distances that are larger than the empirical distance.

For measuring the distance between distributions we can use the Kolmogorov-Smirnov statistics, but in general other goodness of fit measures could be used. The Kolmogorov Smirnov statistics is the maximum distance between the CDF of the data and the fitted model.

$$D = max|S(x) - P(x)| \tag{1.27}$$

First we fit empirical data to get model parameters, and calculate the KS statistics of this fit. Then, large number of syntetic data sets, are generated with model optimized model parameters. Then each syntetic dataset is fitted, and KS statistics is obtained relative to its own model. Then simply we count the fraction when the KS statistics of syntetic distributions is larger than in empirical data, that represents the p-value.

If $p < 0.1$ then we reject the hypothesis that this distribution describes the empirical data, otherwise the model can not be rejected. Failing to reject the hypothesis does not mean the model is correct distribution for the data. There might be other distributions that fit the data equaly good, or even better. For having accurate p-value we need large sample. For small number of syntetic distributions it is possible to have high p-value, even if the distribution is wrong model to the data.

Using goodness of fit it is possible to calculate p-value for any distribution. For example comparing p-value of power-law fit to the p-value of trunctated power-law fit, we can conclude which one is better fit to data. If p-value for powe law is high, while for alterantive distribution it is low, we can conclude that power-law is more probable fit.

The another method called the likelihood ratio test allows us to directly compare two distributions. The distribution with higher likelihood under epmpirical data is better fit. We can calculate the likelihood ratio, or it is easier to obtain the logarithm of likelihood ratio, because its sign determine which distribution is better fit. For given two distributions $p_1(x)$and $p_2(x)$.

The likelihoods are defiend as $L_1 = \prod_{i=1}^{n} p_1(x)$ and $L_2 = \prod_{i=1}^{n} p_2(x)$, or the ratio of likelihoods as $R = \frac{L_1}{L_2} = \prod_{i=1}^{n} \frac{p_1(x)}{p_2(x)}$

Taking the logarithm, we obtain the loglikelihood ratio

$$\mathcal{R} = \sum_{i=1}^{n} [lnp_1(x) - lnp_2(x)] \tag{1.28}$$

As data x_i are independent, by central limit theorem their sum $\mathcal{R}$ becomes normally distributed, with expected variance σ^2. We can approximate the variance as

$$\sigma^2 = \frac{1}{n} \sum_{1}^{n} [(l_i - l_i) - (< l >^{(1)} - < l >^{(2)})]$$

When $R > 0$ the first distribution is better fit to data and when $R < 0$, the other one should be chosen. When $R = 0$, it is not possible to distinguish between two distributions.

The sign of R is not enough criteria to conclude which distribution is better fit. It is a random variable subject to statistical fluctuations. We need loglikelihood ratio that is sufficently positive or negative, and to be sure that its sign is not result of fluctuations.

If we are suspcted thaat the true expectation value of the loglikelihood ratio is zero, meaning that observed sign of $\mathcal{R}$ is simply product of fluctuations and an not be trusted. The probability that measured log likelihood ratio has magnitude as large or larger than observed value R is given as

$$p = \frac{1}{\sqrt{2\pi n\sigma^2}} \int_{-\infty}^{-|\mathcal{R}|} e^{-x^2/2n\sigma^2} dx + \int_{|\mathcal{R}|}^{\infty} e^{-x^2/2n\sigma^2} dx \tag{1.29}$$

Here we use standard two tail hypothesis test, assuming that the null hypothesis is that $R = 0$. If p-value is larger than threshold, the sign of R is not reliable, and the test does not favor any distribution. If p is small, $p < 0.1$ than it is unlikely that observed sign is obtained by chance, so we reject the null hypothesis that $R = 0$.

1.7 Multiplicative processes

Using the multiplicative processes we can generate the power-law but also the log-normal distribution.

The lognormal distribution is generated by processes that economist Gibrat called the law of proportionate effect.

If we start from the organism of size S_0. At each time step, the organism may grow or shrink, according to random variable ϵ,

$$S_t = \epsilon_t S_{t-1}$$

When the state of the system at time t is proportional to the state of the system at previous time step, we have the multiplicative process. The ϵ is proportionality constant that can change over time. The state of the system at time step t is determined by the product of the various ϵ_t and the initial size S_0

$$S_t = \epsilon_t S_{t-1} = \epsilon_t \epsilon_{t-1} ... \epsilon_2 \epsilon_1 S_0$$

The S_t is drawn from the log-normal distribution. If ϵ_t is drawn from the lognormal distribution, then S_t also follows lognormal, as the product of lognormal distributions is again lognormal. Still, the distribution of the ϵ does not determine the distribution of the $S_j t$.

Taking the logarithm of the equation:

$$ln(S_t) = ln(S_0) + \sum_{i=0}^{t} ln(\epsilon_i)$$

The sum of the logarithms of the ϵ_t, according to the Central Limit Theorem follows the normal distribution. The CLT states that the sum of intetically distributed random variables with finite variance converge to the normal distribution. If $ln(S_t)$ is normally distributed, then S_t follows the log-normal distribution.

Using the multiplicative processes it is possible to generate the lognormal distribution. Also similar process may lead to the power-law distribution.

In the Champernowne model, where individuals are divided into classes according to their income. The minimum income is m. People between incomes m and γm are in the first class, in the second class are people with income between γm and $\gamma^2 m$. The individuals can change their class, so it is described as the multiplicative process, but with threshold, as income can not be lower than m. For example, if we fix $\gamma = 2$, and consider that with probability $p_{i,i-1} = 2/3$ the change is from higher to lower class, and with probability $p_{i,i+1} = 1/3$ individual goes to higher class. This process leads to the power-law distribution.

1.8 Fractal analysis

Approach to study complex systems is detecting time-series of selected variables. Some systems are characterised by periodic ot nearly periodic behaviour. In complex systems this periodic behaviour is not limited to one or two characteristic frequencies. They extend over wide spectrum and fluctuations on many time scales as well as well as broad distributions. In these cases dynamics of the system is characterized by scaling laws, which are valid over a wide range of time scales or frequencies. If dynamic of the system can be described with one scaling exponent system is monofractal, otherwise we deal with multifraactal time-series.

Rescaling of time t by a factor a may require rescaling of the time-series values $x(t)$ by factor a^H, to get the self-similaraity. In this case it is:

$$x(t) = a^H x(at)$$

The Hurst exponent characterize the type of self-affinity.

Many records do not exhibit a simple mono fractal scaling behaviour. The scaling behaviour may be more complicated , and different scaling exponents can be found for many interwoven fractal subsets of the time series. In this case the multifractal analysis must be applied.

Two general types of multifractality exist. The multifractality due to a broad probability distribution for values of the time series, the multifractality can not be destroyed. Multifractality due to different long-term correlations of the small and large fluctuations. In this case the probability density function of the values can be regular distribution with finite moments, and the corresponding shuffled series will exhibit non-multifractal scaling as correlations are destroyed with shuffling procedure. If both kinds of multifractality are present, the shuffled series will show weaker multifractality than the original series. Multifractal analysis will reveal higher order correlations, multifractal scaling can be observed if the scaling behaviour of small and large fluctuations is different. Extreme events might be more or less correlated than typical events.

Long and Short-term correlations The time-series are presistent such that a large value is usually followed by large values and small values. Considering the increments $\delta x_i = x_i - x_{i-1}$, of self-affine series $i = 1, .., N$, with N values measured equidistant in time, so δx_i can be either persistent, independent or anti-persistent. For the random walk with $H = 0.5$ the increments are independent of each other. Persistent and anti-persistent increments, where a positive increment is likely to be followed by another positive or negative increment.

For stationary data with constant mean and standard deviation the auto-covariance function can determine the degree of persistence.

$$C(s) = \langle \Delta x_i \Delta x_{i+s} \rangle = \frac{1}{N-s} \sum_{i=1}^{N-s} \Delta x_i \Delta x_{i+s}$$

If the data are uncorrelated the $C(s) = 0$. Short range correlations are described by $C(s)$ declining exponentially

$$C(s) = exp(-s/t_c)$$

such behaviour is typical for increments generated by an auto-regresive process

$$\Delta x_i = c \Delta x_{i-1} + \epsilon_i$$

with random uncorrelated offsets ϵ_i and $c = exp(-1/t_c)$.

For long-range correlations $\int C(s)$ diverges in the limit for long series. In practive this means that caracteristic time can not be defined because it increaseas with N. Contrary to short-range correlations, the correlation function decline as power-law

$$C(s) = s^{-\gamma}$$

This type of behavior can be modeled by Fourier filtering techniques. Long-term correlated, behavior of Δx_i leads to self-affine scalling beahviout characterized by $H = 1 - \gamma/2$.

A direct calculation of the $C(s)$ is difficult due to present noise in the data and nonstationarity. Non-stationarities make the definition of $C(s)$ probalematic, because the average is not well defined, also $C(s)$ fluctuates around zero on large scales s, so it is not possible to obtain the correct correlation exponent γ.

Hurst's rescaled-range analysis The method called rescaled range analysis R/S was proposed by the Hurst. It begins with spliting the time series x_i into non overlaping segments ν of the size s, having $N_s = int(N/s)$ segments. Then is calculated the profile in each segment.

$$Y_\nu(j) = \sum_{i=1}^{j} (x_{\nu s+i} - \langle x_{\nu s+i} \rangle_s)$$

Substracting the averages, constant trends in the data are eliminated. Finally the differences between minimum and maximum value and the standardan deviation in each segment are calculated as:

$$R_\nu(s) = max Y_\nu(j) - min Y_\nu(j)$$

$$S_\nu(s) = \sqrt{\frac{1}{s} \sum Y_\nu^2(j)}$$

Finally, the rescaled range is averged over all segments to obtain the fluctuation function F(s).

$$F_{RS}(s) = \frac{1}{N_s} \sum \frac{R_\nu(s)}{S_\nu(s)} \sim s^H$$

the H is Hurst exponent introduced in the first equation. The values of H that can be obtained by Hurst rescaled analasys are $< 0 H < 2$. Values $H < 1/2$ indicate long-term anticorrelated data, $H > 1/2$ indicated long-term positively correlated data. For power-lae correlations decaying faster than $1/s$, $H = 1/2$, like for uncorrelated data.

On the other hand the standard fluctuation analysis is based on the random walk theory. For time series with zero mean, we consider the global profile, the cumulative sum:

$$Y(j) = \sum x_i$$

, and then study how fluctuations of the profile, in a given time window of size s increase with s.

We first divide each record of N elements into Ns non-overlaping segments of the size s, and another Ns non-overlaping segments starting from the end. Then we calculate the fluctuations in the each segment. In the standard FA we get the fluctuations just from the values of the profile at both endpoints of each segment.

$$F_{FA}^2(\nu, s) = [Y(\nu s) - Y((\nu + 1)s)]^2$$

Then we can average F^2 over all subsequences to obtain the mean fluctuation

$$F_2(s) = [\frac{1}{2N_s}\sum F^2(\nu, s)]^{1/2} \sim s^H$$

For the relevant case of long-term correlations where $C(s)$ follows the power-law behaviour, $F_2(s)$ also increases by power-law.

The fluctuation exponent is indentical with Hurst exponent for monofractal data.

1.8.1 Detrended fluctuation analysis

DFA was introduced by Peng et al. and represents an important method for describing the non-stationary time series. As before methods it is also base on random walk theory, and the FA analysis is special case with linear detrending.

Like in FA method first one calculates the global profile of the time series. DFA deaks with monotonous trends in a detrending procedure. This is done by establishing a polynomial trend within each segment by least-square fitting and subtracting this trend from the original profile, detrending.

$$Y_s(j) = Y(j) - y^m_{\nu,s}(j)$$

The degree of the polinomial can be varied in order to eliminate constant m=0, linear m=1, quadratic m-2, or higher order trends of the profile function. The variance of the detrended profile in each segment gives us mean-square fluctuations

$$F^2(\nu s) = \frac{1}{s}\sum Y_s^2(j)$$

Finaly fluctioations over all segments are averaged, to obtain the mean fluctuations $F_2(s)$ as in eq. F2s, and as before, from the scaling of the fluctuating function we can determine the Hurst exponent.

1.8.2 Multifractality of the signals

Multifractal detrended fluctuation analysis (MFDFA) [**?**, **?**] to estimate multifractal Hurst exponent H(q). For given time series $\{x_i\}$ with length N, first we define global profile in the form of cumulative sum, equation 1.30, where where $\langle x \rangle$ represents average of the time series:

$$Y(j) = \sum_{i=0}^{j}(x_i - \langle x \rangle), \quad j = 1, ..., N \tag{1.30}$$

Subtracting the mean of the time series is supposed to eliminate global trends. The profile of the signal Y is divided into $N_s = int(N/s)$ non overlapping segments of length s. If N is not divisible with s the last segment will be shorter. This is handled by doing the same division from the opposite side of time series which gives us $2N_s$ segments. From each segment ν, local trend $p^m_{\nu,s}$ - polynomial of order m - should be eliminated, and the variance $F^2(\nu, s)$ of detrended signal is calculated as in equation 1.31:

$$F^2(\nu, s) = \frac{1}{s}\sum_{j=1}^{s}\left[Y(j) - p^m_{\nu,s}(j)\right]^2 \tag{1.31}$$

Then the q-th order fluctuating function is:

$$F_q(s) = \left\{\frac{1}{2N_s}\sum_{\nu}^{2N_s}\left[F^2(\nu, s)\right]^{\frac{q}{2}}\right\}^{\frac{1}{q}}, q \neq 0$$

The value of $H(0)$, which corresponds to the limit $H(q), q-0$, cannot be determined directly because of the exponent diverge. Instead logarithmic averaging procedure has to be considered.

$$F_0(s) = \exp\left\{\frac{1}{4N_s}\sum_{\nu}^{2N_s} ln\left[F^2(\nu, s)\right]\right\}, q = 0 \tag{1.32}$$

The fluctuating function scales as power-law $F_q(s) \sim s^{H(q)}$ and the analysis of log-log plots $F_q(s)$ gives us an estimate of multifractal Hurst exponent $H(q)$.

For the monofractal time series, H(q) is independent of q, meaning that scaling is indentical for all segments, and averaging fluctuations gives indentical scaling for all values of q. If small and large fluctuations scale differently, there will be dependence of h(q) on q. Positive values of q, segments with large variance are dominant in the Fq(s), so positive q describes segments with large fluctuations. The negative values of q, H(q) describes the scaling of the segments with small fluctuations.

Also, large fluctuations are characterized by smaller scaling exponent.

Multifractal signal has different scaling properties over scales while monofractal is independent of the scale, i.e., H(q) is constant.

1.9 Dynamical reputation model

Any dynamical trust or reputation model has to take into account distinct social and psychological attributes of these phenomena in order to estimate the value of any given trust metric [**?**]. First of all, the dynamics of trust is asymmetric, meaning that trust is easier to lose than to gain. As part of asymmetric dynamics, in order to make trust easier to loose the trust metric has to be sensitive to new experiences (recent activity or the absence of the activity of the agent), while still maintaining nontrivial influence of old behavior. The impact of new experiences has to be independent of the total number of recorded or accumulated past interactions, making high levels of trust easy to lose. Finally, the trust metric has to detect and penalize both the sudden misbehavior and the possibly long term oscillatory behavior which deviates from community norms.

We estimate dynamic reputation of the Stack Exchange users using Dynamic Interaction Based Reputation Model (DIBRM) [**?**]. This model is based on the idea of dynamic reputation, which means that the reputation of users within the community changes continuously through time: it should rapidly decrease when there is no registered activity from the specific user in the community (reputation decay), and it should grow when frequent, constant interactions and contributions to the community are detected. The highest growth of user's reputation is found through bursts of activity followed by short period of inactivity.

In our implementation of the model, we do not distinguish between positive and negative interactions in the Stack Exchange communities. Therefore, we treat any interaction in the community (question, answer or comment) as potentially valuable contribution. In fact, evaluation criteria for Stack Exchange websites going through beta testing, described in SI, do not distinguish between positive and negative interactions. The percentage of negative interactions in the communities we investigated was below 5%, see Table 1 in SI. Filtering positive interactions would also require filtering out comments because they are not rated by the community, and that would eliminate a large portion of direct interactions between the users of a community, which is essential for estimating their reputation.

In DIBRM, reputation value for each user of the community is estimated combining three different factors: 1) *reputation growth* - the cumulative factor which represents the importance of users' activities; 2) *reputation decay* - the forgetting factor which represents the continuous decrease of reputation due to inactivity; *the activity period factor* - measuring the

length of the period of time in which the change of reputation happened. In case of Stack Exchange communities, the forgetting factor has a literal meaning, as we can assume that past contributions provided by a user are being forgotten by active users as their attention is captured by more recent content.

In line with the the basic dichotomy of reputation dynamics, which revolves around the varying influence of past and recent behavior, DIBRM has two components: *cumulative factor* - estimating the contribution of the most recent activities to the overall reputation of the user; *forgetting factor* - estimating the weight of past behavior. Estimating the value of recent behavior starts with the definition of the parameter storing the basic value of a single interaction I_{b_n}. Cumulative factor I_{c_n} then captures the additive effect of recent successive interactions. The reputational contribution I_n of most recent interaction n of any given user is estimated in the following way:

$$I_n = I_{b_n} + I_{c_n} = I_{b_n}(1 + \alpha(1 - \frac{1}{A_n + 1})) \tag{1.33}$$

Here, α is the weight of the cumulative part and A_n is the number of sequential activities. If there is no interaction at t_n, this part of interactions has a value of 0. Important property of this component of dynamic reputation is the notion of sequential activities. Two successive interactions made by a user are considered sequential if the time between those two activities is less or equal to the time parameter t_a which represents the time window of interaction. This time window represents maximum time spent by the user to make a meaningful contribution (post a question or answer or leave a comment).

$$\Delta_n = \frac{t_n - t_{n-1}}{t_a} \tag{1.34}$$

If $\Delta_n < 1$ is less than one the number of sequential activities A_n will increase by one, which means that the user is continuing to communicate frequently. On the other hand, large values Δ_n greatly increase the effect of the forgetting factor. This factor plays a major role in updating the total dynamic reputation of a user in each time step (after every recorded interaction):

$$T_n = T_{n-1}\beta^{\Delta_n} + I_n \tag{1.35}$$

Here, β is the forgetting factor. In our implementation of the model, the trust is updated each day for every user irrespective of their activity status. Therefore, the decay itself is a combination of β and Δ_n: the more days pass without recorded interaction from a specific user, the more their reputation decays. Lower values of beta lead to faster decay of trust as shown on figure **??**.