Budapest University of Technology and Economics
Institute of Mathematics
Department of Stochastics

Towards a Better Understanding of the Characteristics of Fractal Networks

Students' Scientific Conference (TDK)
Report

Enikő Polyák

Supervisors:

Roland Molontay
Research Fellow
MTA-BME Stochastics Research Group

Marcell Nagy
PhD Student
Department of Stochastics
Budapest University of Technology and Economics

2021

Contents

Chapter 1

Introduction

Networks are present in almost every field of life. An immediate example that comes into mind is the Internet itself, but we could also think of human relationships, which can be modeled by networks. The primary goal of the study of networks is to explore and understand their structure, origin and possible evolution. For example, if we aim to efficiently stop or prevent a pandemic, it is important to explore both the biological structure of the virus and the social interactions of communities. Network theory plays a major role in both of them.

The breakthrough of network science dates back around the millennium, since the rapid and large-scale development of computer science made it possible to store and efficiently analyse complex networks. An observation, that there are properties, which are generally present in a large number of networks regardless of their origin, was also made in these years. One of the most common features of networks is the scale-free property [1], which can be interpreted such that most of the nodes of a network have only a few neighbours, but there is a small number of so-called hub nodes too, with a large amount of connections. Another frequent characteristic is the small-world property [2]. In these networks, the average distances between nodes grow proportionally to the logarithm of the network size, which means in practice that every node can be reached from any other in just a few steps.

Networks can possess other interesting properties too, such as fractality. Fractal scaling is a notion originating from geometry, but around two decades ago it was extended to complex networks as well [3]. Since then, numerous important properties have been shown to be present in fractal networks, such as robustness against intentional attacks [4] and accelerated flow [5]. Consequently, it is in dire need to uncover the underlying mechanisms

causing fractality. Several studies have been created throughout the years focusing on the exploration of the origins of fractality, but it is still an unresolved problem. In this work, we intend to discover which network characteristics influence the emergence of fractality in complex networks. To this end, we review the most influential results in the field and also suggest novel approaches concerning the conclusions of these studies. Furthermore, we propose a completely different perspective for the solution of the problem as well, which utilizes the tools of machine learning. All of the mentioned analyses rely on our large dataset consisting of both real-world a model generated networks, in order to be able to make universal findings.

In the rest of this chapter, we define the most important notions of network theory, including fractality. In Chapter 2, we lay the foundation of our analyses by showing how fractality can be determined in networks, presenting different fractal network models and describing the creation and properties of our dataset, which forms the basis of the analyses. In Chapter 3, we examine the characteristics, which have been associated with fractality, one by one, and finally present our novel machine learning approach for the problem. In Chapter 4, we summarize our findings and propose further research possibilities.

1.1 Important notions

In this section, the most important notions of network theory are presented, which may appear in the later chapters concerning the analyses. We assume that the reader is familiar with basic notions of graph theory. Throughout this paper we consider networks, which can be modeled by simple, undirected, connected graphs. If the originating network is not connected, we take only its largest connected component. Note also that in this paper the words graph and network are used as equivalent terms.

Definition 1 (Degree distribution) *The $P(k)$ degree distribution is defined to be the probability that a uniformly selected node has degree k.*

Definition 2 (Scale-free network) *A network is called scale-free, if its degree distribution follows a power law, i.e.*

$$P(k) \sim k^{-\gamma},$$

where $\gamma \geq 1$.

Definition 3 (Small-world network) *Small-world networks are those for which the average*

path length is proportional to the logarithm of the number of nodes, i.e.

$$l \sim \log N,$$

where l denotes the average of the length of shortest paths, N is the number of nodes in the network.

Definition 4 (Local clustering coefficient) *The C_u local clustering coefficient of vertex u gives the proportion of possible edges among its neighbours that are actually present in the graph. Formally:*

$$C_u = \frac{|\{(s,t) : s,t \in \Gamma_u, (s,t) \in E\}|}{\binom{deg(u)}{2}},$$

where Γ_u is the set of the neighbours of u, E denotes the edge set of the graph and $deg(u)$ is the degree of u.

Definition 5 (Edge betweenness centrality) *The betweenness centrality of an e edge is the number of shortest paths containing e divided by the number of all the shortest paths in the graph, i.e.*

$$c(e) = \sum_{u \neq v} \frac{\sigma_{uv}(e)}{\sigma_{uv}},$$

where σ_{uv} denotes the number of shortest paths connecting vertices u and v, and $\sigma_{uv}(e)$ the number of those, which contain e. We note here that Girvan and Newman defined in [6] the edge betweenness centrality of e as the number of shortest paths containing e however, in practice it is often more useful to consider it with the $1/\sum_{u \neq v} \sigma_{uv}$ normalizing constant.

Definition 6 (Eigenvector centrality) *Let $\boldsymbol{A}$ denote the adjacency matrix of the graph. Then the eigenvector $\boldsymbol{x}$ corresponding to the largest eigenvalue of $\boldsymbol{A}$ contains the centrality values. Consequently, the eigenvector centrality of the vertex u is the component of $\boldsymbol{x}$ corresponding to u. Note that in order to get an appropriate measure, the normalization of the eigenvector is necessary.*

Definition 7 (Assortativity coefficient) *Assortativity coefficient describes the connection pattern of a graph namely, that the similar or the non-similar nodes tend to connect to each other concerning some characteristics. The most common, which is considered is the degree of nodes. The assortativity coefficient defined by Newman [7] is the Pearson correlation coefficient of the degrees of the endpoints of a randomly selected edge. Formally:*

$$r = \frac{\sum_{j,k} jk(e_{jk} - q_j q_k)}{\sigma_q^2},$$

where j and k indicate the remaining degrees (i.e. degree minus 1). Furthermore, q_k is the distribution of the remaining degree, e_{jk} is the joint distribution of the remaining degrees of the two vertices, and σ_q^2 is the variance of the q_k distribution.

1.2 Fractality of networks

Similarly to the case of geometric fractals, fractality of networks can also be defined by the so-called box-covering method, using the length of shortest path between two nodes as the distance metric. The method can be summarized as follows [3]: The nodes of the network are partitioned into boxes of size l_b in a way that any two nodes of a box are at distance less than l_B from each other. Then the minimum number of boxes needed to cover the entire network with l_B sized boxes is denoted by $N_B(l_B)$. With these notations, a network is defined to be fractal, if the relation of $N_B(l_B)$ and l_B follows a power law, i.e.:

$$N_B(l_B) \sim l_B^{-d_B}.$$

The d_B exponent is called the box-dimension or fractal dimension of the network.

Unfortunately, box-covering is proved to be an NP-hard problem [8]. This means that there is no efficient algorithm, which could find the exact solution, i.e. the optimal $N_B(l_B)$ number of boxes. However, there are numerous approximating methods, for a collection and comparative analysis of box-covering algorithms, please refer to [9]. Here, we present only the most well-known and used Compact Box Burning (CBB) algorithm, which we use later for the boxing of our networks. The method goes as follows [8]:

1. Let C be the set of uncovered nodes.
2. Randomly choose a $c \in C$ node, and remove it from C.
3. Remove every node from C, which is at distance at least l_B from c.
4. Repeat steps 2 and 3 until C becomes empty. At this point, the chosen c nodes form a compact box, thus no other nodes could be added to this box.
5. Repeat steps 1-4 until the whole network is covered.

Chapter 2

Data and methods

In order to resolve the common problem of non-generality of results about the origins of fractality, our study relies on a large dataset, consisting of both real-world and model-generated networks. Needless to say that in the study of fractal networks, it has major importance to appropriately recognize the fractal nature of a network. Unfortunately, this task is far from trivial, and most of the solutions rely on visual evaluations. To avoid the uncertainty of these techniques, we apply a more mechanic method to determine the presence of fractality in networks. In this chapter we lay the foundation of our analyses, starting with the determination of fractality. After that, we present the used mathematical network models, and finally we describe our data in detail.

2.1 Determination of fractality

In theory, the determination of the fractality of a network can be done by testing if the $(l_b, N_B(l_B))$ datapoints, resulting from the box-covering method, follow a power law distribution. A statistical framework for the detection of power law behaviour in empirical data was developed by A. Clauset, CR. Shalizi and M.E.J. Newman [10], however, there are some difficulties, which is why this technique is commonly not appropriate for this particular problem. First of all, due to the NP-hard nature of the box-covering method, the use of approximating algorithms is necessary. These algorithms have an uncertainty factor, consequently we cannot guarantee that their outcomes are always accurate, and lower quality data often cause misleading results. Furthermore, statistical tests usually fail for small sample sizes, and for smaller networks or for those with small average distances, the number of points resulting from box covering is not large enough to obtain reliable

information by these tests.

Because of the mentioned difficulties, in practice, the most common technique for detecting the fractal nature of a network is to draw log-log plots of the $(l_b, N_B(l_B))$ datapoints, fit a line to them, and decide about the goodness-of-fit by the mean squared error, the coefficient of determination or by simply looking at the plots. Obviously, these methods and the conclusions drawn from their results are highly influenced by personal decisions. Furthermore, considering a large number of networks the evaluation of plots becomes impracticable. For this reason, we have been searching for a more mechanical way to decide about the fractality of networks.

We use a method, which takes advantage of the observation that while in fractal networks the $N_b(l_B) \sim l_B^{-d_B}$ relation holds, for non-fractal networks $N_B(l_B) \sim e^{-d_e l_B}$ is true [11]. This concept for the determination of fractality was introduced by K. Takemoto in [12]. Here, we apply a modified version of the method presented in [13]. Namely, we fit both a power law and an exponential curve in the form of the mentioned relations to the normalized $(l_b, N_B(l_B)/N)$ points, where N is the number of nodes in the network. The fitting is done by excluding the first point, i.e., when $l_b = 1$, because it is usually an outlier to the distribution of the data. Then the fractality is measured by the ratio of the root-mean-square errors of the two curves:

$$R = \frac{RMSE_{powerlaw}}{RMSE_{exponential}}.$$

The motivation behind the normalization of data points and the use of RMSE is to be able to compare for different networks the goodness-of-fit of each curve separately, not only their ratio.

We could say that if $R < 1$, then the network is fractal, since in this case the power law curve fits better than the exponential one, otherwise it is non-fractal. However, the presence of different properties in networks is usually not pure, especially in real networks. It is a common phenomenon that fractal scaling holds only in an $(l_{B,MIN}, l_{B,MAX})$ range of l_B [10,11]. Often for small l_B values the power law relation prevails, while for large l_B's exponential holds. Fortunately, this metric allows to measure fractality on a continuous scale, the closer R is to 0, the more fractal the network is. However, in order to compare the characteristics of fractal and non-fractal networks, we still need to create a cut-off point. We observed on both real and model generated networks that choosing it to be $R = 0.65$ can be reasonable. We can say that the investigated networks with $R < 0.65$ are rather fractal than non-fractal and vice versa. Figure 2.1 shows a few illustrative examples.

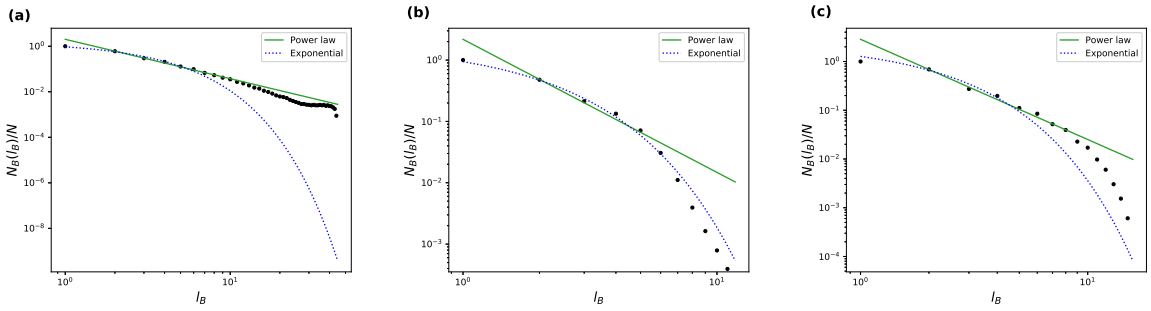

Figure 2.1: Results of fitting power law and exponential curves to some real networks, illustrated on log-log plots. **(a)** Brain network with $R = 0.007$ **(b)** Social network with $R = 3.3$ **(c)** Metabolic network with $R = 0.09$

For the sake of fairness we have to mention that the described method cannot be used for networks for which the outcome of box covering consists of only a few points (e.g., less than 6). However, for these networks, fractal nature can hardly be interpreted anyway. Furthermore, it may also not give appropriate results for some network models. The reason behind it is that in models the fractal and non-fractal scaling often only asymptotically hold. For this reason, we determine the fractality with this method only for real networks, and we stick to the theoretical parameters for models.

2.2 Network models

Mathematical models play a crucial role in the understanding of network properties. Numerous models have been created throughout the years to capture fractal scaling in networks and to discover the relations of fractality to other characteristics. In this section, we describe four such network models, including the connection of their parameters to fractality.

2.2.1 Song-Havlin-Makse model

The most widely known model in the field of network fractality is the Song-Havlin-Makse model, which was introduced in [4]. The main idea behind it is to dynamically grow a network, while influencing the degree correlations, especially the correlation between hubs. The model is defined as follows:

1. The initial graph is a simple structure, e.g. two nodes connected via a link.

2. In iteration step $t+1$ we connect $m-m$ offspring to both endpoints of every edge, i.e. a v node gains $m \cdot deg_t(v)$ offspring, where m is a predefined parameter and $deg_t(v)$ is the degree of node v at the end of step t.

3. In iteration step $t+1$ every (u, v) edge is removed independently with probability p, where p is a predefined parameter. When an edge is removed, it is replaced with x new edges between the offspring of u and v.

The fractality is influenced by the choice of parameter p namely, the generated network is fractal for $p = 1$, and non-fractal for $p = 0$ [4]. The intermediate values develop mixtures between the two properties. Our observation is that the networks resulting from setting $p \geq 0.4$ can be considered fractal, which is illustrated on Figure 2.2.

For the later analyses, we restrict ourselves to the $x = 1$ case in order to reduce the number of possible networks to be generated, and hence the computational demand of the tasks.

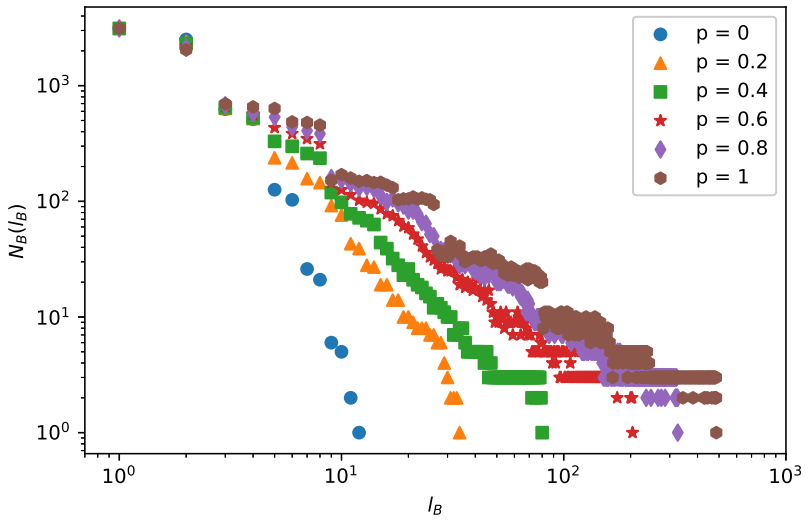

Figure 2.2: Illustration of the fractality of the Song-Havlin-Makse model for different parameter settings.

2.2.2 Repulsion based fractal model

The following model was presented in [14]. It is similar to the SHM model in a way, that starting with an initial graph it dynamically grows through iterations, and edge rewiring

influences the structure of the network. Its uniqueness is that it always generates fractal networks, irrespective of the parameter choices. The evolution of the model is as follows:

1. The initial condition and the growth of the model is the same as step 1 and 2 of the Song-Havlin-Makse model.

2. In iteration step $t+1$ we remove every (u, v) edge with probability

$$p_{uv}^{Y} = 1 - \left| Y - \frac{deg_t(u) + deg_t(v)}{2 \cdot deg_{t,max}} \right|,$$

   where $Y \in [0, 1]$ is a predefined parameter, $deg_t(u)$ is the degree of node u, $deg_{t,max}$ is the maximum degree at step t. When an edge is removed, it is replaced with x new edges between the offspring of its endpoints.

3. We add $deg_t(v)$ edges among the newly generated offspring of every old node v.

The author claims that in this way, when $Y = 0$ repulsive relation is created among small degree nodes, while for $Y = 1$ it is among hubs [14]. As Figure 2.3 shows, the model indeed generates fractal networks for all parameter settings.

As in the case of the SHM model, for the analyses we simplify the model by setting $x = 1$.

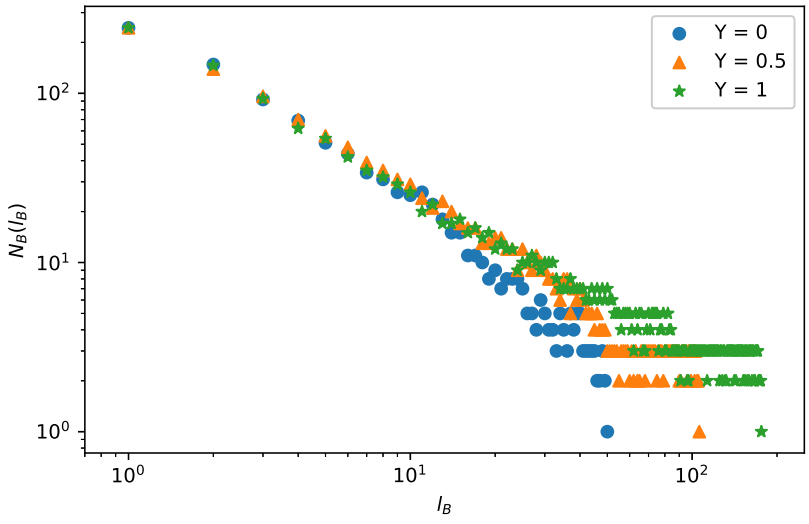

Figure 2.3: Illustration of the fractality of the Repulsion based fractal model for different parameter settings.

2.2.3 Mixture model

The Mixture model was introduced in [14]. The main idea behind it is to take a pure fractal initial graph and rewire some of its edges based on the preferential attachment mechanism to obtain a less fractal network. The steps of the model are as follows:

1. We start with a k-dimensional (practically $k = 2$) grid graph with $n_1 \times n_2 \times \cdots \times n_k$ vertices.

2. We remove every (u, v) edge with a predefined probability p, and replace it by choosing a new endpoint for the edge. The probability of selecting vertex w is proportional to
$$p_w = \frac{1}{1 + \exp\left(-a \cdot \left(\frac{deg(w)}{deg_{max}}\right) - \frac{1}{2}\right)},$$
where a is a positive constant, $deg(w)$ is the degree of node w and deg_{max} is the maximum degree of the current graph. By practical motivation, to avoid multiple edges, we exclude the neighbours of the starting point from the set of possible endpoints. By default, u is chosen to be the starting point of the new edge, however if in this way the graph becomes disconnected, v is chosen instead.

The fractality of the generated network depends on the choice of p. For $p = 0$ the network is purely fractal, and as p grows the model shows a transition from fractal to non-fractal networks [14]. We can say that a cut-off point can be created at $p = 0.01$, as Figure 2.4 demonstrates.

2.2.4 (u, v)-flower

The family of (u, v)-flowers was created by Rozenfeld, Havlin and Ben-Avraham [15]. Similarly to most of the previous models, this model also generates networks through iterations, but the edge replacement procedure is quite different. The model is defined as follows:

1. The initial graph is a cycle consisting of $w = u + v$ nodes and edges, where u and v are predefined parameters, and we can assume that $u \leq v$.

2. In iteration step $t + 1$ every (x, y) edge is replaced by two paths connecting x and y, one with length u and one with length v.

H. D. Rozenfeld et al. showed in [15] that the model generates fractal networks when $u > 1$, and non-fractal ones when $u = 1$. This statement is illustrated on Figure 2.5.

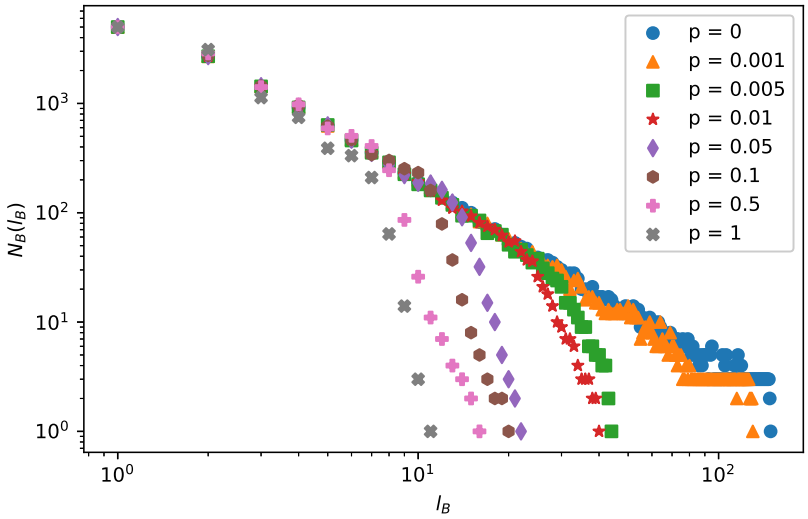

Figure 2.4: Illustration of the fractality of the Mixture model for different parameter settings.

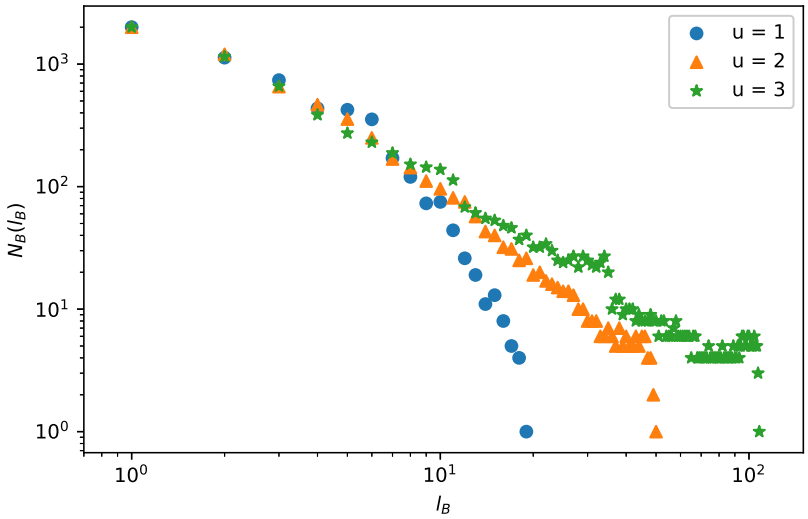

Figure 2.5: Illustration of the fractality of the (u, v)-flower for different parameter settings.

2.3 Data

To thoroughly analyse network properties, search for possible relations among them, and to be able to draw relevant conclusions, it is essential to consider a diverse and large-scale collection of networks as the basis of the analyses. Although mathematical models can give insight to the evolution of networks and some properties in focus, they usually

cannot capture every characteristic of real networks. Striving to be as comprehensive as possible, we generated networks with the models described in Section 2.2 considering various parameter settings, furthermore, we collected a large number of real networks originating from six different domains.

During the selection of parameter settings for the different models, we aimed to get a representative sample of the possible networks, which can be generated, while keeping the number of networks reasonably low for computational purposes. For this reason also we limited ourselves mostly to networks with at most 5000 number of nodes. Fortunately, the number of nodes is deterministic in the parameters for all models, thus this limitation could be easily controlled. Our choices of parameter values can be summarized as follows: For all the cases below, n denotes the number of iterations the models do.

Song-Havlin-Makse model

$$
\begin{aligned}
&n = 1, 2, \ldots, 5 \\
&m = \begin{cases} 1, 2, 3, 5, 10, 20, 50, & \text{if } n = 1 \\ 1, 2, 3, 5, 10, 20, & \text{if } n = 2 \\ 1, 2, 3, 5, 10, & \text{if } n = 3 \\ 1, 2, 3, & \text{if } n = 4 \\ 1, 2, & \text{if } n = 5 \end{cases} \\
&p = 0, 0.2, \ldots, 0.8, 1
\end{aligned}
$$

136 generated networks in total.

Repulsion based fractal model

$$
\begin{aligned}
&n = 1, 2, \ldots, 5 \\
&m = \begin{cases} 1, 2, 3, 5, 10, 20, 50, & \text{if } n = 1 \\ 1, 2, 3, 5, 10, 20, & \text{if } n = 2 \\ 1, 2, 3, 5, & \text{if } n = 3 \\ 1, 2, 3, & \text{if } n = 4 \\ 1, & \text{if } n = 5 \end{cases} \\
&p = 0, 0.5, 1
\end{aligned}
$$

63 generated networks in total.

Mixture model

$$N = n_1 \cdot n_2 = 10, 50, 100, 200, 500, 800, 1000, 1500, 2000, 3000, 5000,$$

and n_1 and n_2 is chosen in a way that $|n_1 - n_2|$ becomes as small as possible

$$p = \begin{cases} 0, 0.00001, 0.0001, 0.001, 0.0025, 0.005, 0.0075, 0.01, \\ 0.05, 0.1, 0.25, 0.5, 0.75, 1 \end{cases}$$

154 generated networks in total.

(u, v)**-flower**

$$1 \leq u \leq v, \text{ such that } w = u + v = 2, 3, \ldots, 10$$

$$n = \begin{cases} 9, & \text{if } w = 2 \\ 8, & \text{if } w = 3 \\ 6, & \text{if } w = 4 \\ 5, & \text{if } w = 5, 6 \\ 4, & \text{if } w = 7, 8, 9, 10 \end{cases}$$

114 generated networks in total. We have to mention that unfortunately, during the preparation of the data some networks got lost. This is the reason why the total number of networks is not equal to the number of possible networks, which can be generated with the described parameter settings.

For those analyses, where the evaluation of results is needed to be done network-wisely and is based on making plots, we restricted ourselves to a smaller number of networks. We created three categories, namely where the number of nodes is approximately 800, 2000, and 5000. In the case of the Song-Havlin-Makse and Repulsion based fractal model, for every category, one realization of the (n, m) pair of parameters was selected, which generates networks with number of nodes corresponding to the categories. All the previously mentioned p values were assigned to these (n, m) pairs, resulting in six generated networks per category for the Song-Havlin-Makse, and three for the Repulsion based fractal model. Similarly, for the Mixture model one realization of the (n_1, n_2) pair was chosen for every category, according to the previously detailed rule. Moreover, we selected six values for parameter p, three below 0.05, and three above it. Finally, in the case of the (u, v)-flower, we considered all the (u, v, n) parameter settings, which generate networks with sizes corresponding to the categories, thus three to five networks got assigned to each category, including both fractal and non-fractal ones.

Real networks were collected from online repositories [16–28]. Table 2.1 gives a short description of the different domains and shows the corresponding number of gathered networks. In total, we work with 275 real-world networks. Some of their main features can be seen in Table 2.2, aggregated by domains. For those analyses, which require visual evaluation, we selected four to six networks from every domain, bearing in mind to have both fractal and non-fractal networks from all size categories presented in the domain.

We decided about the fractality of the networks in the same way as described in Section 2.1. In order to eliminate the randomness of the box-covering algorithm and the network models, we repeated the procedure 15 times. More precisely, in the case of the models, for a particular parameter setting we generated 15 networks and considered their average box-counts, while for the real networks we performed the box-covering 15 times and averaged their outcomes. The resulting class distribution of the datasets are shown on Figure 2.6. There are much more fractal networks amongst the model-generated ones, while real networks are rather non-fractal. However, combining the two datasets the classes become balanced.

Domain	**Description**	**Number of networks**
Brain	Human and animal connectomes (neural connections in the brain)	45
Metabolic	Protein-protein interactions of organisms	47
Cheminformatics	Graph structure of enzymes	45
Infrastructural	Transportation and distribution networks	30
Foodweb	What-eats-what in an ecological community	69
Social	Facebook, Twitter and collaboration networks	39

Table 2.1: Description of the network domains and the numbers of collected networks.

Domain	Number of nodes		Number of edges		Diameter	
	AVG	(MIN, MAX)	AVG	(MIN, MAX)	AVG	(MIN, MAX)
Brain	1073	(65, 2989)	7485	(730, 31548)	14	(2, 43)
Metabolic	815	(11, 2831)	2099	(10, 20448)	9	(1, 22)
Cheminformatics	58	(44, 125)	100	(77, 149)	16	(9, 37)
Infrastructural	1268	(8, 6474)	2396	(7, 15645)	37	(3, 122)
Foodweb	115	(19, 765)	655	(38, 6613)	5	(2, 9)
Social	4465	(86, 9763)	7300	(117, 24806)	14	(6, 28)

Table 2.2: Some of the main features of the collected real networks. The average, minimum and maximum value of the number of nodes, the number of edges and the diameter of the networks by domains.

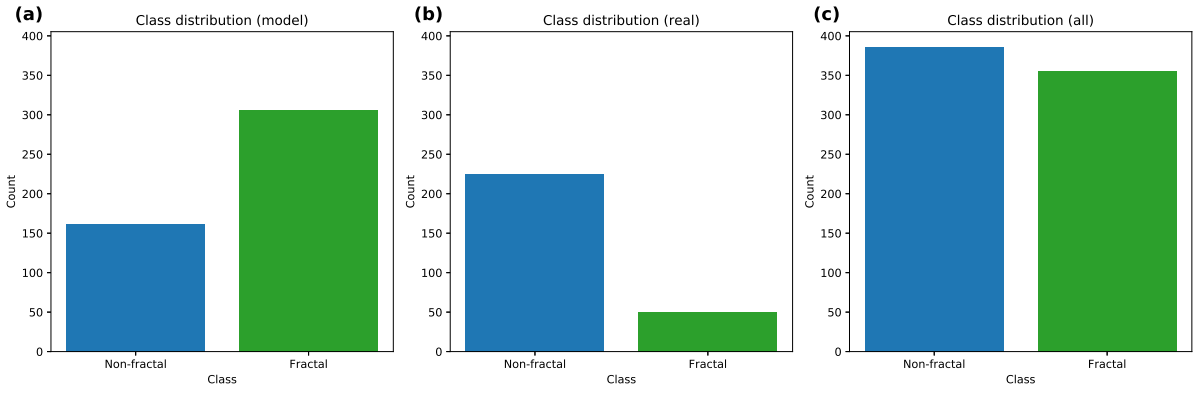

Figure 2.6: Class distribution concerning fractality for **(a)** model generated, **(b)** real-world, **(c)** all of the examined networks together.

Chapter 3

Analysis of network characteristics

In this chapter, we intend to give a comprehensive analysis of the relation of fractality to other network properties. We revisit the widely researched suggestions that fractal scaling originates from disassortativity and/or from the repulsion between hub nodes. We also reconsider some newer questions concerning long-range correlation of nodes and edge betweenness centrality. Where it is possible, we follow the methodology of the corresponding research papers and extend the results to a larger scale relying on our massive dataset. Moreover, we also give our own approach to the examined issues and investigate the relevance of the suggestions from all these perspectives. Finally, we study the connection of fractality to other network characteristics from a whole different point of view. Namely, we are searching for important features of fractal networks with the help of machine learning algorithms.

3.1 Disassortativity and hub repulsion

The first network properties, which were associated with fractality are disassortativity and repulsion between large degree nodes, i.e., hubs. It was suggested that fractal nature of networks originates from these characteristics [4, 29]. There are numerous papers dealing with this statement, some support it [30], and there are more that confute it [14, 31–33].

In most of the studies, the concepts of disassortativity and hub repulsion coincide, although the latter can be considered only as the practical interpretation of the former. For this reason, we rather separate the two notions and measure the assortativity of a network by the classic assortativity coefficient, and define a different hub repulsion score. For this, let N_{hub} denote the number of hubs in the network and E_{hub} the number of edges

among these hubs. Then the hub repulsion score (HRS) is defined to be one minus the ratio of the number of realized and possible edges among hubs. Formally:

$$HRS = 1 - \frac{E_{hub}}{\binom{N_{hub}}{2}}.$$

In this way, HRS is small for those networks, in which hubs tend to connect to each other (no repulsion), and large, when there are only a few or no edges among them (strong repulsion).

During the analysis, we define hubs as the nodes whose degree is above the 90th percentile of the degree distribution, i.e., the top 10% of nodes according to their degree. Furthermore, both the assortativity coefficient and the hub repulsion score is averaged over 15 realizations of the network models for each parameter setting.

Results

In the case of the network models, the change in the assortativity coefficient is observed as a function of the parameter, which influences the fractality of the network. Figure 3.1 illustrates the results on three examples for every model. On the one hand, it can be said that three of the four models support the statement that fractal networks are disassortative. It has to be mentioned though, that while the (u, v)-flower also supports that the difference in fractal and non-fractal networks can be found in their assortativity pattern, the Song-Havlin-Makse model fails on this conjecture, since it always generates diasassortative networks. On the other hand, the Mixture model is a complete counterexample for the mentioned beliefs, because it not only generates instances for assortative fractal networks, but we can observe a positive connection between fractality and assortativity, as it was also pointed out in [14]. Finally, real networks do not show any pattern concerning the assortativity of fractal and non-fractal networks. They are often disassortative regardless of fractality, but among the examined fractal networks there are as much assortative as disassortative, which is well illustrated on Figure 3.2. Overall, we can support the conclusion of [31] namely, that fractality is independent of the assortative mixing.

Regarding the hub repulsion, we can also say that some models support the conjecture, that this property may lie behind fractality. Figure 3.3 shows how the hub repulsion score characterizes the different cases of the (u, v)-flower. It can be seen on both subfigures that in the $u = 1$, i.e., the non-fractal case, the scores are much lower, than in the fractal cases. Furthermore, the scores are not just higher for the fractal networks, they are high in

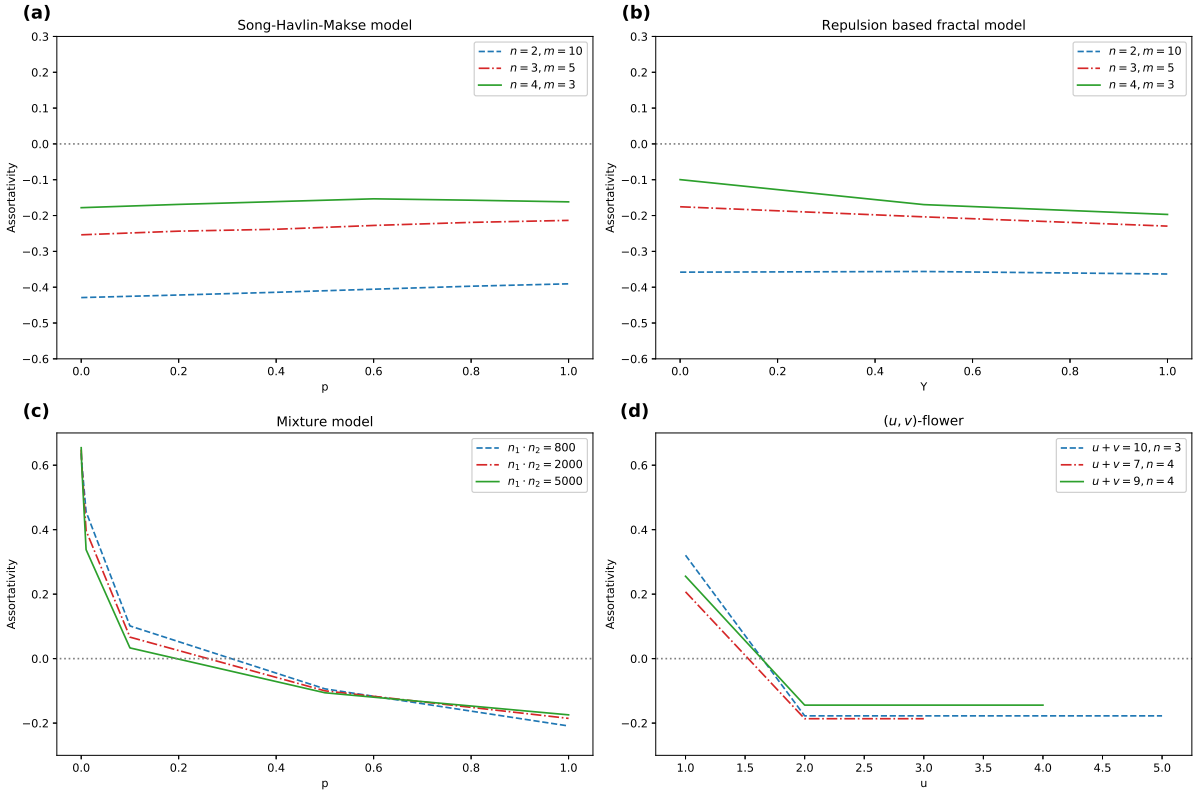

Figure 3.1: Assortativity as a function of the parameter, which influences fractality for the **(a)** Song-Havlin-Makse model, **(b)** Repulsion based fractal model, **(c)** Mixture model, **(d)** (u, v)-flower. The grey dotted line at Assortativity $= 0$ is a guide for the eye.

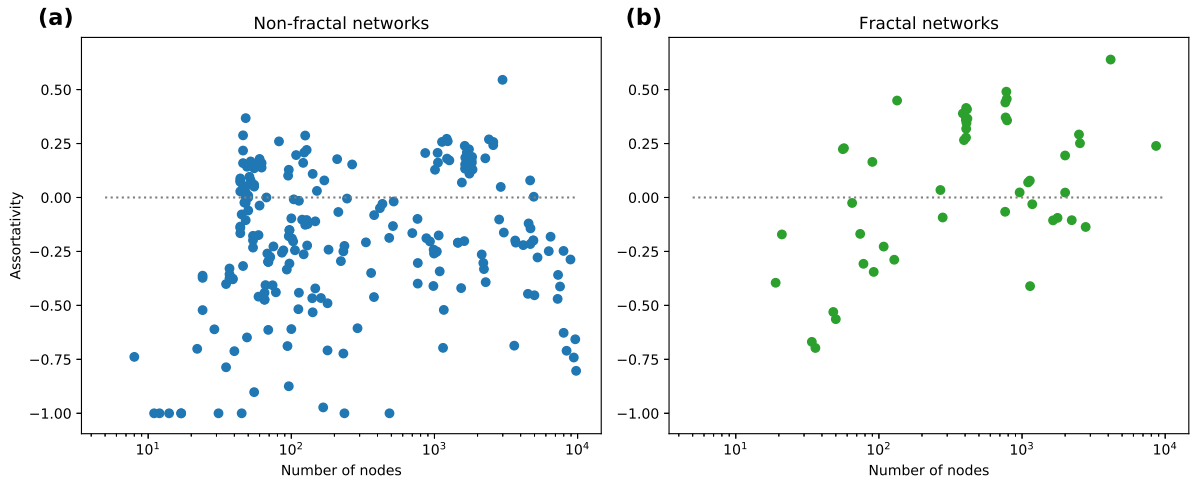

Figure 3.2: Assortativity of real networks plotted against the number of nodes on a semi-log scale. Subfigure **(a)** shows the non-fractal, subfigure **(b)** the fractal networks. The grey dotted line at Assortativity $= 0$ is a guide for the eye.

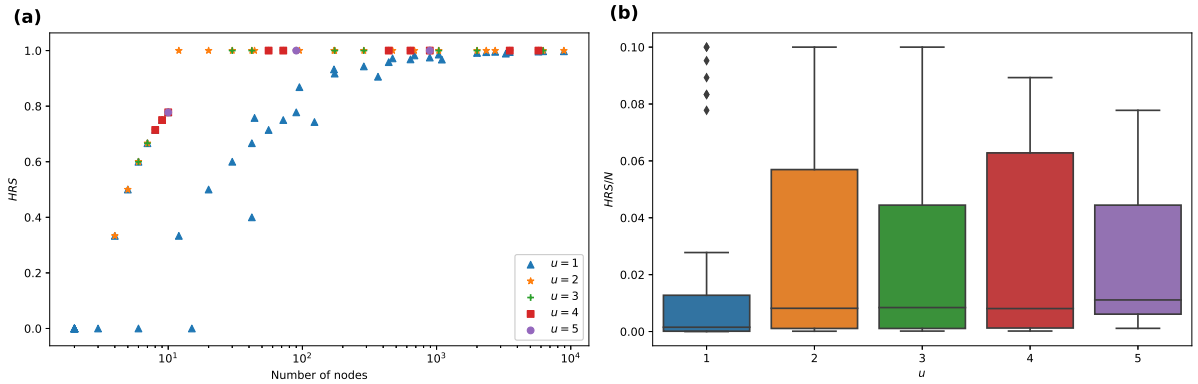

Figure 3.3: Plots of hub repulsion score for the (u, v)-flower. **(a)** HRS as a function of the number of nodes on a semi-log scale. **(b)** Boxplots of the normalized hub repulsion scores for different u parameters. The normalization is done by dividing by the number of nodes in the network.

general. However, it can be observed that the HRS gets close to 1 for large networks, for the non-fractal cases too. Similar behaviour is presented by the Repulsion-based fractal model and the Song-Havlin-Makse model as well, consequently we have to assume that the score may not be reliable in larger scales. Nevertheless, we can still use it for smaller networks and observe that the Song-Havlin-Makse model shows that hub repulsion is much stronger in the fractal networks than in the non-fractal ones. Concerning the Repulsion-based fractal model, we can say that the HRS and Y show a positive relation, and that fractal networks do not necessarily have to possess a strong hub repulsion property. It has to be mentioned that for those networks for which hub repulsion cannot be interpreted (because there are no hubs), this hub repulsion score practically cannot get much below 1, because of the large number of possible edges among the considered nodes. This is the reason why for most cases of the Mixture model it cannot give useful measures.

Lastly, as Figure 3.4 shows, real networks fully contradict the conjecture that hub repulsion may cause fractality. There are fractal networks with low HRS and non-fractal ones with high score too, and there is no significant difference in the distribution of the values for the two classes. We would like to note here how important it is to take into consideration the network sizes when comparing the hub repulsion scores. There is a tremendous difference between the two boxplots on Figure 3.4 **(b)** and **(c)**, and the reason behind it is that smaller networks can have smaller HRS more easily, while large networks cannot obtain significantly less than 1, as it was pointed out earlier. Therefore

having a bunch of small and only a few large networks in one group, and reversely in the other group, the pure (mean) scores may be misleading.

In conclusion, we can say that, similarly to disassortativity, hub repulsion also seems to be independent of fractality.

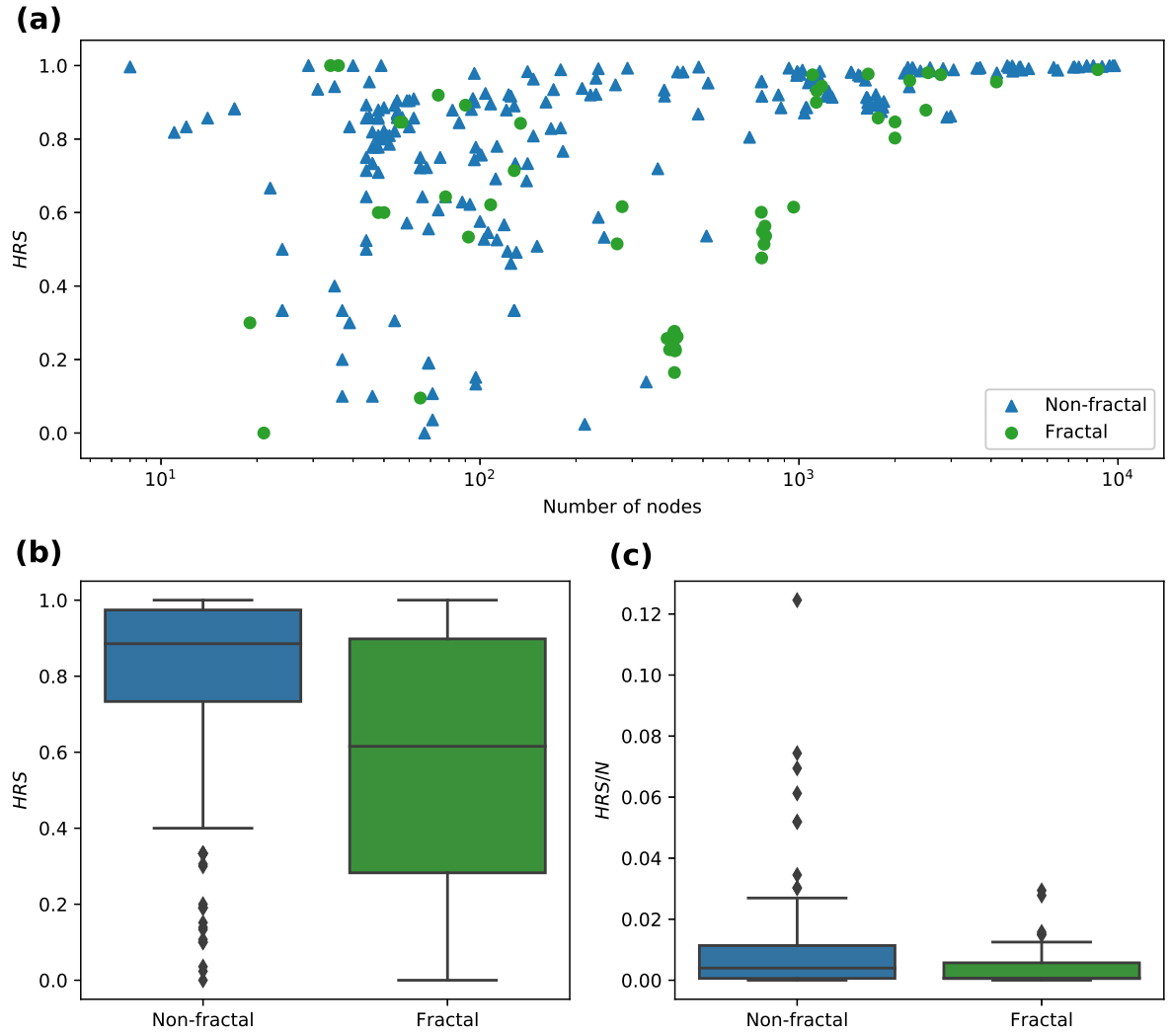

Figure 3.4: Plots of hub repulsion score for real networks. **(a)** HRS as a function of the number of nodes on a semi-log scale. **(b)** Boxplots of HRS in the fractal and non-fractal class. **(c)** Boxplots of the normalized hub repulsion scores for the two classes. The normalization is done by dividing by the number of nodes in the network.

3.2 Long-range correlation

Besides the direct relations of nodes, long-range correlations can also be captured and analysed in networks. This extension of degree anticorrelation has already been associated with fractal scaling [32,34]. Both studies suggest, based on different approaches, that there is a connection between long-range negative correlation and fractality. Here, we apply both of these methods on our data in addition to a more immediate extension of neighbour-level degree correlation measures, introduced in [35].

In [34] a fluctuation analysis approach was proposed to measure long-range correlations. We follow the steps of this method, which can be summarized as follows:

1. Consider all shortest paths in the network with length d. For all of these paths calculate the average degree of the nodes on the path.
2. Calculate $F(d)$, which is the standard deviation of the averages calculated in step 1.
3. Repeat step 1 and 2 for all possible d.
4. Examine if the relation of $F(d)$ and d follows power law with exponent α. If so, $-\frac{1}{2} < \alpha < 0$ suggests positive, $-1 < \alpha < -\frac{1}{2}$ negative long-range correlations.

An extension of the concept of hub repulsion to long-range scales was proposed in [32]. They examined how the distribution of hub distances look like in fractal and non-fractal networks. The procedure can be summed up by the following steps:

1. Calculate the distance of all pairs of hubs.
2. For all distance l calculate $\hat{P}(l)$, which is the number of hub pairs separated by a shortest path of length l.
3. Calculate $\tilde{P}(l)$ by dividing $\hat{P}(l)$ by the number of possible edges among hubs, i.e. by $\binom{N_{hub}}{2}$.

In this way, $\tilde{P}(l)$ is the probability that a randomly selected pair of hubs is at distance l from each other. In order to be consistent with the results of [32], for this analysis we cut off the hubs at the 98th percentile of the degree distribution.

The third approach, which we consider here for capturing long-range correlations in networks, was introduced in [35]. It extends the notion of neighbour connectivity to nodes at a distance larger than one. The main concept of the method can be summarized as follows:

1. Fix m, and for every node v average the degree of nodes, which are at distance m from v.

2. Calculate $\langle k_m \rangle(k)$ by taking the average of the outputs of step 1 over nodes with degree k.

3. Examine the relation of k and $\langle k_m \rangle(k)$.

Following the details of the analysis of [35], we consider the values of m up to 5, and assume power law relation between k and $\langle k_m \rangle(k)$.

Results

The results of the fluctuation analysis are illustrated for some model-generated and real-world networks on Figure 3.5. Generally, it can be said that for the (u, v)-flower and the Song-Havlin-Makse model the relation of d and $F(d)$ of the fractal networks follow power law with exponent less than $-\frac{1}{2}$, while in the non-fractal cases the mentioned relation is rather exponential, which support the observations of [34]. However, for the Repulsion-based fractal model $F(d)$ do not seem to have power law distribution. It may not be immediately visible from Figure 3.5**(b)**, but testing if power law or exponential distribution is more appropriate for the data, we get that exponential fits better. For this purpose, we use the *powerlaw* Python package [36]. In the case of the Mixture model, none of the previously mentioned distributions seem to fit for the results of the fluctuation analysis for the fractal networks. Among real-world networks there are some cases, where the expected behaviour of $F(d)$ can be observed, as Figure 3.5**(c)** shows. However, there are examples, where power law relation cannot be detected, thus long-range correlations cannot be concluded, as it is illustrated on Figure 3.5**(d)**. In conclusion, we can say that long-range anticorrelation captured by fluctuation analysis is not a universal property of fractal networks.

Concerning the distribution of hub distances, we can say that most of the examined models support the suggestion of [32], that in fractal networks hub distances have a wide range, while in non-fractal networks hubs cannot get far from each other. However, a surprising observation can be made based on the Song-Havlin-Makse model. As Figure 3.6**(a)** illustrates, the range of hub distances widens as p grows, but in the pure fractal $p = 1$ case it falls back to the level of non-fractal networks. Consequently, this model seems to contradict, that there have to be larger hub distances in fractal networks, than

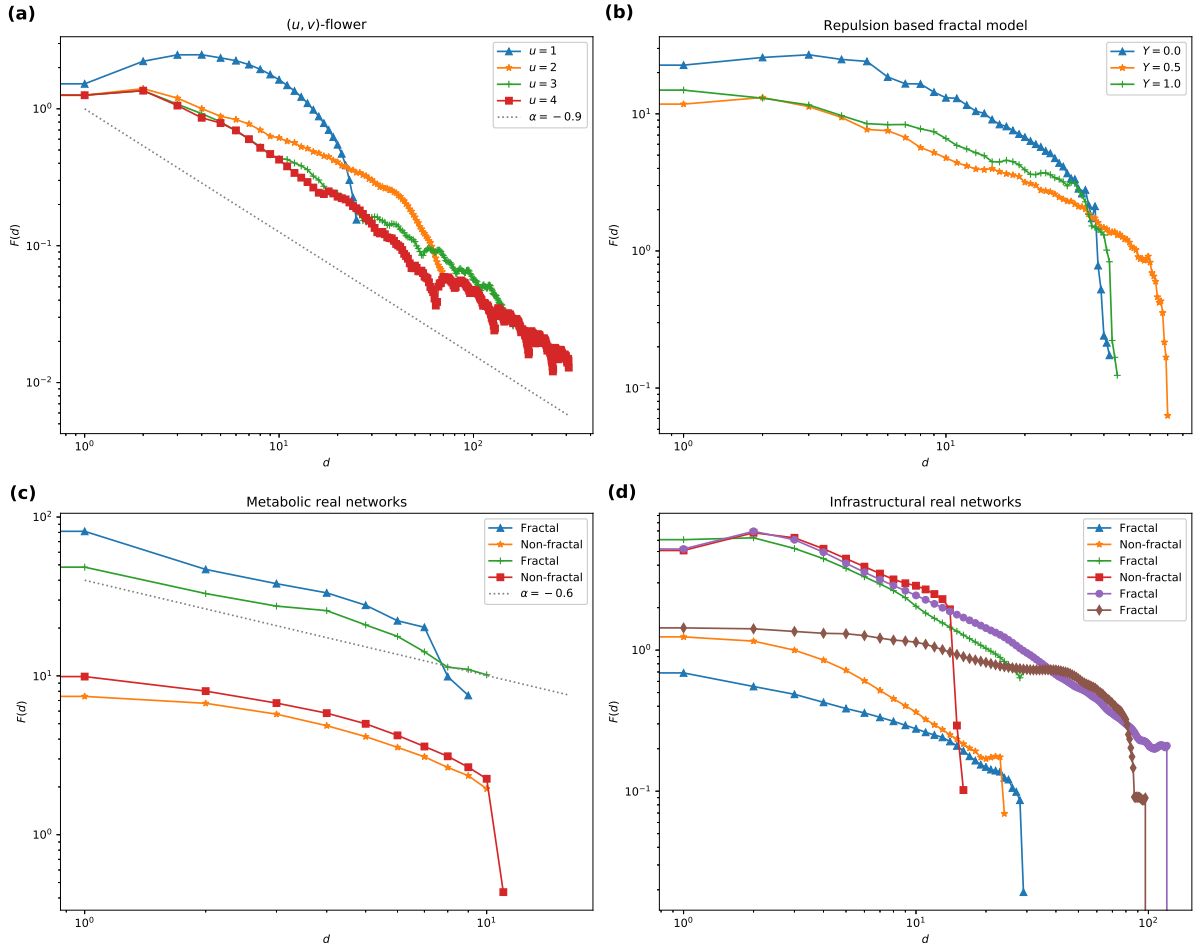

Figure 3.5: Results of the fluctuation analysis on a log-log scale for some examples of the **(a)** (u, v)-flower, **(b)** Repulsion based fractal model. **(c)** illustrates the cases of some metabolic, **(d)** infrastructural real networks. The dotted grey lines on subfigures **(a)** and **(c)** are guides for the eye.

in non-fractal ones. Real networks also turn out to be counterexamples for this conjecture. Although, there are some cases, where the hoped property can be observed. The first two subplots of Figure 3.6**(b)** give an example for that, where two networks of the same size are considered, and the fractal one clearly possesses larger hub distances, than the non-fractal one. However, most of the time both fractal and non-fractal networks have small hub distances, thus this property seems to be independent of fractality. An example for this suggestion is given on the third and fourth subplots of Figure 3.6**(b)**. They show that two fractal networks of the same size can differ completely concerning their hub distances.

Finally, by the results of the third approach of capturing long-range correlations, we

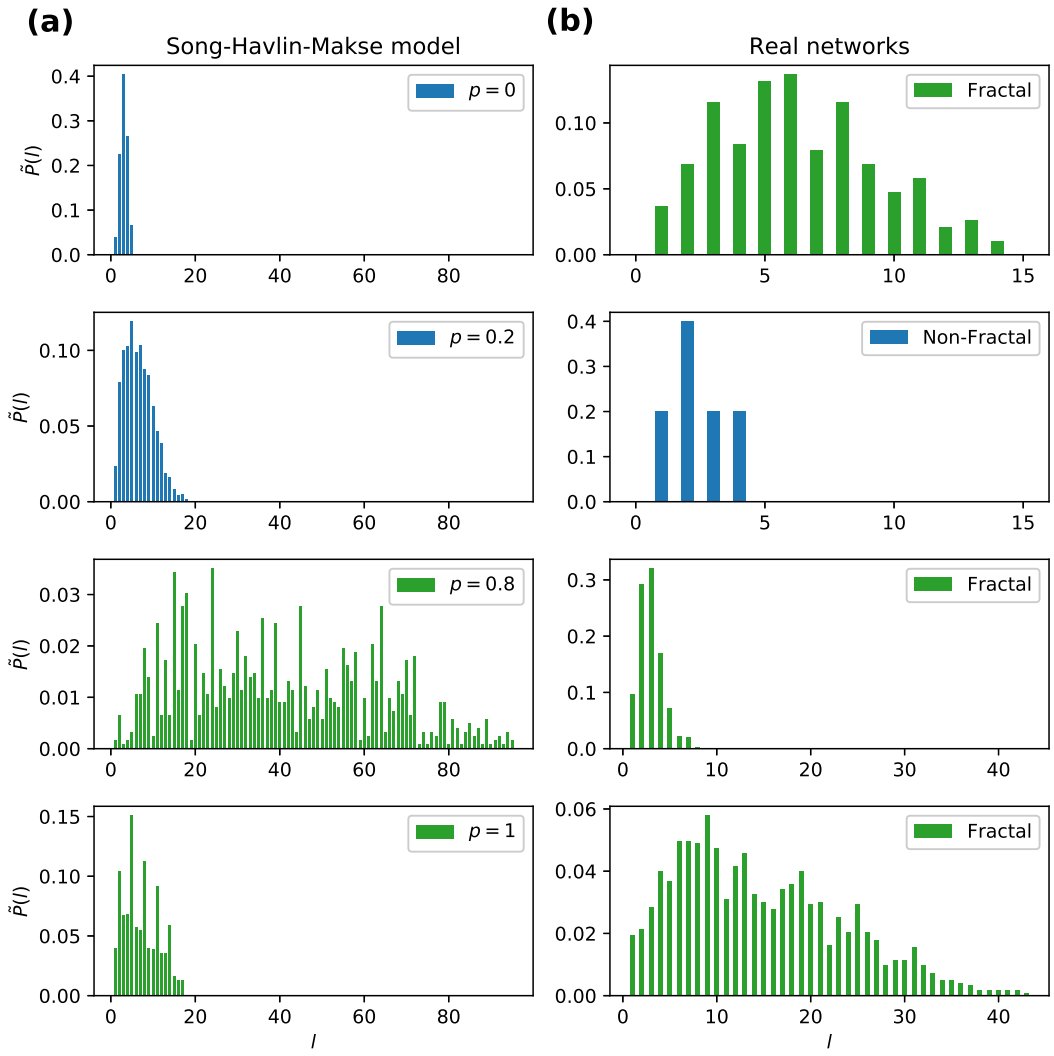

Figure 3.6: Distribution of hub distances for **(a)** some cases of the Song-Havlin-Makse model, **(b)** some real networks. On subfigure **(b)**, for the first two subplots two networks with similar sizes are considered, and similarly to the third and fourth subplots.

can say that there is no connection between the fractality and the long-range correlation profile of networks. It can be observed that in the case of the Repulsion-based fractal model and the Song-Havlin-Makse model, usually no correlation can be detected for $m \geq 3$ and until that, the correlation profile does not change. For the (u, v)-flower, at $m = 3$ or $m = 4$ the reverse of the $m = 1$ case can be observed for all networks, independently of fractality. Networks generated by the Mixture model preserve their correlation profile for all m distances, i.e., fractal networks have positive degree correlations, while non-fractal networks have negative, even in the long-range scale. This phenomenon is well

illustrated on Figure 3.7. Degree correlations of the real networks are usually preserved or reversed for larger distances, but do not seem to disappear. However, these processes are independent of the fractality of the networks. To illustrate this, Figure 3.8 shows two brain networks, one of them is fractal, the other one is not, and their correlation profile is exactly the same for all m distances. Lastly, the networks considered on Figure 3.9 show that fractal networks with negative correlation on the direct neighbour level can show positive correlation by stepping a little farther, at distances of $m = 2$ in these particular cases.

Overall, we can conclude from the results of all three approaches, that fractality seems to be independent of long-range correlation profiles.

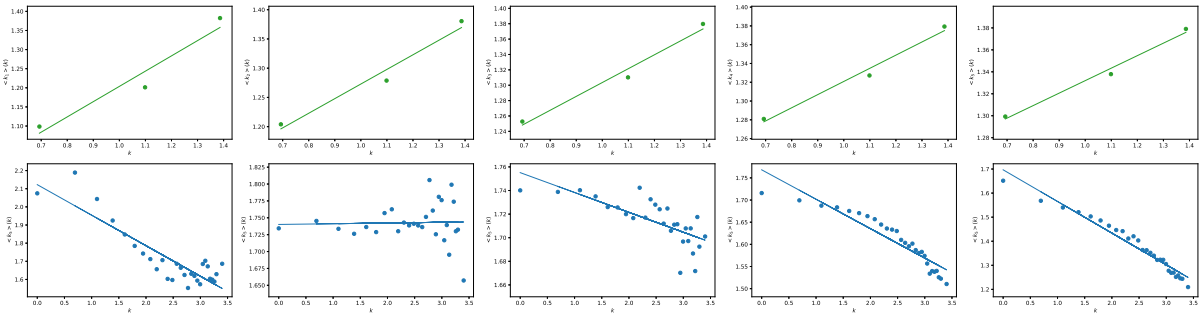

Figure 3.7: Degree correlations of the two extreme cases of the Mixture model at distances from 1 to 5. The first row corresponds to the $p = 0$, the second to the $p = 1$ case. $\langle k_m \rangle(k)$ is plotted against k on a log-log scale, and the line fitted to the log-transformed data is also provided.

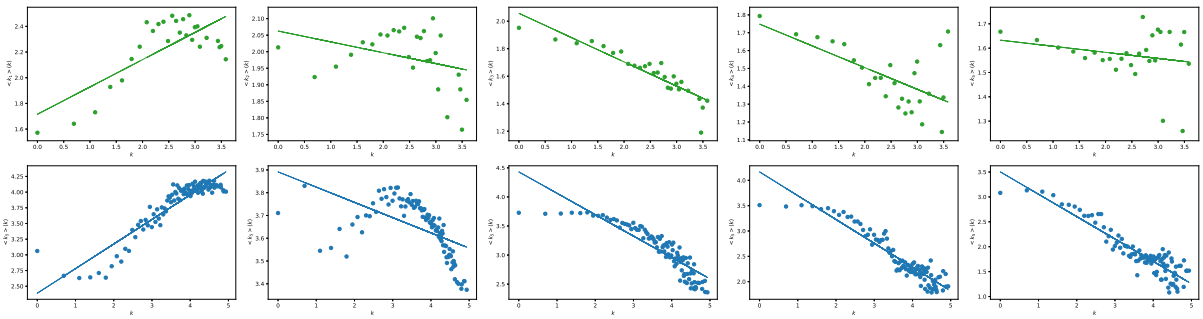

Figure 3.8: Degree correlations of a fractal (top row) and a non-fractal (bottom row) brain network at distances from 1 to 5. $\langle k_m \rangle(k)$ is plotted against k on a log-log scale, and the line fitted to the log-transformed data is also provided.

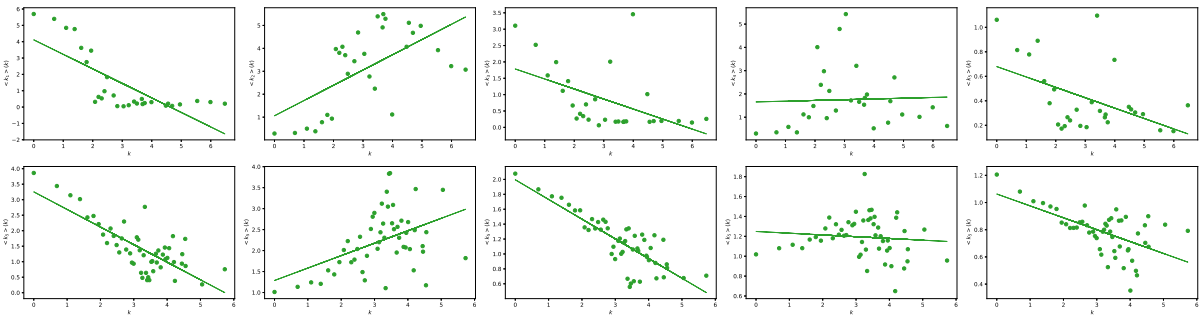

Figure 3.9: Degree correlations of two fractal social networks at distances from 1 to 5. $\langle k_m \rangle(k)$ is plotted against k on a log-log scale, and the line fitted to the log-transformed data is also provided.

3.3 Edge betweenness centrality

In [33], the authors reported that even a small number of edges with high betweenness centrality (BC) can destroy the fractal scaling of a network. Although, they investigated this conjecture from the point of minimum spanning trees, here we rather study the suggestion generally on the exact networks. Consequently, we examine the question that fractal networks can or cannot have edges with high betweenness centrality. To this end, for every network we divide the edges into four categories according to their betweenness centrality by the following percentile cutpoints: 95, 85, 70. Then we take the average of the edge betweenness centralities in every category and examine if they are lower for fractal networks, than for non-fractal ones, and if the average drops significantly after the first category for fractal networks.

Results

Concerning the network models it can be said that some of them generate networks with edges having small betweenness centrality, and some can create edges with quite large BC too. However, these are model-specific properties, thus conclusions can also be made only on each model separately. Figures 3.10 **(a)** and **(b)** show the resulting average edge betweenness centralities aggregated for the two extreme cases of the Song-Havlin-Makse model. In the pure fractal $p = 1$ case the average BC of the top 5% of the edges is roughly the same, independently of the other parameters, and can be considered high in general. The averages in the other categories drop significantly for some networks, but not necessarily for every fractal network. In the pure non-fractal $p = 0$ case, the averages

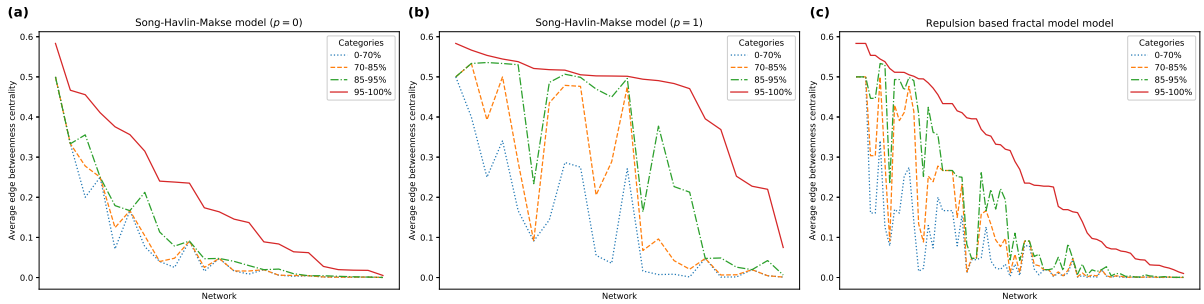

Figure 3.10: Average edge betweenness centralities in the different categories created by the BC of the edges. **(a)** Song-Havlin-Makse model in the $p = 0$ case, **(b)** Song-Havlin-Makse model in the $p = 1$ case, **(c)** Repulsion based fractal model. On the x-axis, the networks are ordered by the average edge betweenness centralities.

in every category can take on values from a wide range, but they are generally, especially in the highest category less than the ones for the fractal networks. Similar observations can be made on the Mixture model, with the note that in that case the edge betweenness centralities are low in general for all parameter settings.

For the (u, v)-flower there seems to be no significant difference in the averages between the fractal and non-fractal networks, except for a few extreme cases, where they are all 1 for some non-fractal networks. Figure 3.10**(c)** illustrates the average edge betweenness centralities in the different categories for networks generated by the Repulsion-based fractal model. It can be said that there are fractal networks, for which the averages in all categories are high, and there are such as well, where they are low. Significant drops can be detected in about half of the cases. Finally, Figure 3.11 shows the results for the real networks. It can be observed, that the averages are higher in general for the fractal than non-fractal networks. Note that they also seem to be more stable across the different categories for non-fractal networks, significant drops are more common for fractal ones. Consequently, there are fractal networks, in which indeed only a few edges with high BC occur, but it is not a universal property.

As a summary, we can conclude that fractal networks can have edges with high edge betweenness centrality too. Comparing a large number of networks from the fractal and non-fractal classes, we could say that the averages of the BCs in the different categories are often higher for the fractal networks, but this is not a universal difference between the two classes by no means.

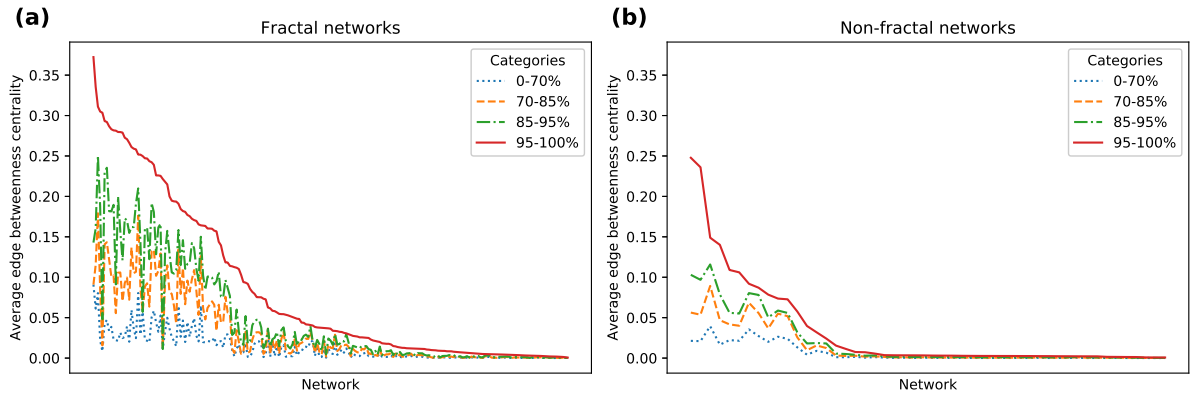

Figure 3.11: Average edge betweenness centralities in the different categories created by the BC of the edges. **(a)** Fractal real networks, **(b)** Non-fractal real networks. On the x-axis, the networks are ordered by the average edge betweenness centralities.

3.4 A machine learning approach

In the previous sections, the main concept in discovering the origins of fractality was to take a network characteristic and examine if there is any difference between fractal and non-fractal networks concerning this particular property. Here, we propose a different approach to uncover the underlying mechanisms behind fractality. Namely, we address this problem as a binary classification task. We intend to distinguish fractal and non-fractal networks based on a few selected network characteristics, with the help of machine learning algorithms, and identify the most important features. In this way, we could conclude which properties cause the fractal scaling of networks.

There are numerous machine learning algorithms, which can be used for solving classification problems. Here, we apply three widespread methods, each of them are decision tree based, namely, simple decision tree, random forest and XGBoost. During the selection of explanatory variables we are bearing in mind to get a collection of characteristics, which represents the structure of the networks, but it is not too correlated. Moreover, we aim to make these metrics as independent of the network size as possible, hence where it is reasonable, normalization is also performed. The selection is carried out by [28] and [37]. In addition, we also consider some metrics resulting from the previous analyses. The list of the final explanatory variables together with their description can be found in Table 3.1.

We consider three datasets to perform the task on, one consisting of the model-generated networks, one of the real networks, and one which concatenates the two sets, thus including all examined networks. More precisely, we drop the small networks from

Name	Description
avg_path_logsize	Average of the length of shortest paths, divided by the logarithm of the number of nodes
avg_deg	Average degree
assortativity	Assortativity coefficient
avg_clust	Average of the local clustering coefficients
max_eigen	Maximum of the eigenvector centralities
skew_deg_dist	Skewness of the degree distribution
diam_logsize	Diameter, divided by the logarithm of the number of nodes
max_deg_n	Maximum degree, divided by the number of nodes minus 1
hub_repulsion	Hub repulsion score (see: Section 3.1)
ebc_avg_95	Average of the edge betweenness centralities over the top 5% of edges (see: Section 3.3)

Table 3.1: Name and description of the chosen explanatory variables, for the classification task.

all dataset, i.e. the ones whose number of nodes is less than 100, because in most of these cases fractality can hardly be defined, as it was also mentioned earlier. We use $\frac{2}{3}$ of the datasets for training and the remaining $\frac{1}{3}$ for testing. Two evaluation metrics are used to measure the performance of the algorithms, accuracy and the Area Under the (ROC) Curve (AUC). To read about the definition and usage of these measures, as well as the basics of data science, we suggest the following introductory book: [38]. The parameter optimisation for the algorithms is carried out based on the latter one, because for unbalanced classes accuracy can often be misleading. For the data preparation, training, and evaluation of algorithms, we use the *scikit-learn* [39] and *XGBoost* [40] Python packages.

To identify, which are the important variables in the decision making process of the algorithms, we calculate the *permutation importance* score of the features. These values show, how much the performance of the model decreases if the values of a given attribute are randomly permuted. For details concerning the determination of the scores, please refer to the documentation of the *permutation importance* function[1] of the *scikit-learn* package.

[1] `https://scikit-learn.org/stable/modules/permutation_importance.html`

	Decision tree			Random forest			XGBoost		
	model	**real**	**all**	**model**	**real**	**all**	**model**	**real**	**all**
Accuracy	0.86	0.93	0.91	0.95	0.74	0.95	0.86	0.89	0.91
AUC	0.95	0.91	0.95	0.99	0.95	0.98	0.97	0.81	0.96

Table 3.2: Accuracy and AUC scores of the different algorithms on the fractal/non-fractal classification task, on the three considered datasets.

Results

The performances of the models measured on the test sets are summarized in Table 3.2. It can be said that all of the algorithms can solve the problem with high accuracy and AUC score, thus we can conclude that fractal and non-fractal networks indeed differ in the considered network characteristics. To see, which are the distinguishing features, we look at the *permutation importance* scores for every algorithm and dataset. We can observe that in the case of the dataset of the model generated networks, the normalized average path length has the most significant importance for all algorithms. In addition to it, the normalized diameter, maximum degree and the assortativtiy coefficient also seem to be important for some methods. In the case of real networks, the assortativity coefficient and the normalized diameter turn out to be the leading features, with the addition of the average clustering coefficient for the simple decision tree. When all of the networks are combined together and classified to fractal and non-fractal categories, not surprisingly, the set of important variables come together from the previous two cases. Namely, the normalized average path length, the assortativity coefficient, and the average clustering coefficient turn out to be the important characteristics, consistently for all algorithms.

From the results detailed above, we can conclude that the magnitude of distances in a network can have a connection to fractality. It may not be the distances between hubs, which influence fractality, but the distances between any of the nodes generally. Furthermore, although assortativity alone is not enough to separate fractal and non-fractal networks, together with other properties, it could contribute to the distinction. It sounds reasonable in general too that the joint presence of some properties causes the rising of another property, in this particular case, fractality.

Chapter 4

Summary

In this work, we investigated which characteristics could cause the emergence of fractal scaling in complex networks. Our analyses relied on a large dataset of both real-world and model generated networks, in order to prevent making conclusions based on coincidences. Concerning the disassortativity of fractal networks, we have found that although most of the considered mathematical models suggest that fractality is in a strong relation with disassortativity, there is also one, which supports the reverse of the statement. Consequently, we can conclude that the fractality of a network is independent of its assortativity, which is suggested by the real networks as well. Similar observations could be made in the case where hub repulsion was measured directly. However, a modification on the proposed Hub Repulsion Score should be made, to capture the property reliably for the full range of networks, which is planned for future work.

The possible connection of long-range anticorrelation to fractality was reviewed using three different methods proposed in the literature for capturing long-range patterns in networks. Just as in the case of neighbour-level anticorrelations, here we have found that correlation of node degrees is also independent of the fractal scaling. For all of the three methods we could find both models and examples of real networks, which support the suggestion of anticorrelation in fractal networks, even in the long-range scale, but there are numerous counterexamples as well. The suggestion of the connection of edge betweenness centrality with fractality was also reviewed. We examined whether fractal networks can or cannot possess edges with large betweenness centrality. We have come into the conclusion that fractal networks show no tendency to have edges mostly with small betweenness centrality, which statement is supported by almost all of the considered models and real networks as well.

Finally, we proposed a novel perspective for the study of the origins of fractality. We formulated a classification task with the goal to distinguish fractal and non-fractal networks based on other network properties. We solved the problem with state-of-the-art machine learning algorithms and identified the characteristics with high distinguishing ability. By these results, we suggest that the joint presence of different properties may be necessary for the emergence of fractality in a network. Furthermore, a feature related to the average distance of a network is possibly one of the essential properties in recognizing fractal scaling.

An important direction of further studies is to directly examine the possible connections of the proposed joint properties to fractality. For these analyses the extension of the dataset with additional models and real networks may be reasonable. Another interesting question, which is loosely connected to the previously mentioned problem, is if there is a conflicting relation between fractality and the small-world property. This suggestion was first proposed in [41], but since then examples have been shown for networks, which are fractal and small-world at the same time, for example in [4]. Thus, it seems that the two properties do not necessarily contradict each other however, average distances may still have connection to fractality, by taking into consideration other characteristics as well.

Bibliography

[1] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[2] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[3] C. Song, S. Havlin, and H. A. Makse, "Self-similarity of complex networks," *Nature*, vol. 433, no. 7024, pp. 392–395, 2005.

[4] C. Song, S. Havlin, and H. A. Makse, "Origins of fractality in the growth of complex networks," *Nature physics*, vol. 2, no. 4, pp. 275–281, 2006.

[5] L. K. Gallos, C. Song, and H. A. Makse, "Scaling of degree correlations and its influence on diffusion in scale-free networks," *Physical review letters*, vol. 100, no. 24, p. 248701, 2008.

[6] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[7] M. E. Newman, "Assortative mixing in networks," *Physical review letters*, vol. 89, no. 20, p. 208701, 2002.

[8] C. Song, L. K. Gallos, S. Havlin, and H. A. Makse, "How to calculate the fractal dimension of a complex network: the box covering algorithm," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, no. 03, p. P03006, 2007.

[9] P. T. Kovács, M. Nagy, and R. Molontay, "Comparative analysis of box-covering algorithms for fractal networks," *Applied Network Science*, vol. 6, no. 73, 2021.

[10] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[11] H. D. Rozenfeld, L. K. Gallos, C. Song, and H. A. Makse, "Fractal and transfractal scale-free networks," in *Encyclopedia of Complexity and Systems Science*, pp. 3924–3943, Springer, 2009.

[12] K. Takemoto, "Metabolic networks are almost nonfractal: A comprehensive evaluation," *Physical Review E*, vol. 90, no. 2, p. 022802, 2014.

[13] T. Akiba, K. Nakamura, and T. Takaguchi, "Fractality of massive graphs: Scalable analysis with sketch-based box-covering algorithm," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pp. 769–774, IEEE, 2016.

[14] M. Nagy, "Data-driven analysis of fractality and other characteristics of complex networks," *Master's Thesis*, 2018.

[15] H. D. Rozenfeld, S. Havlin, and D. Ben-Avraham, "Fractal and transfractal recursive scale-free nets," *New Journal of Physics*, vol. 9, no. 6, p. 175, 2007.

[16] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "Biogrid: a general repository for interaction datasets," *Nucleic acids research*, vol. 34, no. suppl_1, pp. D535–D539, 2006. Available: `https://thebiogrid.org`.

[17] A. Clauset, E. Tucker, and M. Sainz, "The colorado index of complex networks," *URL https://icon. colorado. edu*, 2016.

[18] J. G. D. Vázquez and R. Naik, "Interaction web database." `https://iwdb.nceas.ucsb.edu/resources.html`, 2003.

[19] J. Kunegis, "KONECT – The Koblenz Network Collection." `http://konect.uni-koblenz.de`, 2013. Accessed: 2020-06.

[20] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015. `http://networkrepository.com`.

[21] A. Barabási, *Network Science*. Cambridge University Press, 2016. Available: `http://networksciencebook.com`.

[22] N. Kasthuri and J. Lichtman, "Neurodata's graph database." https://neurodata.io/project/connectomes/, 2008.

[23] Transportation Networks for Research Core Team, "Transportation networks for research." https://github.com/bstabler/TransportationNetworks.

[24] The Centre for Water Systems (CWS) at the University of Exeter, "Centre for water systems." http://emps.exeter.ac.uk/engineering/research/cws/resources/benchmarks/.

[25] A. Cho, J. Shin, S. Hwang, C. Kim, H. Shim, H. Kim, H. Kim, and I. Lee, "WormNet v3: a network-assisted hypothesis-generating server for Caenorhabditis elegans," *Nucleic Acids Research*, vol. 42, no. W1, pp. W76–W82, 2014. Available: https://www.inetbio.org/wormnet/.

[26] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, *et al.*, "Topological structure analysis of the protein–protein interaction network in budding yeast," *Nucleic acids research*, vol. 31, no. 9, pp. 2443–2450, 2003. Data: Vladimir Batagelj and Andrej Mrvar (2006): Pajek datasets. http://vlado.fmf.uni-lj.si/pub/networks/data/.

[27] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, no. 6833, pp. 41–42, 2001. Data: http://moreno.ss.uci.edu/data.html#pro-pro.

[28] M. Nagy and R. Molontay, "Data-driven analysis of complex networks and their model-generated counterparts," *arXiv preprint arXiv:1810.08498*, 2018. Data: https://github.com/marcessz/Complex-Networks.

[29] S.-H. Yook, F. Radicchi, and H. Meyer-Ortmanns, "Self-similar scale-free networks and disassortativity," *Physical Review E*, vol. 72, no. 4, p. 045105, 2005.

[30] Z.-Z. Zhang, S.-G. Zhou, and T. Zou, "Self-similarity, small-world, scale-free scaling, disassortativity, and robustness in hierarchical lattices," *The European Physical Journal B*, vol. 56, no. 3, pp. 259–271, 2007.

[31] L. Kuang, B. Zheng, D. Li, Y. Li, and Y. Sun, "A fractal and scale-free model of complex networks with hub attraction behaviors," *Science China Information Sciences*, vol. 58, no. 1, pp. 1–10, 2015.

[32] Y. Fujiki, S. Mizutaka, and K. Yakubo, "Fractality and degree correlations in scale-free networks," *The European Physical Journal B*, vol. 90, no. 7, pp. 1–9, 2017.

[33] Z.-W. Wei and B.-H. Wang, "Emergence of fractal scaling in complex networks," *Physical Review E*, vol. 94, no. 3, p. 032309, 2016.

[34] D. Rybski, H. D. Rozenfeld, and J. P. Kropp, "Quantifying long-range correlations in complex networks beyond nearest neighbors," *EPL (Europhysics Letters)*, vol. 90, no. 2, p. 28002, 2010.

[35] M. Mayo, A. Abdelzaher, and P. Ghosh, "Long-range degree correlations in complex networks," *Computational Social Networks*, vol. 2, no. 1, pp. 1–13, 2015.

[36] J. Alstott, E. Bullmore, and D. Plenz, "powerlaw: a python package for analysis of heavy-tailed distributions," *PloS one*, vol. 9, no. 1, p. e85777, 2014.

[37] M. Nagy and R. Molontay, "On the structural properties of social networks and their measurement-calibrated synthetic counterparts," in *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 584–588, IEEE, 2019.

[38] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*. Pearson Education India, 2016.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[40] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), pp. 785–794, ACM, 2016.

[41] G. Csányi and B. Szendrői, "Fractal–small-world dichotomy in real-world networks," *Physical Review E*, vol. 70, no. 1, p. 016122, 2004.