

On the Use of Scale-Free Networks for Information Network Modelling

Claudia Hauff and Andreas Nürnberger
Otto-von-Guericke-University of Magdeburg
Faculty of Computer Science
Universitätsplatz 2, 39106 Magdeburg, Germany
Phone: +49-391-67-18487, Fax: +49-391-67-12018
email: claudiahauff@web.de, nuernb@iws.cs.uni-magdeburg.de

ABSTRACT: The notion of scale-free networks was created in the late 1990s when researchers at the University of Notre Dame investigated the distribution of the number of links connected to a node, i.e. the degree distribution of a network. For many real-world networks, they found a strong deviation from the random network model based on random graphs that had prevailed in network modelling until then. This finding led to the development of a novel approach which models the evolution of a network and is able to reproduce the observed degree distribution. In this paper, we briefly review the underlying principles of random graphs and scale-free networks. Furthermore, we give an overview of real world examples of scale-free networks from a biological and applicational perspective and present briefly an application of scale-free networks for the analysis of two types of information networks.

KEYWORDS: scale-free networks, information networks, network growth, network ageing

INTRODUCTION

The notion of scale-free networks was created in the late 1990s when researchers at the University of Notre Dame investigated the distribution of the number of links k connected to a node, i.e. the degree distribution $P(k)$ of a network, of real networks such as the World-Wide Web or metabolic networks. In the World-Wide Web, web pages are the nodes and hyperlink references represent the links while in metabolic networks substrates are the nodes and a link exists between two nodes when the corresponding substrates interact. With $P(k) \sim k^{-\alpha}$ they found a strong deviation from the random network model by Erdős and Rényi [17, 18, 19] (which predicts a Poisson distribution) that had prevailed in network modelling until then. This finding led to the development of a novel approach which models the evolution of a network and is able to reproduce the observed degree distribution. The model is based on two principles: growth and preferential attachment. At each time step, a new node enters the network and connects to a specific number of nodes already present in the network. This simple model was later extended to take into account actions such as the removal of links and nodes, link rewiring as well as aging of nodes. This kind of network construction leads to a characteristic degree distribution defining scale-free behaviour. Networks exhibiting this behaviour can be found in the field of technology (the World-Wide Web, the Internet), in social networks (citation networks, actor networks) and in many biological and chemical networks.

In the following, we first introduce static network models based on random graphs as defined by Erdős and Rényi. Then we review the underlying principles of scale free networks and discuss real world examples from a biological and applicational perspective. Finally, we present an application of scale free networks for the analysis of two types of information networks: a subset of the Web [24] and a subset of the citation network of scientific publications.

STATIC NETWORK MODELS

Creating models that offer a good approximation of real networks is one of the aims of network modelling. Only a meaningful model provides a good basis for retrieving useful information from it. A well-known example of the importance of model accuracy is the modelling of the Internet in order to evaluate new routing protocols which shall increase the Internet's efficiency. Clearly, it is not feasible to test new protocols on the Internet directly and instead simulators need to be used. The results obtained with the simulators will only hold for the Internet if the model is correct.

Between 1959 and 1961 two Hungarian mathematicians, Paul Erdős and Alfréd Rényi, developed a random graph model [17, 18, 19] (short ER model). By introducing probabilities into graph theory they were able to investigate the properties of large graphs, something had not been possible before. In classic graph theory the research focuses on graphs with a relatively small number of nodes and edges. For decades, the ER model was the state of the art in network modelling, primarily because few was known about the topology of large real-world networks. It was assumed that the ER model yields a reasonable estimate. In the 1990s however, this started to change and the value of the ER model for modelling real-world networks was questioned, as more and more data on large real-world networks was made available digitally.

The discovery that the World Wide Web exhibits a very different behaviour from that predicted by the ER model [9], renewed the interest in network science and initiated many research projects. Meanwhile, it is generally accepted that many systems and processes (both artificial and natural) can be seen as specific type of network and thus follow certain laws of networks. In 1998 and 1999 respectively, two new network models were developed that tackle different shortcomings of the ER model: the Small-World network model by Watts and Strogatz [37] (WS model) and the Barabási-Albert model (BA model), the first and simplest scale-free network model, which is presented in the Section 3 on scale-free networks.

The ER and the WS model, which we discuss first in the following, are classified as *static* network models since they assume a fixed number of nodes, while in the scale-free network models the number of nodes changes over time.

ERDŐS – RÉNYI MODEL

Erdős and Rényi defined models of static networks based on the theory of random graphs. Therefore, they provided two definitions of a random graph:

- (1) A probability space over all graphs with n nodes and m edges is defined and each graph $G(n,m)$ has the same probability of being drawn at random.
- (2) Alternatively, a random graph $G(n,p)$ can be generated by carrying out Bernoulli experiments. Given a fixed number n of distinct nodes, for each possible pair of nodes a Bernoulli experiment is performed: with probability p an edge is added between the node pair and hence with probability $1-p$ it is not.

The two formulations are equivalent. We will concentrate on the second and determine several network statistics from it. Since parallel edges and loops are forbidden, $n(n-1)/2$ node pairs are tested. Thus, the resulting graph contains approximately $p[n(n-1)/2]$ edges (each edge counts twice towards the total degree). The average degree \bar{k} of the graph takes the form $\bar{k} = p[n(n-1)/n] = p(n-1)$. The average shortest path length of a random graph scales logarithmically

$$\ell_{rand} \propto \frac{\ln n}{\ln \bar{k}}. \quad (1)$$

This is a crude estimate: about \bar{k}^a nodes are within distance a from a node. Given that in total there are n nodes, $\bar{k}^\ell \propto n$ and solving for ℓ yields Equation (1). What this means is, that even in large networks, the distance between any two nodes is small (compare with Figure 2 (c)).

It was proofed by Bollobás [12], that the degree distribution of a random graph constructed in such a way follows a Poisson distribution:

$$P(k) = \frac{e^{-\bar{k}} \bar{k}^k}{k!}. \quad (2)$$

The deviation from the characteristic degree \bar{k} is small and the probability that a node has a considerably higher or lower degree than \bar{k} approaches zero.

The *generalized random network model* is more flexible: it is able to produce graphs with arbitrary degree distributions. The construction procedure is as follows: given a fixed number n of nodes and a degree distribution $P(k)$, for each node v its degree $d(v)$ is derived from the distribution function $P(k)$ and v receives $d(v)$ stubs. In a second step, randomly chosen stubs of distinct nodes are connected.

SMALL-WORLD NETWORKS

The ER model has an inherent problem when simulating large real-world networks that was long ignored. The probability of two nodes being connected is independent of the connections among all other nodes. This issue was addressed by Watts and Strogatz [37]. They argue that in real networks there is a dependence between edges. As an example, they reasoned that in social networks with people represented by nodes and edges indicating acquaintanceship or friendship, the probability that two people with a common friend know each other is higher than without a common acquaintance. In order to measure this phenomenon, the *clustering coefficient* C of a network was introduced. It describes the probability that two nodes are connected when they have a common neighbour. The equation for the clustering coefficient of node v_i follows immediately:

$$C_i = \frac{2y}{z(z-1)}. \quad (3)$$

with y as the number of connections existing among v_i 's z nearest neighbours and $z(z-1)/2$ as the total number of possible connections between the neighbours (Figure 1). The clustering coefficient C of the entire network is the average over all nodes' clustering coefficients:

$$C = \frac{1}{|V(G)|} \sum_i C_i. \quad (4)$$

In a graph generated by the ER approach the clustering coefficient is $C = \bar{k} / n$, an extremely small value if the network size n is large and p is small.



Figure 1: Example of the clustering coefficient calculation. Node 1 has a clustering coefficient of 0, its two neighbours are not connected. The clustering coefficient of node 2 evaluates to $C_2=2/3$ with $y=2$ and $z=3$.

Watts and Strogatz proposed a procedure that generates networks with the desired attributes: an average shortest path length that scales logarithmically as in the ER model and a large clustering coefficient as seen in real networks. They propose the following: a fixed number of nodes n are laid out in a regular 1-D lattice and each node is connected to its $z \geq 4$ nearest neighbours (Figure 2 (a)). The clustering coefficient of the lattice is high but so is also the average shortest path length as one has to travel along the ring to reach more distant nodes. Then, with probability p each edge of the lattice is rewired to a randomly chosen node (Figure 2 (b)). Already for small p this introduces the necessary shortcuts to lower the average shortest path length to a value similar to the ER model while at the same time keeping the clustering coefficient high, because only a small number of the edges are rewired. When p reaches 1 (Figure 2 (c)) and very few of the original edges remain, the lattice dissolves into a random graph. Thus one can say, that small-world networks are generated by an interpolation between a regular lattice and a random graph. The resulting degree distribution decays exponentially. While in the original model the edges were rewired, similar results are achieved when new edges are added and randomly distributed.

SCALE-FREE NETWORKS

Albert-László Barabási and his research group at the University of Notre Dame were among the first to note the deviation of real networks from the Erdős-Rényi model and also from the Watts-Strogatz model. While the clustering coefficient and the average shortest path length behave similarly to the Watts-Strogatz model, the degree distribution $P(k)$ is different and follows a power law form with degree exponent γ :

$$P(k) \propto k^{-\gamma}. \quad (5)$$

What exactly does this mean? In the two models introduced above the degree of any node is not very dissimilar from the average degree \bar{k} and thus the network has a characteristic degree. A network with a power-law distribution lacks

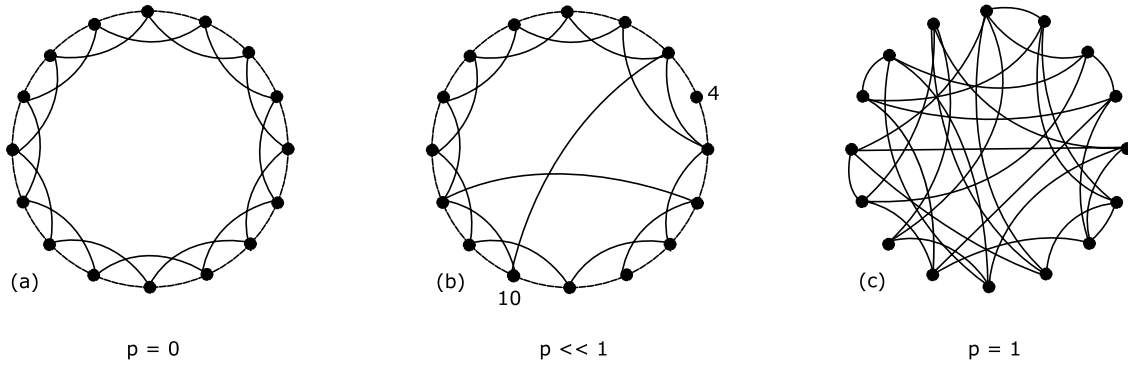


Figure 2: Construction procedure of the Small-World network model. (a) The starting point is a regular lattice with in this case $z = 4$ nearest neighbors per node. Each node's clustering coefficient is equal to $C_i = 1/2$. (b) Rewiring edges with a small probability p decreases $\bar{\ell}$ while keeping C high. The two labelled nodes 4 and 10 are now only a distance of 2 away due to the introduction of a single shortcut. In the regular lattice their distance was 3. (c) Increasing p to 1 results in the creation of a random network with low ℓ and low C .

such a characteristic degree. There are a large number of nodes with very few connections and a small number of nodes with a very high degree. Barabási and his collaborators identified two necessary conditions for the creation of network models with a power law degree distribution: *growth* and *preferential attachment*. The network models are no longer static, but dynamic. The numbers of nodes and edges increase over time and the attachment of edges to nodes occurs preferentially. This concept is best explained with an example: the World Wide Web, a scale-free network, is continuously growing by the addition of new web pages (the nodes of the network) which contain hyperlinks pointing to other pages (the edges). Most of those hyperlinks will point to well-known and highly connected pages such as Google.com or Amazon.com for instance while very few hyperlinks will link to unknown pages. The principle of 'popularity is attractive' applies here: a highly connected web site has a high probability of receiving further connections and is connected to *preferentially*.

EXAMPLES

Scale-free networks can be found in many different areas. Social networks were among the first to be investigated. Citation networks [29], actor networks [8] and scientific collaboration networks [7, 31, 36] belong into this category. Two of the technological networks examined are the World Wide Web [5, 9] and the Internet [20, 40]. Biological networks have aroused a lot of interest lately, they include metabolic networks [21, 25, 30], protein-protein interaction networks [26, 35], protein domain networks [38] and gene expression networks [2, 11]. Other networks where one would not readily suspect a (scale-free) network structure are the web of words [22], the network of sun corona explosions [33], the network of earthquakes [1] and the medieval inquisition [32]. This list is only a small excerpt, but it hopefully shows that scale-free networks appear in manifold domains.

Discovering a scale-free network in real-world data is not the exception anymore. On the contrary, it seems that a scale-free network structure can be extracted from nearly every data set available as even comic book characters are not safe from being investigated [3]. Having said that, it has to be pointed out that there are different opinions within the community as to when a data set exhibits scale-free behaviour. A very simple method is to plot the degree distribution on a log-log plot and if the distribution follows a straight line a power law distribution is present. A given empirically estimated degree distribution might be classified by one research group as a power law whilst another classifies it as a stretched exponential degree distribution. In several cases an examination of a single data set by different research groups yielded widely different estimations of the degree exponent. A major problem is the lack of standardized methods for the detection of a scale-free behaviour and the subsequent calculation of the degree exponent γ . In [27, 34] for example it is claimed that the biological networks whose degree distributions have been investigated so far often do not exhibit pure power-laws, although some are found to follow a power-law with an exponential cut-off. The authors attribute the different findings to the simple mathematical methods used until now to calculate the degree exponent such as linear fit of the data on a log-log plot. They argue that the more accurate maximum likelihood method produces different results. A similar argument is presented in [23]. In the next section, the metabolic network is presented in more detail as a network that was found to be scale-free.

The Metabolic Network

A metabolic network describes the chemical changes within an organism or cell that occur in order to produce energy and materials needed to sustain the organism's or cell's life processes. In general, such a network is developed as follows: the substances on which enzymes act (the so-called substrates) form the nodes of the network and a connection is established between two nodes if a metabolic reaction occurs with both of the node's corresponding substrates taking part. Fell et al. [21] did not distinguish between reaction educts and reaction products and therefore applied undirected links. They analyzed the structure of the core metabolic network (275 substrates) of the bacterium *Escherichia coli* and reported that the probability of a substrate to take part in k reactions follows a power-law. While they discarded common coenzymes such as ATP due to their ubiquitous nature, Jeong et al [25] included them in their investigation into the metabolic networks of 43 different organisms from all three domains of life: archaea, bacteria and eukaryotes. They constructed a directed bipartite graph for each organism from data provided by the WIT database, a pathway-genome database which allows the prediction of a given metabolic pathway based on the annotated genome and biochemical knowledge.

In Figure 3 (a) a part of the *Escherichia coli*'s metabolic network and its metabolic network degree distributions are displayed. As stated before, the substrates are represented by nodes but here they are not directly linked with each other, instead they are connected to educt-educt complexes (indicated by the black boxes; the acting enzymes are given by their enzyme commission number). An outgoing link indicates an educt for a reaction and an incoming link denotes a product which emerges as a new node of the network. For instance, ATP and D-ribose 5-phosphate are the input substrates for a metabolic reaction catalyzed by the enzyme 2.7.6.1 which produces AMP and 5-phosphate-alpha-D-ribose 1-phosphate. The in-degree distribution (the probability that a substrate is produced by k reactions) and the out-degree distribution (the probability that a substrate participates as educt in k reactions) both follow a power-law.

In Figure 3 (b) the in- and out-degree distributions of *Escherichia coli* are illustrated with an estimated degree exponent of $\gamma_{in} = \gamma_{out} = 2.2$. The results for all 43 organisms were similar to the ones presented here; their degree distributions all follow a power-law. While the top connected substrates were nearly identical for all studied organisms, species-specific differences were detectable among the lowly ranked substrates. However, the authors themselves pointed out that the data used was incomplete and probably erroneous – from the 43 organisms only 25 had been fully sequenced, and a sizable portion of potential protein encoding gene sequences had not been functionally assigned.

A study with an improved data set was later carried out by Ma et al. [30] who investigated 65 fully sequenced organisms and improved the metabolic network data by manually cleaning it. Similar to Fell et al. they did not take into account metabolites such as ATP and constructed edges between interacting substrates: an undirected edge in the case of a reversible reaction and a directed edge in the case of an irreversible reaction. Four major subsets within the metabolic network were identified by them: a product subset, a substrate subset, an isolated subset and a fully connected subset with usually less than a third of all nodes. The latter one is the core of the network and follows a power-law for all organisms. In contrast to Jeong et al. they also found different average path lengths for the three domains of life and attributed it to the removal of the common coenzymes before the construction of the network, since they participate in so many metabolic reactions that a short path between most substrates can be constructed through them.

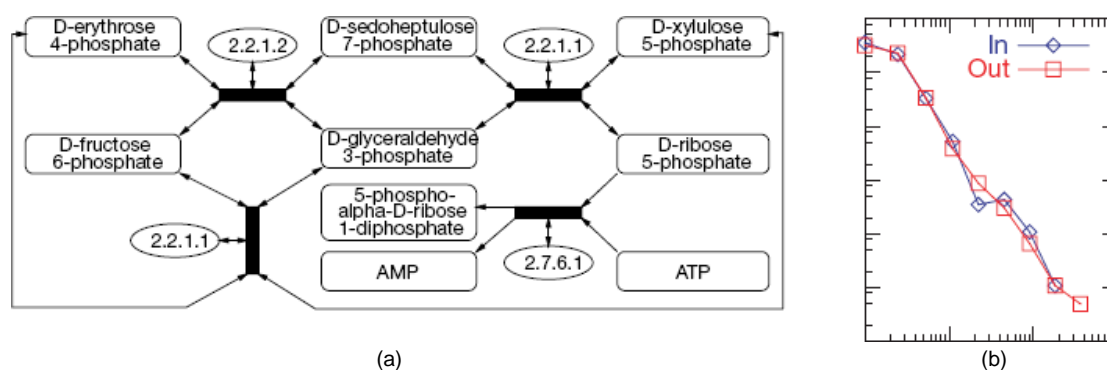


Figure 3: (a) Part of the bipartite metabolic reaction graph from *Escherichia coli*. The white boxes represent the substrates, the black boxes the educt-educt complexes and the ellipses the enzymes. (b) The in- and out-degree distribution of *Escherichia coli*. (Figures (a) and (b) taken from [1] copyrighted by Nature; Reproduction permitted by Nature Publishing Group <http://www.nature.com/>).

Simulations of random failures of nodes showed a robustness that is characteristic for scale-free networks [4]. Randomly removing a number of substrates of *Escherichia coli*'s metabolic network does not alter the diameter of the network as demonstrated by Jeong et al. in computer simulations. A similar conclusion was reached by Edwards et al. [16] who conducted a mutagenesis study of *Escherichia coli*: "a large number of the central metabolic genes can be removed without eliminating the capability of the metabolic network to support growth".

THE BARABÁSI-ALBERT MODEL

The first model developed to explain the observed power-law degree distributions was the BA model. As stated before, its two main elements are *growth* and *preferential attachment*. It is constructed by continuously injecting nodes and edges into the network. Starting with a small number n_0 of nodes, at each time step one node with m ($m \leq n_0$) undirected edges attached to it joins the network. The free ends of the new edges are distributed preferentially among the nodes already in the network. Formally, the probability p that node s ($s \leq t$) which entered the network at time s with degree $k(s,t)$ receives a new edge at time t is written as

$$p = \frac{k(s,t)}{\sum_u k(u,t)}. \quad (6)$$

Barabási and Albert performed extensive numerical simulations with the simple model described above and observed that a network created by repeating the two steps growth and preferential attachment until t is sufficiently large, results in a power-law degree distribution with a fixed degree exponent $\gamma = 3$.

Continuum Approach

Barabási and Albert also developed an analytical approach for their model. The degree k of a node, although a discrete variable, is here treated as a continuous variable. The degree distribution $P(k)$ is derived for an exemplified undirected network constructed according to the BA model to which at each time step one node and one edge is added.

At time t the total degree of the network is equal to $2t$, as there are t edges in the network. From Equation (6) we see that the change $\partial k(s,t)/\partial t$ of degree $k(s,t)$ at time t is proportional to

$$\frac{\partial k(s,t)}{\partial t} = \frac{k(s,t)}{2t}. \quad (7)$$

Equation (7) is a differential equation with the initial condition $k(s=t, t) = 1$. A node that enters the network at time $s=t$ has one edge attached to it and hence a degree of one. Solving the equation yields the mean degree of a node s at time t :

$$k(s,t) = \left(\frac{s}{t}\right)^{\frac{1}{2}}. \quad (8)$$

The degree distribution $P(k,t)$ at time t is proportional to $\partial s(k,t)/\partial k$ and with $-1/t$ as normalization term, it can be derived by solving Equation (8) for $s(k,t)$:

$$P(k,t) = -\frac{1}{t} \frac{\partial s(k,t)}{\partial k} = \frac{2}{k^3} \propto k^{-3}. \quad (9)$$

This confirms the results obtained in the numerical simulations. In the case of employing a uniform attachment function, the above calculations will lead to an exponential degree distribution. Therefore, preferential attachment is a necessary condition to reach a power-law degree distribution.

By passing $k(s,t)$ to the continuum limit, a very convenient method was found to derive $P(k)$. The *Master Equation Approach* [14] and the *Rate Equation Approach* [28] avoid this simplification and derive $P(k)$ utilizing discrete linear equations. In [14] the exact solution for the Barabási-Albert model is presented. Whilst the calculations are considerably more complex, the obtained results are very similar to those reached when applying the continuum approach.

EXTENSIONS

The degree exponents estimated from real world data usually do not reach the degree exponent that is predicted by the BA model. This is not really surprising when comparing the actions possible in real networks with the actions the BA model accounts for (node and edge addition only). The Internet for instance allows the deletion of nodes (routers are taken off the network), the rewiring or deletion of edges (the physical connections between routers are changed) and the connection of two old nodes (previously unconnected routers are connected). A second drawback is the undirected nature of the BA model and as demonstrated in the metabolic network example, directed networks also regularly occur.

To account for these drawbacks, several network models have been proposed that extend the BA model. A directed network model with variable in- and out-degree exponents in the range $(2, \infty)$ was proposed by Dorogovtsev et al. [14]. In such a model, in- and out-degree are determined separately. For the in-degree the target ends of the edges are of importance, while the source ends, which can be anywhere within or outside the network, are ignored. The construction procedure is very similar to the BA model: at each time step one node and m edges with the target ends distributed preferentially enter the network. Since a node enters the network without incoming links attached, the preferential attachment function from Equation (6) needs to be altered, otherwise a new node has a probability of zero to receive incoming links in future. The probability of attracting incoming links is now proportional to $k_{in} + A$ with $A > 0$. The term A is called the *initial* or *additional* attractiveness of a node. Applying the continuum approach again, the mean in-degree becomes

$$k_{in}(s, t) = A \left(\frac{s}{t} \right)^{-\beta} - A \quad (10)$$

with $\beta = 1/(1+a)$ and $a=A/m$ the in-degree distribution $P(k)$ has the form

$$P(k) \propto k^{-(1+1/\beta)} = k^{-(2+a)}. \quad (11)$$

Network models with a degree exponent below 2 can be explained by *accelerated growth* [15]. A network exhibits accelerated growth when its number of edges grows faster than the number of nodes. The non-stationary average degree follows a power law $c_0 t^b$, where b is called the growth exponent. It is clear that $b < 1$, otherwise the average degree would increase indefinitely. There are two general processes that lead to accelerated growth. In the first place when the network grows, the number of links a new node enters the network with can also grow. This is the case, for example, in the World Wide Web, where the amount of pages increases over time, there is more to cite and hence, the average number of links on a page increases. A second possibility is the addition of new links between old nodes. The actor and collaboration networks can be named here. Two established actors or authors may be working together for the first time and therewith a new link is established between them without the introduction of a new node.

Other factors that were introduced include the ageing and physical limitations of nodes [6], weighted edges [39, 10] and random linkage to name just a few. Ageing of nodes is an important factor in networks involving humans. In such networks nodes cannot receive an unlimited number of edges and usually the ability to attract edges decreases as a function of the node's age until it has reached zero. As an actor for instance gets older he is very likely to participate in fewer film projects and obviously when he dies, the respective node in the actor network stops receiving any new edges. Not only social networks have underlying limitations, technological networks can also be restricted. On the Internet, a router cannot attach to an unlimited number of other routers; the bandwidth is here the limiting factor. Weighted edges [39] are a generalization of the binary relationships between nodes considered so far. The different attachment strength between nodes naturally leads to networks with weighted edges. The meaning of the weights depends on the specific network under consideration. On the Internet the load between routers can be taken as a weight, in social networks it could be the degree of acquaintanceship between two people and in metabolic networks the number of reactions between two substrates.

APPLICATIONS

The gain in knowledge about the true structure of complex networks has so far been rarely exploited. Much work has been concentrated on developing network models that resemble real networks as closely as possible, but few applications have been proposed that make use of that additional knowledge. One exception is the research in eradicating epidemics where the knowledge has been applied to identify highly connected nodes that should be treated first in order to decrease a virus' spreading rate [13].

We adopt a different approach and hypothesize that it is possible to gain valuable information by comparing a real-world network with its corresponding network model. The model is created from statistics derived from the real-world network such as the age of the nodes, the network size t , the average degree m and the degree exponent γ . While the degree distribution of the model and the real-world network will be the same or at least very similar, on the individual node level the degrees will almost certainly be different. Then, for each node s of the network the expected number of links, that is, the degree of the node s in the model at time t (Equation (8) for the BA model), is calculated. It is compared with the actual number of links of the node in the real-world network and conclusions are drawn from it. A node with more links than expected is said to be of above average importance, while a node with fewer links than expected is less important. This idea has so far been tested on the WT2g collection [24], a subset of the World Wide Web, which contains almost 250.000 web pages and was crawled in 1997. In contrast to traditional link structure algorithms such as PageRank which calculates importance scores based on a snapshot of the Web, scale-free network models offer a principled approach to take the age of a web page into account. A web page for example that has accumulated 1000 incoming hyperlinks within 2 months after it was uploaded is of more importance in our approach than a web page that has the same 1000 pages referencing it, but was uploaded two years ago. PageRank would assign the same score to both web pages. For each document of the WT2g collection, given its age, the actual and expected numbers of hyperlinks were calculated and an importance scores was derived. These scores were then converted to a probability distribution and used as a document prior in the language modelling approach to information retrieval. Experiments were performed on the ad-hoc task with the titles TREC topics 401-450. The results obtained showed small improvements (1.37%) over the uniform document prior and large gains when compared to traditional link structure algorithms that consider only the link structure and ignore the age of the documents. However, due to the poor statistics available from the data – the age of each document could often not be determined correctly – the study can only be seen as preliminary and the obtained results remain inconclusive.

The same idea can be applied to the citation network of scientific publications where importance scores are derived for each publication based on the actual number of papers referencing it and the expected number of incoming references. A tool has been developed that visualizes a subset of the citation network and the corresponding importance scores depending on the user's query. It is hoped that such scores support users by their search for scientific publications. The first evaluations are very promising. However, a more detailed study is still necessary. A screenshot of the tool is given in Figure 4 (<http://www.uni-magdeburg.de/hauff/vipf>).

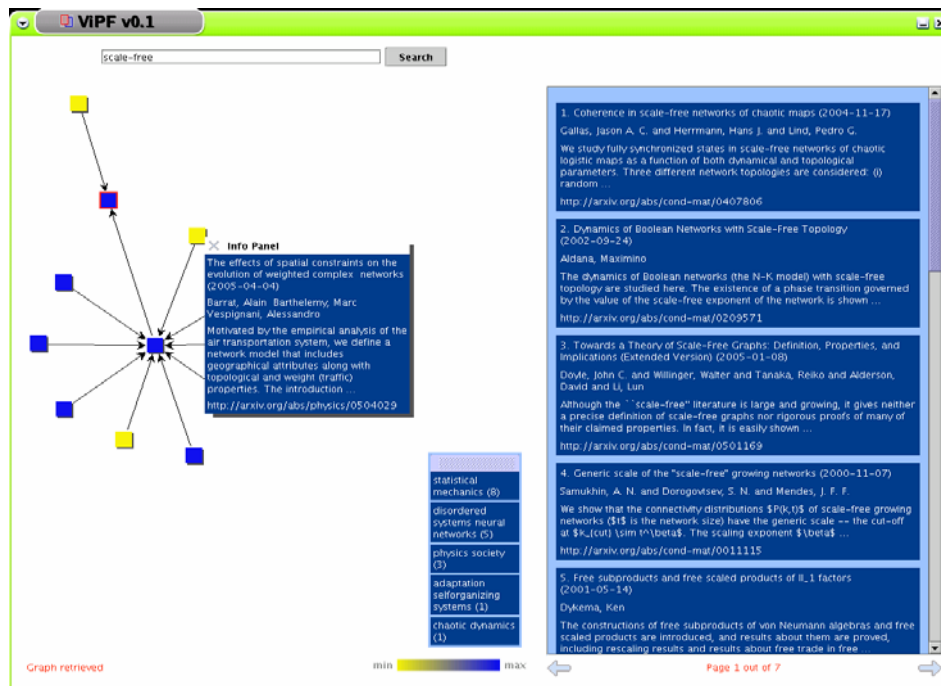


Figure 4: A screenshot of the tool ViPF designed to support the user in his search for scientific publications. Apart from the retrieval of papers by content (right panel) for each paper a part of its citation graph is visualized (left panel). The colour of a node indicates its fitness value which is derived from the comparison actual vs. expected number of citations pointing to the paper. A double-click on an arbitrary node opens a window with further information (title, authors, abstract, link) about the respective paper.

CONCLUSIONS

In this paper, we motivated and critically discussed the basic principles of network modelling approaches that are able to correctly model many natural and artificial real world networks. Despite the advances in the theory of scale-free networks though, few applications have been proposed. Here, we have introduced a simple idea of how to utilize the information gained by comparing properties of real-world networks with the formal (idealized) model derived from it in order to obtain further information about specific nodes of the network

REFERENCES

- [1] S. Abe, and N. Suzuki, Scale-free network of earthquakes, *Europhys. Lett.*, 65(4), pp. 581-586, 2004.
- [2] H. Agrawal, Extreme Self-Organization in Networks Constructed from Gene Expression Data., *Phys. Rev. Lett.*, 89, 268702, 2002.
- [3] R. Alberich, J. Miro-Julia, and F. Rosselló, Marvel Universe looks almost like a real social network, *preprint cond-mat/0202174*, 2002.
- [4] R. Albert, H. Jeong, and A. Barabási, Error and attack tolerance of complex networks, *Nature*, 406, pp. 378-382, 2000.
- [5] R. Albert, H. Jeong, and A. Barabási, Diameter of the World Wide Web, *Nature*, 401, pp. 130-131, 1999.
- [6] L. Amaral, A. Scala, M. Barthélémy, and H. Stanley, Classes of small-world networks, *Proc Natl Acad Sci U S A*, 97(21), pp. 11149-11152, 2000.
- [7] A. Barabási, H. Jeong, R. Ravasz, Z. Néda, T. Vicsek, and A. Schubert, On the topology of the scientific collaboration networks, *Physica A*, 311, pp. 590-614, 2002.
- [8] A. Barabási, and R. Albert, Emergence of Scaling in Random Networks, *Science*, 286, pp. 509-512, 1999.
- [9] A. Barabási, R. Albert, and H. Jeong, Scale-free characteristics of random networks: the topology of the world-wide web, *Physica A*, 281, pp. 69-77, 2000.
- [10] A. Barrat, M. Barthélemy, and A. Vespignani, Weighted Evolving Networks: Coupling Topology and Weight Dynamics, *Phys. Rev. Lett.*, 92, 228701, 2004.
- [11] S. Bergmann, J. Ihmels, and N. Barkai, Similarities and differences in genome-wide expression data of six organisms, *PLoS Biol.*, 2(1), pp. 85-93, 2004.
- [12] B. Bollobas, Random graphs, *Combinatorics, Proceedings Swansea 1981, London Math. Soc. Lecture Note Ser.* 52, pp. 80-102, 1981.
- [13] Z. Dezső, and A. Barabási, Halting viruses in scale-free networks, *Phys. Rev. E*, 65, 055103, 2002.
- [14] S. Dorogovtsev, J. Mendes, and A. Samukhin, Structure of growing networks: Exact Solution of the Barabási-Albert model, *Phys. Rev. Lett.*, 85, pp. 4633-4636, 2000.
- [15] S. Dorogovtsev, and J. Mendes, Accelerated growth of networks, *Contribution to 'Handbook of Graphs and Networks: From the Genome to the Internet'*, 2002.
- [16] J. Edwards, and B. Palsson, The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities., *Proc Natl Acad Sci U S A*, 97(10), pp. 5528-5533, 2000.
- [17] P. Erdős, and A. Rényi, On random graphs, *Publicationes Mathematicae*, 6, pp. 290-297, 1959.
- [18] P. Erdős, and A. Rényi, On the evolution of random graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, pp. 17-61, 1960.
- [19] P. Erdős, and A. Rényi, On the strength of connectedness of a random graph, *Acta Mathematica Scientia Hungary*, 12, pp. 261-267, 1961.
- [20] M. Faloutsos, P. Faloutsos, and C. Faloutsos, On power-law relationships of the Internet topology, In: *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pp. 251-262, 1999.
- [21] D. Fell, and A. Wagner, The small world of metabolism, *Nat. Biotechnol.*, 18(11), pp. 1121-1122, 2000.

- [22] R. Ferrer-i-Cancho, and R. Sole, The small world of human language, *Proceedings of The Royal Society of London. Series B, Biological Sciences*, 268(1482), pp. 2261-2265, 2001.
- [23] M. Goldstein, S. Morris, and G. Yen, Problems with fitting to the power-law distribution, *Eur. Phys. J. B*, 41, pp. 255-258, 2004.
- [24] C. Hauff, and L. Azzopardi, Age Dependent Document Priors in Link Structure Analysis, In: *Proc. of 27th European Conference on IR Research (ECIR 2005)*, pp. 552-554, 2005.
- [25] H. Jeong, B. Tombor, R. Albert, Z. Oltvai, and A. Barabási, The large-scale organization of metabolic networks, *Nature*, 407, pp. 651-654, 2000.
- [26] H. Jeong, S. Mason, A. Barabási, and Z. Oltvai, Lethality and centrality in protein networks, *Nature*, 411, pp. 41-42, 2001.
- [27] R. Khanin, and E. Wit, How Scale-Free are Biological Networks, *preprint*, 2005.
- [28] P. Krapivsky, and S. Redner, Organization of growing random networks, *Phys. Rev. E*, 63, 2001.
- [29] S. Lehmann, B. Lautrup, and A. D. Jackson, Citation Networks in High Energy Physics, *Phys. Rev.*, 68, 026113, 2003.
- [30] H. Ma, and A. Zeng, The connectivity structure, giant strong component and centrality of metabolic networks, *Bioinformatics*, 19(11), pp. 1423-1430, 2003.
- [31] M. Newman, Coauthorship networks and patterns of scientific collaboration, *Proc Natl Acad Sci U S A*, 101, pp. 5200-5205, 2004.
- [32] P. Ormerod, and A. Roach, The Medieval inquisition: scale-free networks and the suppression of heresy, *eprint arXiv:cond-mat/0306031*, 339, pp. 645-652, 2004.
- [33] M. Paczuski, and D. Hughes, A heavenly example of scale-free networks and self-organized criticality, *Physica A*, 342, pp. 158-163, 2004.
- [34] M. Stumpf, P. Ingram, I. Nouvel, and C. Wiuf, Statistical model selection methods applied to biological networks, *Transactions in Computational Systems Biology*, 2005.
- [35] A. Wagner, The Yeast Protein Interaction Network Evolves Rapidly and Contains Few Redundant Duplicate Genes, *Mol. Biol. Evol.*, 18, pp. 1283-1292, 2001.
- [36] C. Wagner, Six case studies of international collaboration in science, *Scientometrics*, 62(1), pp. 3-26, 2005.
- [37] D. Watts, and S. Strogatz, Collective dynamics of 'small-world' networks, *Nature*, 393, pp. 440-442, 1998.
- [38] S. Wuchty, Scale-free Behavior in Protein Domain Networks, *Mol. Biol. Evol.*, 18(9), pp. 1694-1702, 2001.
- [39] S. Yook, H. Jeong, and A. Barabási, Weighted Evolving Networks, *Phys. Rev. Lett.*, 86(25), pp. 5835-5838, 2001.
- [40] S. Yook, H. Jeong, and A. Barabási, Modeling the Internet's large-scale topology, *Proceedings of the Nat'l Academy of Sciences* 99, pp. 13382-13386, 2002.