Cent. Eur. J. Comp. Sci. • 1(1) • 2011 • 137-153
DOI: 10.2478/s13537-011-0009-x

VERSITA

Central European Journal of **Computer Science**

Good versus optimal: Why network analytic methods need more systematic evaluation

Review Article

Katharina A. Zweig*

Network Analysis and Graph Theory, Interdisciplinary Center for Scientific Computing (IWR), Ruprecht-Karls-University Heidelberg, Speyerer Straße 6, 69118 Heidelberg, Germany

Received 02 Feb 2011; accepted 07 Mar 2011

Abstract: Network analytic method designed for the analysis of static networks promise to identify significant relational patterns that correlate with important structures in the complex system the network is derived from. In this mini review, three groups of network analytic methods are discussed: centrality indices, network motifs, and clustering algorithms. We show that so far these methods have mainly been used in a descriptive way, but that they show promising possibilities to be used for prediction and classification. We thus conclude the article with a discussion of how benchmark sets and evaluation criteria could look like to realize this promise.

Keywords: network analysis • network motifs • centrality indices • clustering algorithms • ground truth

1. Introduction

Being the social species that we are, the importance of relationships for our day-to-day life is evident, and as a species, we have developed a fine sense for the social position of ourselves and others. Network analysis is a set of tools to quantify this intuitive notion. Towards this aim, different measures have been developed to find, e.g., the most central vertices [46, 47], to extract a hierarchy on the vertices [19], to find tightly knit clusters [27], or to even predict likely future relationships [49]. One of the most important but also challenging aspects of this budding field is its inherent interdisciplinarity: the systematic use of the term *social network* in sociology started in the mid 1950s [28] together with the design of network analytic tools to understand the individual behavior of actors in a network, and the success or failure of certain individuals within it [67]; At the same time in **mathematics**, a special area of graph theory gained momentum, namely *extremal graph theory* and *random graph theory* [8], with first applications to questions concerning the structure of social networks [22]. **Computer science** meanwhile developed numerous algorithms and data structures for computing graph theoretic measures, and also laid the technical basis for two of the largest networks, the internet and the World Wide Web. At this time, **physics** had already perfected its understanding of systems composed of a myriad of individual particles that are – albeit – of a very simple nature like gas atoms or magnets. The global behavior of these

* *E-mail: katharina.zweig@iwr.uni-heidelberg.de*

so-called *complex systems* was studied in a field called *statistical physics*. Today it is acknowledged that all of these fields build the basis for network analysis, but only at the end of the last century two seminal papers combined all four perspectives into one [6, 68] and thus attracted the attention of a large range of scientists. The field promises to make the structure of complex systems observable by focusing on a *complex network* derived from it. That means, instead of trying to understand **all** types of relationships between **all** kinds of entities in something as complex as a human cell, a complex network is in most cases composed of only one type of entity and one kind of relationship beetween the instances of this entity. An example are so-called *protein-protein-interaction networks* in which two proteins are connected if there is evidence for close physical contact between them. The basic assumption is that the structure captured in this simple network might reveal information on something that is much more important but also much more difficult to obtain, namely the function of a protein in the cell. In this sense, the complex network is a *proxy* of the corresponding complex system and it is assumed that significant structures in the network correspond to significant structures or processes in the full system. Even if the correspondence between the structures were not perfect but rather noisy, many of the significant structures in the network are easy enough to compute to make even a weak correspondence profitable. Thus, if a certain level of correspondence can be established, network analysis promises a rather cheap way to identify candidate answers for those questions that would be too difficult to answer in the full complex system.

The buzz around network analysis, which is sometimes even called the *new science of networks* [5], is tremendous and the field has seen more than one hundred thousand publications using network analytic methods in the last decade[1]. The promise of the field to enlighten our understanding of fundamental human behavior and the delicate interwoven nature of biological systems has, however, not yet been accomplished. This is mainly caused by the fact that for many of the intuitive notions about typical network structures, many methods have been proposed to identify them, but a rigorous evaluation of when which method is applicable is still missing at large. Thus, many analyses of complex networks are merely descriptive and correlations of significant relational patterns with important structures in the corresponding complex system are established on an anecdotal basis.

Network analysis has evolved into a very broad field where, next to the analysis of static networks, the modelling of networks has become a focus [18], as well as the analysis of processes on complex networks [7, 25], or the design of networks with wanted features [66]. In this mini review we will focus on the structural analysis of static networks by discussing three of the main network analytic approaches, the intuition behind them, and the few approaches to validate their results. We will particularly discuss *bootstrap* and *permutation methods* which are used as a null-model to validate the significance of the identified network structures. We will also summarize findings that the choice of the most suitable null-model is not trivial and will influence the results heavily. We thus propose that network analysis not only needs statistical validation but more benchmark sets with a gold standard solution on which new network analytic methods can be tested.

The article is organized as follows: Section 2 provides necessary definitions. Section 3 gives a quick history of the evolution of network analysis as a descriptive tool box. It also discusses more detailed how a complex network is a proxy of a complex system and why that matters. Section 4 introduces the terms *ground truth* and *golden standard solutions* as applicable to network analysis and presents two successful examples, whose validation approaches can be generalized to other network analytic measures. Section 5 describes a selection of network analytic measures and some of the few approaches to validate them, which are not yet systematically applied. The examples show that evaluation of network analytic measures can be done and that there is a strong need to do it. We thus conclude the review with a discussion of possible ways to construct benchmark sets and suitable evaluation criteria for network analytic measures in Section 6.

2. Definitions

A *graph* $G = (V, E)$ is composed of a set of vertices V and a set of edges $E \subseteq V \times V$. Edges can be *directed* ($e = (x, y)$), i.e., represented by an ordered tuple, or *undirected* ($e = \{x, y\}$), indicated by an unordered set. Additionally, the edges

[1] *Google scholar shows more than* 128,000 *hits with the search term "network analysis" as of January 2011.*

can be weighted by a weight function $\omega : E \to \mathbb{R}$. A graph is said to be *bipartite* if there are two subsets of vertices V_1, V_2 with $V_1 \cup V_2 = V$ such that there is no edge between vertices from the same subset, i.e., $\forall\ x, y \in V_1,\ \ (x, y) \notin E$ and $\forall x, y \in V_2,\ \ (x, y) \notin E$. A subgraph $G' = (V', E')$ of G, denoted by $G' \subseteq G$ is composed of any subset of vertices $V' \subseteq V$ and a subset $E' \subset E$ of edges between vertices in V'. The *degree* $deg(v)$ of a vertex v is defined as the number of edges it is contained in. The *density* of a graph or subgraph is defined as the ratio between the number of edges in the (sub-)graph and the number of possible edges in the graph. Any ordered set of edges $(e_1 = (v_1, w_1), e_2 = (v_2, w_2), ..., e_k = (v_k, w_k))$ with $w_i = v_{i+1}, \forall 1 \leqslant i < k - 1$ is called a *path* $P(v_1, w_k)$ between v_1 and w_k. The *length* $L(P(x, y))$ *of a path* $P(x, y)$ is defined as the sum of the weights of the edges in it (weighted graph) or the number of edges in it (unweighted graph). The *distance* $d(v, w)$ between two nodes is defined as the minimal length of any path between them; if there is no such path it is defined to be ∞. Paths with minimal length between the vertices they connect are called *shortest paths*. A *random walk* on a graph is the result of a process in which a particle starts at some vertex v and chooses one of the neighboring vertices at random to which it then proceeds. In the simplest case all neighbors are chosen uniformly at random. A *clique* is a graph with n vertices in which every vertex is connected to every other vertex. A *clustering* of a network G is a set of subsets of vertices, called *clusters*. Clusterings can be differentiated by whether they allow for multiple membership (*overlapping clusters*) or not; in the latter case a *clustering* is simply a *partition* of the vertices in G.

3. Network analysis – an introduction

Network analysis has evolved from a set of methods to describe the individual role of all vertices in a small network to describe the universal, statistical behavior of vertices in a large and complex network. In this section we will briefly review this evolution and indicate how how the predictive power of the field's tools can be evolved.

The first decades of social network analysis were dominated by finding explanations for the individual role of certain network positions, and researchers concentrated on the detailed analysis of a few small, carefully created networks [67]. A typical example for such a network is the well-analyzed Florentian marriage network that displays the marital relationships between the 16 most important families in 15th century Florence, Italy [61]. A typical question to ask in such a network is who are the most *central* nodes in the network, and dozens of measures have been introduced to emphasize different aspects of a central position in a network [47], as we will discuss below.

In contrast to this, networks in physics were used as model for different interaction topologies of indistinguishable units, e.g., atoms in magnetic material. Here, the spin of an atom is influenced by the orientation of neighboring atomic spins as well as by thermal fluctuations. A typical question to ask is under which conditions the whole piece of material shows a certain level of magnetization. The two most commonly used placements of the atoms and thus the definition of their neighborhood are (a) multidimensional grids and (b) the random graph, in which every atom is connected to a randomly selected subset of other atoms [65]. In these networks, the very nature of the material is not taken into account: basically all magnetic materials behave in the same way, i.e., the model is *universal*. Furthermore, the energy state of individual atoms does not influence the global behavior; rather, the model is based on a certain *stochastic* distribution of energy states that fully determines the behavior of the system. Thus, physics concentrated on systems in which *local, individual behavior* determined by the *local environment* result in *global behavior* which can moreover be transferred to other systems and thus can be considered to be *universal behavior*. In 1998, Watts and Strogatz were the first that tried to explain a commonly observed global phenomenon, which is called the *small-world effect*, in three different networks by a new kind of network model [68]. The small-world effect describes the phenomenon that in social networks humans are on the one hand members of tightly knit social groups but that on the other hand everyone is connected to everyone else by only a short chain of acquaintances. They showed that this can be easily explained by a model that scales between the very ordered multi-dimensional grid and a random graph. In 1999, Barabási and Albert came up with a model that explains the emergence of so-called *scale-free networks*, i.e., those in which the probability $P(k)$ to draw a node with k edges is proportional to $k^{-\gamma}$, where γ is a constant. In both of these works, the perspective was to give a *stochastic model* of how individuals behave locally that explains a globally emerging behavior. Moreover, both of the papers showed that the observed network structures emerge in very different kinds of networks, namely social, biological,

and technical networks[2]. With this they showed the *universality* of their finding.

3.1. Complex networks as proxies of complex systems

On the example of protein-protein interaction networks we want to show how a complex network is a proxy of a complex system: Proteins are small molecules in a cell, which are active at different times and places within a cell (see Fig. 1). Some of them come into close contact, thereby exchanging information in a structured way and inducing or prohibiting actions of each other. Biological functions such as cell division or DNA duplication need these concerted actions of groups of interacting proteins to get under way. To understand the biological role of a single protein is a very tedious and costly task; thus any kind of prediction on the biological function of a new protein based on less expensive data is very much appreciated. Fig. 1(I) displays a schematic way of how the real interactions of proteins could look like in a cell: dotted lines represent interaction within the same biological function and the line with the vertical end represents an inhibition, which is not likely to connect two proteins with the same biological function. While the full complex system contains even more information like the dynamic and local pattern of a protein in the cell (Fig. 1(I)), this complex information is not always available for all of the proteins. Simple information, which **is** readily available, is whether two given proteins can come close enough into contact to possibly interact with each other. This information can be gathered in high-throughput assays and displayed by a complex network between the according proteins (Fig. 1(II)). Of course, by mapping the complex system of a cell onto the complex protein-protein interaction network, some information is lost and erroneous information is added:

1. Both edges denoted by a in Fig. 1(II) are missing because the yellow and the pink protein only interact with the red one if they interact with each other. In the high-throughput assay only pairs of proteins are tested, and thus only the interaction between the yellow and the pink protein is detected.

2. Edge b is misleading since it does not denote an interaction between two proteins with the same function.

3. Edge c is misleading since although the yellow and the orange protein do come into close physical contact in the assay, they are normally in two different compartments of the cell and would not interact with each other.

4. Edge d in Fig. 1(II) might be missing out of methodological problems, e.g., because the signal is too weak to be detected.

Missing edges that should be there are called *false-negative* edges and additional edges are called *false-positives*. Edge c is an edge that reports a true interaction but it does not connect proteins with the same biological function but rather the opposite, i.e., it carries distorted information. Thus, the reduction to a complex network contains distorted information, and it introduced false information and neglected true information. There is a more fundamental question to this reduction: physical contact is not necessary for two proteins to share the same biological function and physical contact occurs between proteins with different biological functions. Can this network really reveal information about groups of proteins with the same biological function? If we had the information of which proteins interact with each other because they share the same biological function, we could expect that those with the same function would build tightly knit groups of densely interconnected proteins. The most common approach is thus to find such *clusters* of proteins that are densely interconnected in the protein-protein interaction network as shown in Fig. 1(III) (see subsection 5.3 for an overview on clustering methods). The question is then whether the clusters in the protein-protein interaction network resemble the wanted clusters of proteins with the same biological function.
In general, network analytic methods seem to be quite robust against added random edges and random deletion of edges and one can often see that those proteins that end up in the same clusters share biological function to various extents [62]. It is thus assumed that those proteins with an unknown function might have the same biological function as those who are in the same cluster. Still, the question is whether the resulting clusters are really the optimal ones that predict the biological function correctly for most proteins. This has, up to now, not been evaluated quantitatively, neither

[2] *It is often claimed that the degree distribution of real-world networks is scale-free, i.e., $P(k) \simeq k^{-\gamma}$ [3, 6]. This is however, difficult to show empiricially due to the finiteness of the data sets [21]. It is thus safer to speak of networks with a highly skewed degree distribution.*

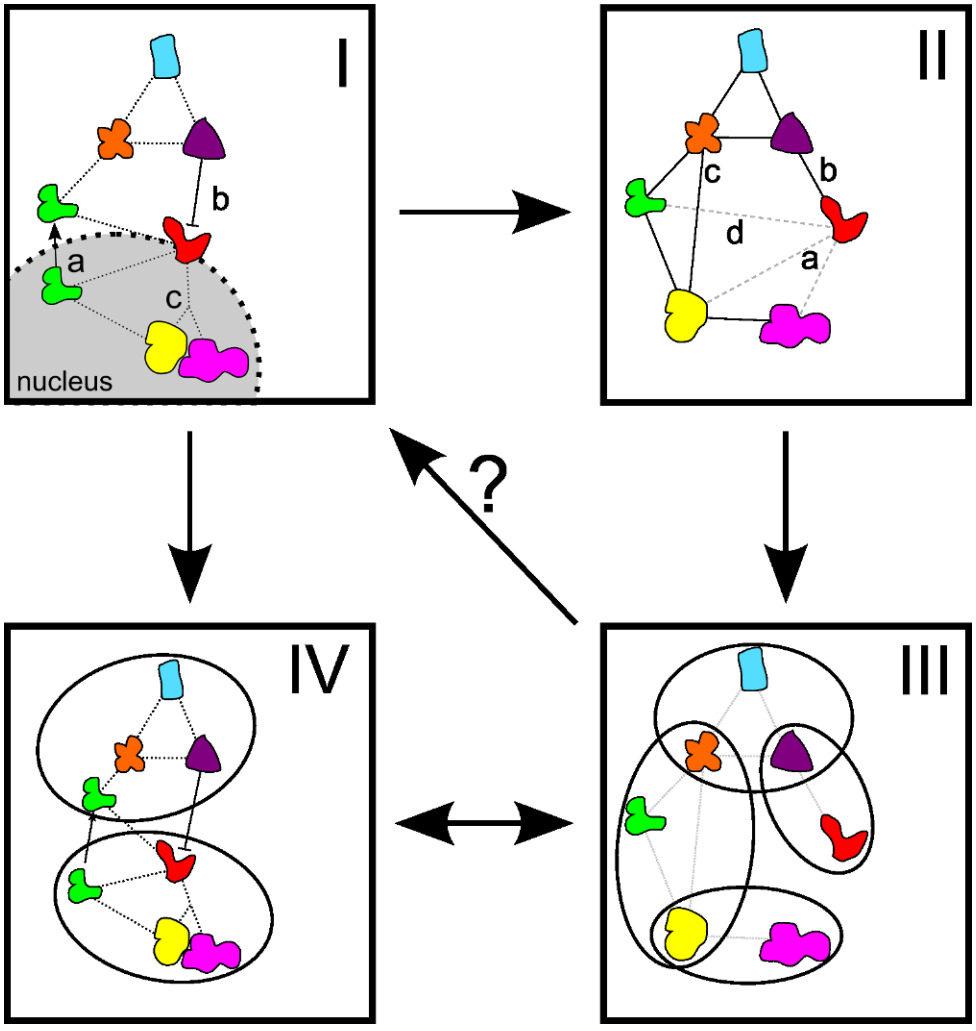

Figure 1. I) A schematic view on proteins in a cell. Some are located in the nucleus, some are found in both, nucleaus and cytoplasm of the cell; edge a denotes that the green protein can leave the nucleus. Proteins can interact with each other in various way, e.g., inhibit each other (edge b). General interactions are represented by dotted lines. Edge c shows a protein complex (yellow and pink protein) that interacts with the red protein once it has formed. II) A complex network that represents part of the information in I. Some edges are added, others are missing. III) The result of a clustering method on II. Does it really represent groups of proteins with the same biological function, i.e., what is the correlation between III and I? IV) Only the comparison of the clusters in III with a known or desired clustering, derived from the compex system itself, as shown in IV can evalute this.

for the afforementioned clustering algorithm by Palla et al. [62], nor for clustering algorithms in general. Moreover, most measures from the major classes of network analytic methods for static networks follow the same path: the complex system at hand is too complex to process. Thus, it is reduced onto a complex network. This process might lead to distortions, addition, and deletion of edges, and in general, the observed relationship might not be totally equivalent to the relationship that is of interest. Nonetheless, the method is then applied to the complex networks with the assumption that significant patterns in it correlate with significant patterns in the corresponding complex system. In Fig. 1 this corresponds to the question whether the information in (*III*) is of any use to understand (*I*).

There are basically two ways to evaluate whether the results reveal new information on the complex system of interest: The first is to analyze both steps, from *I* to *II* and to *II* to *III*, namely how closely the complex network resembles the wanted complex network and how well the method deals with distorted, added, and deleted information. In the case of the protein-protein interaction network this step already fails enormously: it is reported that in high-throughput assays up to 50% of all edges are false-positive [23]. No method we know has shown that it can deal with such a high

percentage of unreliable data. However, even if the complex network was almost error free, the method itself could still be less than optimal in identifying the clusters or other structural patterns of interest. Thus, the connection between the results of the method and their correlate in the complex system of interest would still be unclear. In this mini-review we propose that the only way to evaluate the correlation between a network analytic result and its correlate in the complex system is to test the method against so-called ground truth data or a gold standard solution, as described in the following section.

Thesis 1

Complex networks are proxies of a complex system of interest that can be analyzed by various network analytic methods. The intuition is that statistically significant relational patterns in the complex network point to correlating structures in the complex system it is derived from. The mapping of the complex system onto a complex network often results in distorted, added, and missing information. Network analytic results might thus not always point to the intended straightforward correlates in the complex system of interest. It is thus necessary to evaluate the quality of any network analytic method to reveal interesting patterns in the complex system it is derived from.

4. Ground truth and gold standard solutions in network analysis

Describing the universal behavior of networks has been a very fruitful aspect, but so far, the concrete insights into biological networks or the working of neural networks have been almost inconsequential [42]. We are however convinced that network analysis shows a very promising approach to understand *specific* structural elements in a given network if benchmark sets would be provided and quality evaluating methods from machine learning and data mining would be applied rigorously [42, Ch. 7]. In the following, we will use the term *ground truth* to determine any kind of knowledge about the entities or their relationships in a complex network that can be used to evaluate the performance of an algorithm that was developed to discover significant relational patterns in a network. We will use the term *gold standard solution* if ground truth is not available but humans can agree on a substantial part of the wanted result or on the exact result on a part of the data. Again, the gold standard can then be used to assess the quality of a algorithm. We want to illustrate both approaches on two exemplary algorithms, one for the prediction of future links in co-authorship networks and one for computing similarity values for vertices in a bipartite graph.

4.1. Ground truth for link prediction

In their paper from 2007, Liben-Nowell and Kleinberg examine so-called *co-authorship networks* which display an edge between two scientists if they have co-authored at least one paper [49]. These networks evolve in time as new cooperations emerge. The authors examined whether the network structure in which two non-cooperating authors at time t are positioned can predict whether they will start to cooperate until time $t' > t$. To evaluate this, Liben-Nowell and Kleinberg created co-authorship networks established by publications from 1994-1997 in five different communities, and artificially split each of the data sets into two time intervals, from 1994 – 1996 and 1997 – 1999. The network resulting from the first time interval was then used to measure various structural relationships between any two scientists that did not yet cooperate. For each measure, all pairs of scientists were then ranked by the according value and the top ranked pairs were predicted to have published a common article in the second time interval. Since the data was artificially reduced to the first time interval but was fully known, the prediction could be evaluated given the *ground truth* of those authors who truly wrote their first common paper in the second time interval. With this information, it could be quantitatively evaluated which structural feature yields the best prediction for which community. Interestingly, it turned out that despite its simplicity, the number of common co-authors was a very good predictor for all of the communities. With this idea of testing a network analytic measure against a part of the network data that was hidden from the algorithm, Liben-Nowell and Kleinberg allowed for a benchmark against which new algorithms can and must be compared. The state of the art for co-authorship networks is now a supervised learning approach that combines a number of the structural measures to improve the link prediction quality [50]. For networks from other domains like biological or technical networks similar benchmark data sets are still missing and thus it is not clear whether the resulting prediction algorithm is universal or not.

In our second example, ground truth itself is not available, but a gold standard solution on a subset of vertices was introduced to measure the quality of similarity indices for products in market basket analysis.

4.2. Golden standard solution for similarity indices in user-film-ratings

In market basket analysis, a set of customers and the products they buy is used to determine subsets of products that are more often bought together than expected [35, Ch. 5]. The intuition is that these products are also inherently similar to each other and thus this information can be used to recommend the products to customers that bought other products of the same set. This idea of finding similar entities can be generalized to any kind of bipartite data, e.g., bipartite relations between different types of biological entities, board members and companys, or even citizens and their answers to questions in a census. So far, the analysis of the resulting subsets has been rather anecdotal, e.g., when examining U.S. census data on subsets of commonly co-occurring answers the best rated combinations were: "five year olds don't work, unemployed residents don't earn income from work, [and] men don't give birth" [15]. In the simpler case of finding two products that are bought together more often than expected, a similarity index f is needed that assigns a real-value to each pair of products p_i, p_j. For each product p_i all other products p_j can then be ranked by their value $f(p_i, p_j)$, and, e.g., the top five ranked products can be used as recommendations for those customers that like product p_i. Since market-basket data can be displayed as a bipartite graph between products and customers, it is obvious that a similarity index can be based on network analytic structural measures. Without a reference set that defines a gold standard solution on at least a subset of the products it is, however, difficult to assess whether network analytic similarity indices result in better recommendations or not.

With this respect, we have compared one of the classic similarity indices used for market basket analysis, called $leverage_{\mathrm{SIM}}$, with a new, network analytic index called $leverage_{\mathrm{FDSM}}$ on a dataset comprised of $17,770$ films and $20,000$ users [73]. The problem is that for a given film p_i many subsets of 10 other films *a posteriori* might sound plausible. Table 1 shows the top five recommendations for two films, a romantic sitcom called "Friends" (Season 1) and a Sci-Fi/Action series called "Star Gate SG-1" (Season 1), ranked by both similarity indices. Looking at the ranks of the classic index $leverage_{\mathrm{SIM}}$, most humans would certainly assess that the recommendations are reasonable: for the emotional sitcom the recommendations contain blockbusters that are in general liked by women ("Miss Congeniality", "Pretty Woman"), while for the Sci-Fi series action and sci-fi blockbusters are recommended ("Armaggedon", "Independence Day").

Table 1. For the two films "Stargate SG-1, Season 1" and "Friends, Season 1" we show the top five ranked other films according to two similarity indices (top two rows: classic similarity index $leverage_{\mathrm{SIM}}$, lower: new, network based similarity index $leverage_{\mathrm{FDSM}}$; as described by Zweig [73]).

film	1st	2nd	3rd	4th	5th
Friends, Season 1	Miss Congeniality	Forrest Gump	Pretty Woman	Pirates of the Caribbean: The Curse of the Black Pearl	Friends: Season 3
Stargate SG-1, Season 1	Independence Day	Armageddon	Pirates of the Caribbean: The Curse of the Black Pearl	Men in Black II	Lord of the Rings: The Fellowship of the Ring
Friends, Season 1	Friends: Season 4	Friends: Season 3	The Best of Friends: Season 2	The Best of Friends: Season 1	Friends: Season 5
Stargate SG-1: Season 1	Stargate SG-1: Season 2	Stargate SG-1: Season 3	Stargate SG-1: Season 4	Stargate SG-1: Season 6	Stargate SG-1: Season 5

A posteriori, the recommendations seem to be reasonable and intuitive. On the other hand, one might argue that the most similar film to a part of a series is another part of the same series and define this as gold standard solution for the subset of films that are part of a series. It can be seen that the ranking based on the new index is much better

at filtering these other parts of the series out of the set of 17,770 films[3]. Note that the rankings can also be used to generate graphs, namely so-called *one-mode projections* of the underlying bipartite graph [72]: Fig. 2 shows the local network around the first season of "Star Trek: Deep Space Nine", where there is a connection between it and another film X if either X is among the ten highest ranks of it or if it is among the ten highest ranked films of X. It can be seen that the old method (Fig. 2(a)) produces a less dense local network containing also some blockbusters, while the new method produces a dense network been clearly related films from the Star Trek series.

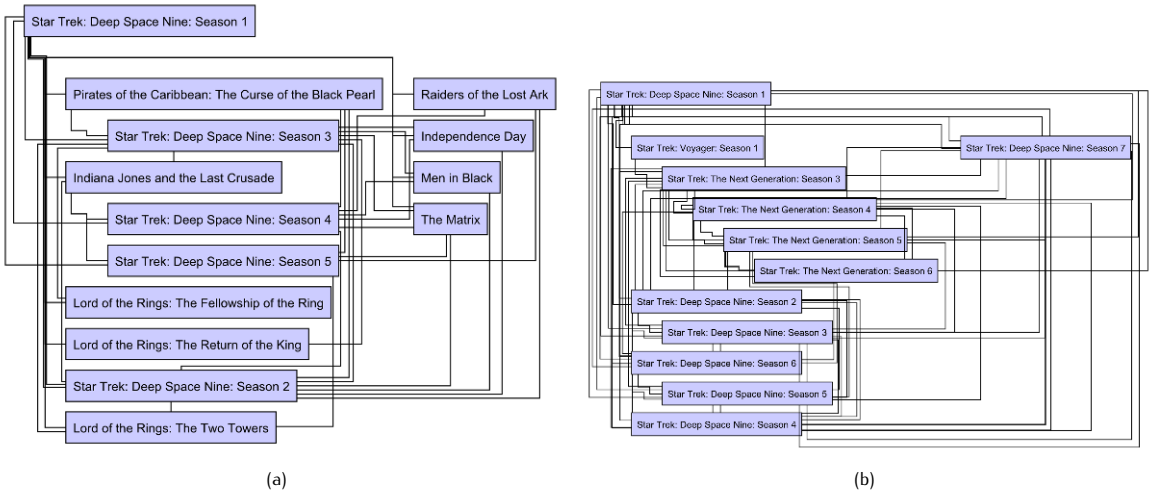

Figure 2. Local networks between films created from two similarity indices, (a) $leverage_{\text{sim}}$ and (b) $leverage_{\text{FDSM}}$ [72]. The first season of "Star Trek: Deep Space Nine" is connected to its ten highest ranks and all series are connected with it for which this film is among the ten highest ranks.

It might be arguable whether a good recommendation method should really rank other parts of the same series highest; the main point here is that assessing the quality of a ranking a posteriori and in a large space of possible rankings is not good enough, since the human mind is very good at spotting patterns in many of the possible subsets of products. Rather, it is necessary to define a gold standard solution on a subset of the graph and to prove that the algorithm is able to classify the vertices accordingly. This allows the field to both discuss the chosen gold standard solution and to improve algorithms with respect to them.

Thesis 2

Link prediction and the creation of one-mode projection of bipartite graphs based on similarity indices can be evaluated by ground truth and gold standard solutions. The performance of a link prediction algorithm is evaluated by testing it on a fully known data set. The idea is that if the prediction is good on this fully known network, it will also be good on other, similar networks. The similarity index was tested on a gold standard solution defined on a part of the network. It is assumed that the similarity index works equally well on other parts of the network. These examples can be used as a paradigm for the evaluation of other network analytic measures for the analysis of static networks.

We have shown on two examples that the definition of ground truth or gold standard solutions in a network analytic setting can help to design better algorithms. We will now summarize the state of the art for three of the most important groups of network analytic measures and also report on how they have been evaluated so far.

[3] *For a full set of results on all series in the data set, see www.ninasnet.de/Projects/OMP_Recommendations/10BestRecommendations.html*

5. Network analytical measures and their evaluation

Several network analytic methods have been proposed over the last 50 years to quantify certain intuitions humans have about the structural composition of networks. Recent textbooks list many of them [25, 37, 45, 55], and the most important belong to the following three categories of methods:

1. Centrality measures [46, 47];
2. Network motifs [2];
3. Clustering methods [27] and block models [48, 60, 64, 67].

The idea behind these measures is that they identify structures in the network that correlate with structures in the complex system the network is derived from. This is especially interesting if the correlating, external structure would otherwise be difficult to observe or even be unattainable: centrality measures try to identify those vertices that are most important for the work of the complex system under observation; methods for the computation of network motifs try to identify those subgraphs of a network that are most likely correlated with a special process on the complex system under observation; and methods that compute blocks or clusters in a network finally try to identify groups of vertices and a pattern of relationships within and between these groups such that most edges can be explained by this pattern. If this can be achieved, it is assumed that the identified groups also share a role or function in the complex system under observation. In the following, we will briefly sketch some instantiations of the three categories of network analytic measures, show with what structure they are believed to correlate, and how this correlation has been evaluated so far.

5.1. Centrality Measures

Among the first network analytic measures were so-called *centrality indices*, and by now dozens of them have been proposed [46]. Due to their generality, there is not yet a tight definition of what constitutes a centrality index. The minimum requirements are:

1. A vertex' centrality value only depends on the graph's structure and not on external properties of the vertex;
2. Its value increases with the perceived importance of the vertex [46].

The intuition behind the idea of quantifying centrality is that a well-designed centrality measure captures the importance of a vertex in the network's structure and that this correlates with the perceived importance of the corresponding entity in the complex system under observation. The most intuitive proposal for a centrality index is to use the degree or the weight of the sum of incident edges as a measure for the centrality of a vertex (*degree centrality* [46]). Although this most basic centrality index is often astonishingly good in identifying important vertices [38], it is clear that it does not suit all complex systems equally well: a boss in a company does not need to know absolutely the most of her employees, but she is still the most central person regarding decision making or information dissemination in the company. For these situations, other centrality measures were developed, like *Katz' influence measure* [41], which measures the number of paths from a given vertex, weighted by an attenuation factor β that quickly decreases with the length of the paths. In yet another respect, the secretary of the boss might be perceived as most important because she provides—at least for most employees—the only access to the boss; this perspective is best captured by a different centrality measure, the *betweenness centrality* [29]. Next to these most widely known centrality indices a large range of other indices has been proposed, each with its own motivating example [46]. Although the set of centrality indices so far resisted to be categorized in any comprehensive way, Borgatti has introduced a very important view on when to use which centrality index [11]: he proposes that centrality indices can only be chosen with respect to a process of interest that uses the network's structure. Examples for these processes are gossiping, flow of goods, or flow of information. He then proposes that there are mainly two dimensions that determine a suitable centrality index, namely the question of whether the type of object flowing over the network can be split (like money) or reproduced (like information) and what kind of paths are used by the flow, e.g., shortest paths or random walks.
Although there are many different centrality measures and there is also a theory on when to use which, there is overall a lack of proof that the centrality indices proposed so far are of substantial value in revealing new information. Most published analyses use a set of different centralities on the same network and the subsequent analysis of the results is

rather anecdotal: the vertices are sorted by the chosen centrality index or indices, and the ranking is then explained in parts by external knowledge or by comparing the rankings of different centrality indices [26]. A good example for this is the application of various centrality indices to the marital network between the most important families in medieval Florence mentioned above [61, 67]. With almost all centrality indices, the Medici family is assigned the highest centrality which coincides with historical knowledge about the family's importance. The problem is that the other ranks are only briefly discussed and not evaluated with respect to some ground truth ranking.
Social networks might prove to be essentially more difficult in establishing even gold standard solutions, but there are certainly networks in which an importance ranking is known beforehand or in which at least a binary classification into important and less important vertices can be achieved. This ground truth or gold standard solution can then be used to evaluate the ranking imposed by a given centrality index. One network in which this was possible and done, is the protein-protein interaction network of yeast. For some proteins of this network it is known whether their removal from the genome is lethal for the resulting organism. In their paper "Lethality and centrality in protein networks" Jeong et al. show that there is a correlation between the degree of a protein and its lethality [38]: From those 93% of proteins with a degree of at most 5, only 21% are lethal, but from the 0.7% of proteins with a degree of > 15, around 63% are lethal. This is one of the few examples where an evaluation of a centrality index and its correlation with known facts was actually tested, but it was not used to establish the best centrality index to identify lethal proteins. Correspondingly, there is also not much work analyzing when a centrality measure does **not** correlate with its long-believed external correlate. One of the few examples analyzes whether the betweenness centrality is a good measure for the load of routers in the Internet [36]. According to the theory by Borgatti, the betweenness centrality could be suitable because IP packets are not normally reproduced and are mainly sent on shortest paths. But Holme can show that under various models of packet flow in the internet there is no high correlation between the betweenness and the load of a router.

Observation 1

In summary, many centrality indices have been proposed and there is a good theory on when which centrality index should be used, but their correlation with an externally perceived importance has not yet been thoroughly evaluated in real-world settings.

5.2. Network motifs

Network motifs are statistically significant subgraphs that characterize subsets of vertices that are connected in the same way. It is assumed that these motifs bear function since they occur more often than expected in a random graph model. The precise definition is as follows: let H be any graph and let G denote the graph of interest. H is called a *network motif* of graph G if it occurs in G more often than expected in comparison to a *null-model*. A *null-model* in this respect is any kind of randomized network that maintains certain structural aspects of a graph. The most simple one is the classical *Gilbert random graph*[4] in which every edge has the same probability $0 \leqslant p \leqslant 1$ to be established. Given a graph $G = (V, E)$, the *according* random graph is built on the same number of nodes $n = |V|$ and p is defined such that the *expected number of edges* equals $m := |E|$:

$$p := \frac{2m}{n(n-1)} \tag{1}$$

for the undirected case and

$$p := \frac{m}{n(n-1)} \tag{2}$$

for the directed case. Thus, this model maintains the number of nodes and expectedly maintains the number of edges, while all other structural features are perturbed. It can thus be seen as the most basic null-model for any given graph G. For this random graph model, the expected number of certain subgraphs, e.g., cliques, in dependence on the parameter p is known explicitly [10].

[4] *Actually, this random graph model is often called the Erdős-Rényi random graph model, but it was first published by Gilbert [9, 31].*

Other models take more structural features into account: One main feature that distinguishes almost all real-world networks from this random graph model is their *degree distribution*, i.e., the probability $P(k)$ that a node chosen uniformly at random from V has degree k. The degree distribution is Poisson distributed in the classic random graph model but it is highly skewed for most others, i.e., there are many nodes with a small degree but there are some with a much higher degree than expected in a Poisson distribution [24]. The most common null-model is thus a random graph that maintains not only the number of nodes and edges but also results in exactly or expectedly the same degree sequence [17, 56]. That is, for each vertex v with degree $deg(v)$ in G, vertex v will have the same or expectedly the same degree in any instance $G' = (V, E')$ of the random graph model. An instance of the first type can be most easily produced by the *stub method* [58]: every vertex is assigned $deg(v)$ many 'half-edges' or *stubs* and in each steps two of the remaining stubs are chosen uniformly at random, connected to an edge, and removed from the pool of stubs until all stubs are connected. An instance of the latter type can be produced by connecting each two vertices x, y with probability $deg(x)deg(y)/(2m)$. It is easy to see that the expected degree of v is then $deg(v)$ [17]. For these types of graph models, much less is known about their typical properties. In their article, Chung and Lu discuss the average distance of vertices in the random graph model with a given expected degree sequence. If other properties are of interest, it is possible to create a large set of samples, i.e., of instances from the random graph model, and to measure the property of interest in them. By the law of large numbers, the observed average value of the property can be used as an approximation of the expected value of the property in the whole ensemble of possible instances.

If the underlying graph is bipartite, and the random graph model is needed to maintain this bipartiteness together with the degrees and the number of nodes and edges, there is even less known about the properties. Moreover, there is no easy way to determine a probability for an edge between two nodes x and y and thus constructing an instance from this model uniformly at random is more complicated. It can either be implemented as a Markov chain sampling [16, 32] or a so-called *importance sampling* [1].

There are even more specialized random graph models, e.g., ones in which the number of certain subgraphs is maintained [51, 52], those in which the evolution of a network is modeled [6, 68], or those that are determined by a spatial embedding of the vertices [44, 69]. In general, the more structural features they take into account, the less is known about their expected structural properties and very often, an efficient algorithm to construct instances uniformly at random is often missing. For many random graph models $G(\vec{p})$, determined by a vector $\vec{p}$ of parameters, and a network property $\mathcal{T}$, there is no closed formula for the expected value of $\mathcal{T}(G)$ for a given $\vec{p}$. An example is the *expected number of common neighbors* (co-occurrence) of two nodes x, y in the bipartite fixed degree sequence model sketched above [73] or the *expected number of occurrences of a given subgraph*. Thus, in this field there are many open questions that need to be analyzed in the following years.

Once a suitable random graph model is chosen as a null-model, the number of occurrences of all subgraphs of interest is computed for the original graph G. This is of course a costly task [39] and still a matter of active research. Other approaches try to estimate the number of subgraphs in the original graph by different sampling methods [34, 40, 63]. When the number is known, it is compared to the *expected number* of the occurrence of the same motif(s) in the null-model. If this is not known by a closed formula, it is approximated by the average of this number in a large sample of instances from the null-model, and the difference is statistically evaluated, e.g., by the *z-score* as proposed by Milo et al. [53]. Note that this can of course only be computed if the occurrence can be assumed to be normally distributed. If the occurrence of any subgraph G' is found to be significantly larger in G than in the according null-model, it is then said to be a *network motif*. The choice of Milo et al. in their first paper on network motifs was to use a null-model in which the degree sequences were maintained while the edges were completely perturbed [53]. An interesting comment on their work by Artzy-Randrup et al. showed that some of the network motifs that were significant with respect to that null-model were absolutely insignificant with another, geometrical null-model [4]. The main implication of their article is not that the geometrical null-model was in any way better suited as a null-model but just that the question of when which random graph model can be used as a null-model is an open question that has not been resolved until today.

Wasserman and Faust already required that ranking vertices by their centrality indices alone is not good enough if the index's value cannot be assessed statistically [67, p. 728]. In this sense, network motif methods are already much better since they compare the number of occurrences of a motif with that of a perturbed data set. The article by Artzy-Randrup et al. however shows that this internal quality measure might not be good enough since the choice of the null-model is so crucial. Thus, it would also be beneficial for the development of network motif computing methods to establish a gold standard solution describing which kind of subgraphs need to be contained in the computed set to accept the algorithm.

Observation 2

Network motif detecting algorithms are even harder to evaluate in their quality since they rely on an additional step, the choice of the appropriate random model. It is very difficult to argue why which random graph is best suitable as shown above. Thus, these methods need evaluation by ground truth or gold standard solutions even more than others.

5.3. Clustering methods and block models

Another common set of methods in network analysis focuses on finding so-called *clusters* [27] or to assign vertices to so-called *blocks* [60]. Again, the notion of *clusters* is based on an intuitive assumption: it has been hypothesized that in social systems there are groups of people that have a higher chance to be connected to each other than to people of another group; this assumption is called the *homophily* assumption. Clustering methods try in general to find subsets of vertices in which the probability that any two of them are connected is high. *Block models* try to find subsets of vertices that show a similar relational pattern with vertices assigned to the same or a different block; especially, vertices in the same block do not need to be connected at all. The block model is thus the more general concept to explain the observed relational pattern in a given network. The block model can be abstracted from a given network, but more often it is derived from external knowledge about the vertices. This is especially common in social networks, where certain attributes of vertices might be known as in the following example: In a data set concerning 48 political campaigns and their donors it was known who donated to which campaign and who was affiliated with the workers/democratic party [5]. It can then be tested whether the probability that two 'democrat'-donors donate to a common campaign is higher than that a democrat-donor and a non-democrat donor choose the same campaign. In this setting, the roles are assigned by external attributes of the vertices, and the question is whether the network's structure can at least partly be explained by a homophily assumption within the groups of vertices with the same attributes.

In contrast, clustering algorithms for networks are oblivious of external attributes and thus try to identify groups such that the density of edges within groups is much higher than in the whole graph. The idea is that if there are groups within which there are many edges, there must be an external correlate, e.g., a common attribute, that is responsible for the high density of edges within the group. In this sense, block models and clustering algorithms can be seen as complementary: where the first (often) starts from external knowledge, the latter first defines groups with a relational pattern that are then correlated with external attributes. An example for the latter is given by Palla et al. in a paper in which they introduce a new clustering algorithm and apply it to a protein-protein interaction network. Their algorithm takes only the adjacency matrix of the network into consideration and is oblivious of any external attributes of the proteins. The authors find that within one of the identified clusters most proteins are associated with a certain biological function known as "ribosome biogenesis/assembly" [62]. Another protein with the name Ycr072c was also found in this cluster but its biological function was not known at the time of publication; Palla et al. suggested that their algorithm might predict that this protein is also important in the ribosome's biogenesis and assembly. This would indeed be an enormous information gain: the information used in a protein-protein interaction network is relatively cheap to obtain, compared with the time-consuming and expensive proces of excluding possible biological functions of a given protein until the one is found in which it is engaged.

The promise of clustering methods and, more general, algorithms that identify blocks of vertices with a similar relation pattern is strong, and thus it needs to be evaluated whether the simple approach is actually able to discover groups of entities with a simple relational pattern. There are principally four different methods to evaluate whether these models achieve what they are assumed to do:

1. **Bootstrapping or permutation methods:** In cases where there is no ground truth, the data itself can be used to generate subsamples or it can be perturbed as described above for network motifs [6]. The resulting samples can then be evaluated with respect to their edge density within or between the clusters or blocks.

[5] *Hanneman R., Riddle M., Introduction to social network methods, online publication of the University of California, Riverside; http://faculty.ucr.edu/~hanneman/nettext/, 2005*

[6] *Hanneman R., Riddle M., Introduction to social network methods, online publication of the University of California, Riverside; http://faculty.ucr.edu/~hanneman/nettext/, 2005, Ch. 18*

2. **Clustering and block model quality measures:** Block models can be evaluated by a set of different quality measures, e.g., by comparing the observed edge density with the expected edge density in some null-model [67, Ch. 16]. Similarly, a clustering's quality can be quantified by so-called *clustering quality measures*, e.g., the *coverage* which measures the percentage of the total weight of edges within clusters with regard to the sum of all edge weights [30, Section 8.1]. These measures can also be used to compute a clustering with optimal quality, and in this case they also uniquely define what a clustering is. A special case which combines method 1 and 2 is a cluster quality measure called *modularity* [57]: it computes the coverage but reduces it by the *expected* edge density within the computed clusters. It can be used with any non-overlapping clustering computed by an algorithm or it can be tried to compute a clustering with optimal modularity. This is done with various heuristic methods [33, 57] since optimizing the modularity is NP-hard [12].

3. **Artificial networks:** Another accepted method is to generate artificial networks with a predefined clustering or block model structure as ground truth and to partly perturb these networks or to add random edges to it [13, 14, 59]. The clustering (or block model) identified by the proposed algorithm can then be compared with the original clustering C_O hidden in the artificial network. Especially, the known clustering implies the characterization of all pairs of vertices x, y into two classes: those that are in the same cluster (block) and those that are not. Given the algorithmically computed clustering C, each pair of vertices is now either *true positive*, i.e., it is in the same cluster (block) both in C_O and C, *true negative*, i.e., not in the same cluster (block) in both cases, or *false positive* and *false negative*. With this notion, classic prediction quality measures like *specificity* (percentage of true negative cases of all negative cases) and *sensitivitity* (percentage of true positive cases of all positive cases) can be computed.

4. **External knowledge (applicable to clustering algorithms):** Clustering algorithms should be tested on real-world networks with a known clustering as a gold standard solution. There are only very few examples of this in the literature and most of them are very small: the first one is Zachary's karate club [59, 70] whose members split into two independent group after some time. Another well-known data set is the network of political books as defined by the links "customers who bought this book also bought" provided by Amazon and manually labeled by Valdis Krebs (http://www.orgnet.com/divided.html). It is, however, so simple that almost any clustering algorithm will be able to correctly divide it into clusters that consist of only one political color. Another small data set with known clustering was proposed by Zahoránszky et al., which describes a set of 128 molecules, manually labeled into six distinct clusters of similar molecules based on their scaffold ring system [71]. As with the artificial networks, such a gold standard solution can then easily be used to test the outcome of a newly proposed clustering algorithm.

Observation 3

In general, many network analytic measures have been proposed for the tasks of identifying central vertices and for groups of vertices with the same or similar relational patterns. Many applications like recommendation systems or biology would profit from optimal clustering algorithms designed for their specific needs. So far, a posteriori analysis of the resulting clusters shows that they have a reasonable structure but it is difficult to assess whether they are optimal for the task at hand. We conclude that for most of the methods, standard benchmark sets and ground truth solutions are missing although their benefit could be tremendous.

In the following section we will discuss how network analysis can evolve from a rather descriptive tool box to a predictive science with great benefits for the analysis of complex systems.

6. Discussion: From description to prediction

As we have sketched above, network analysis has evolved from a descriptive tool box on single small networks from sociology to a descriptive analysis method for universal network structures in single large networks from very different fields. But network analysis promises to be more than just a descriptive tool box: anecdotal evidence has shown that centrality indices sometimes really coincide with the perceived importance of entities in the complex system they are contained in [38], that automatically computed clusterings contain clusters of remarkably similar objects [20], and that network motifs found in different networks designed for the same process are remarkably similar [54]. The next step will

be to evolve the field into a truly predictive one by creating and openly sharing benchmark sets with known ground truth or wanted gold standard solutions. Other fields like the development of efficient satisfiability algorithms with their SAT competitions and the resulting benchmark sets[7], the machine learning community with various acknowledged benchmark sets[8], or the community concerned with pattern recognition in images[9] have all greatly profited from the discussion on what is considered as a benchmark, and subsequent improvement of algorithms.
The examples from section 4 have shown that a quantitative evaluation of network analytic results and their correlation to significant groups or patterns in the complex system of interest is possible. Link prediction algorithms as discussed in section 4.1 basically result in a ranking which is then compared to a ground truth ranking in some data set where this information is available. It is assumed that, if the quality is high in the known data set, the algorithm will also perform well on similar data sets where this information is not available. This general approach can be applied to all network analytic measures that result in a ranking of nodes or subgroups, such as centrality measures or network motif identifiers. A similarity index between films as described in section 4.2 results in rankings for each node that can then be used to derive local networks (or clusters) around each film. Although it is in general difficult to build ground truth clusterings for something as complex as films, there is a subset of films for which a gold standard solution could be devised: the series. By comparing the clusters around films that are part of a series with the gold standard, algorithms can be evaluated in their quality. It is then assumed that if they perform well on this subset of data, they will also perform well on other parts of the data. Of course, a ranking can also be compared with a partial ranking on a subset of data instead of with a full ranking in a similar data set and the performance of a clustering algorithm can be tested on a different network with known ground-truth instead of a subset with known ground truth.

Thesis 3

Network analytic methods such as centrality indices, clustering methods, and network motif identifiers need to be thoroughly evaluated with respect to their ability to identify statistically significant patterns in a complex network and to quantify whether these correlates with relevant patterns in the complex system from which the network was derived. Since network analytic methods basically result in rankings of single nodes or groups of nodes, or they identify blocks and clusterings of nodes, they can quantitatively compared with ground truth or gold standard rankings and clusterings by standard machine learning quality measures [35]. As shown in the examples, ground truth or gold standard solutions can be derived from either a) a part of the complex system like the series in the film set or b) a very similar complex system and its corresponding network like the co-authorship networks of different communities. In order to prove the usability of network analytic methods in biology, economy, sociology, and various other fields of application, it is thus necessary and possible to systematically build benchmark data sets and to develop a systematical framework of quality evaluation in the design of network analytic methods.

In network analysis, it is an open question of how similar two networks need to be in order to approximate an algorithm's performance on one with that on the other. It is also clear that not all possible partial knowledge about a complex system is equally well suited to assess an algorithm's performance on the rest of the data. Still, without starting to evaluate network analytic methods we will not be able to address these important open questions in the future.
In summary, network analysis has measured and described many different networks so far with great success, but to make the next step we need to quantify the correlation between the significant structures found in a complex network and the intended correlates in the complex system they are derived from. To quote Lord Kelvin: "In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be."[43, p.73]

[7] *http://www.satcompetition.org/*
[8] *http://archive.ics.uci.edu/ml/*
[9] *http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/*

References

[1] Admiraal R., Handcock M.S., networksis: A package to simulate bipartite graphs with fixed marginals through sequential importance sampling, J STAT SOFTW, 2008, 24, 1-21

[2] Alon U., An Introduction to Systems Biology - Design Principles of Biological Circuits, Chapman & Hall/CRC/Taylor & Francis, Boca Raton, FL, 2007

[3] Nunes Amaral L.A., Scala A., Barthélémy M., Stanley H.E., Classes of small-world networks, P NATL ACAD SCI USA, 2000, 97, 11149-11152

[4] Artzy-Randrup Y., Fleishman S.J., Ben-Tal N., Stone L., Comment on "network motifs: Simple building blocks of complex networks" and "superfamilies of evolved and designed networks", SCIENCE, 2004, 305, 1107

[5] Barabási A.-L., Linked – The New Science of Networks, Perseus, Cambridge MA, 2002

[6] Barabási A.-L., Albert R., Emergence of scaling in random networks, SCIENCE, 1999, 286, 509-512

[7] Barrat A., Barthélemy M., Vespignani A., Dynamical process on complex networks, Cambridge University Press, Cambridge, 2008

[8] Bollobás B., Modern Graph Theory, Springer Verlag, Heidelberg, Germany, 1998

[9] Bollobás B., Random Graphs, Cambridge Studies in Advanced Mathematics 73, 2nd edition, Cambridge University Press, London, 2001

[10] Bollobás B., Extremal Graph Theory, Dover Publications, 2004

[11] Borgatti S.P., Centrality and network flow, SOC NETWORKS, 2005, 27, 55-71

[12] Brandes U., Delling D., Gaertler M., Goerke R., Hoefer M., Nikoloski Z., Wagner D., Maximizing modularity is hard, Technical report, Dynamically Evolving, Large-Scale Information Systems DELIS-TR-379, 2006

[13] Brandes U., Gaertler M., Wagner D., Experiments on graph clustering algorithms, In: Di Battista G., Zwick U. (Eds.), Proceedings of the 11th European Symposium on Algorithms, LECT NOTES COMPUT SC, 2832 (15-20 September 2003, Budapest) 568-579

[14] Brandes U., Gaertler M., Wagner D., Engineering graph clustering: Models and experimental evaluation, ACM JOURNAL OF EXPERIMENTAL ALGORITHMICS, 2007, 12, Article 1.1

[15] Brin S., Motwani R., Ullman J.D., Tsur S., Dynamic itemset counting and implication rules for market basket data, In: Peckham J. (Ed.), Proceedings ACM SIGMOD International Conference on Management of Data (13-15 May 1997, Tucson, Arizona, USA), ACM Press, 1997, 255-264

[16] Brualdi R.A., Algorithms for constructing (0,1)-matrices with prescribed row and column sum vectors, DISCRETE MATH, 2006, 306, 3054-3062

[17] Chung F., Lu L., The average distances in random graphs with given expected degrees, P NATL ACAD SCI USA, 2002, 99, 15879-15882

[18] Chung F., Lu L., Complex Graphs and Networks, American Mathematical Society, Providence, RI, 2006

[19] Clauset A., Moore C., Newman M.E.J., Hierarchical structure and the prediction of missing links in networks, NATURE, 2008, 453, 98-101

[20] Clauset A., Newman M.E.J., Moore C., Finding community structure in very large networks, PHYS REV E, 2004, 70, 066111

[21] Clauset A., Shalizi C.R., Newman M.E.J., Power-law distributions in empirical data, SIAM REVIEW OF MODERN PHYSICS, 2009, 51, 661-703

[22] de Sola Pool I., Kochen M., Contacts and influence, SOC NETWORKS, 1978/79, 1, 5-51

[23] Deane C., Salwiński L., Xenarios I., Eisenberg D., Protein interactions: two methods for assessment of the reliability of high throughput observations, MOL CELL PROTEOMICS, 2002, 1, 349-356

[24] Dorogovtsev S.N., Mendes J.F.F., Evolution of Networks, Oxford University Press, New York, NY, 2003

[25] Easley D., Kleinberg J., Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Cambridge University Press, New York, NY, 2010

[26] Faust K., Wasserman S., Centrality and prestige: A review and synthesis, QUANTITATIVE ANTHROPOLOGY, 1992, 4, 23-78

[27] Fortunato S., Community detection in graphs, PHYS REP, 2010, 486, 75-174

[28] Freeman L., The development of social network analysis, Empirical Press, Vancouver, 2006

[29] Freeman L.C., Centrality in networks: I. conceptual clarifications, SOC NETWORKS, 1979, 1, 215-239
[30] Gaertler M., Clustering, Network Analysis: Methodological Foundations, In: Brandes U., Erlebach T. (Eds.), LECT NOTES COMPUT SC, vol. 3418, Springer-Verlag, New York, 2005
[31] Gilbert E.N., Random graphs, ANN MATH STAT, 1959, 30, 1141-1144
[32] Gionis A., Mannila H., Mielikäinen T., Tsaparas P., Assessing data mining results via swap randomization, ACM TRANSACTIONS ON KNOWLEDGE DISCOVERY FROM DATA, 2007, 1, 167-176
[33] Girvan M., Newman M.E.J., Community structure in social and biological networks, P NATL ACAD SCI USA, 2002, 99, 7821-7826
[34] Gonen M., Shavitt Y., Approximating the number of network motifs, LECT NOTES COMPUT SC, 2009, 5427, 13-24
[35] Han J., Kamber M., Data Mining - Concepts and Techniques, Morgan Kaufmann, San Francisco, 2006
[36] Holme P., Congestion and centrality in traffic flow on complex networks, ADV COMPLEX SYST, 2003, 6, 163-176
[37] Jackson M.O., Social and economic networks, Princeton University Press, 2010
[38] Jeong H., Mason S.P., Barabási A.-L., Oltvai Z.N., Lethality and centrality in protein networks, NATURE, 2001, 411, 41-42
[39] Kashani Z.R.M., Ahrabian H., Elahi E., Nowzari-Dalini A., Ansari E.S., Asadi S., Mohammadi S., Schreiber F., Masoudi-Nejad A., Kavosh: a new algorithm for finding network motifs, BMC BIOINFORMATICS, 2009, 10, 318-329
[40] Kashtan N., Itzkovitz S., Milo R., Alon U., Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs, BIOINFORMATICS, 2004, 20, 1746-1758
[41] Katz L., A new index derived from sociometric data analysis, PSYCHOMETRIKA, 1953, 18, 39-43
[42] Keller E.F., Revisiting "scale-free" networks, BIOESSAYS, 2005, 27, 1060-1068
[43] Kelvin W.T., Electrical units of measurement, Popular Lectures and Addresses (Vol. 1), London MacMillian, 1889
[44] Kleinberg J., The small-world phenomenon: An algorithmic perspective, Proceedings of the 32nd ACM Symposium on Theory of Computing (21-23 May 2000, Portland, OR, USA), 2000, 163-170
[45] Kolaczyk E.D., Statistical Analysis of Network Data: Methods and Models, Springer Verlag, Heidelberg, 2009
[46] Koschützki D., Lehmann K.A., Peeters L., Richter S., Tenfelde-Podehl D., Zlotowski O., Centrality Indices, In: Brandes U., Erlebach T. (Eds.), Network Analysis – Methodological Foundations, LECT NOTES COMPUT SC, 2832, Springer Verlag, New York, 2005
[47] Koschützki D., Lehmann K.A., Tenfelde-Podehl D., Zlotowski O., Advanced Centrality Concepts, In: Brandes U., Erlebach T. (Eds.), Network Analysis – Methodological Foundations, LECT NOTES COMPUT SC, 2832, Springer Verlag, New York, 2005
[48] Lerner J., Role Assignments, In: Brandes U., Erlebach T. (Eds.), Network Analysis – Methodological Foundations, LECT NOTES COMPUT SC, 2832, Springer Verlag, New York, 2005
[49] Liben-Nowell D., Kleinberg J., The link-prediction problem for social networks, J AM SOC INF SCI TEC, 2007, 58, 1019-1031
[50] Lichtenwalter R.N., Dame N., Lussier J.T., Chawla N.V., New Perspectives and Methods in Link Prediction, Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (25-28 July 2010, Washington DC, DC, USA), ACM, 2010
[51] Milo R., Kashtan N., Itzkovitz S., Newman M.E.J., Alon U., On the uniform generation of random graphs with prescribed degree sequences, preprint available at http://arxiv.org/abs/cond-mat/0312028
[52] Milo R., Kashtan N., Itzkovitz S., Newman M.E.J., Alon U., Subgraphs in networks, PHYS REV E, 2004, 70, 085102
[53] Milo R., Itzkovitz S., Kashtan N., Levitt R., Alon U., Response to comment on "network motifs: Simple building blocks of complex networks" and "superfamilies of evolved and designed networks", SCIENCE, 2004, 305, 1107
[54] Milo R., Itzkovitz S., Kashtan N., Levitt R., Shen-Orr S., Ayzenshtat I., Sheffer M., Alon U., Superfamilies of evolved and designed networks, SCIENCE, 2004, 303, 1538-1542
[55] Newman M.E.J., Networks: An Introduction, Oxford University Press, 2010
[56] Newman M.E.J., The structure of scientific collaboration networks, P NATL ACAD SCI USA, 2001, 98, 404-409
[57] Newman M.E.J., Fast algorithm for detecting community structure in networks, PHYS REV E, 2004, 69, 066133
[58] Newman M.E.J., Modularity and community structure in networks, P NATL ACAD SCI USA, 2006, 103, 8577-8582
[59] Newman M.E.J., Girvan M., Finding and evaluating community structure in networks, PHYS REV E, 2004, 69, 026113
[60] Nunkesser M., Sawitzki D., Blockmodels, In: Brandes U., Erlebach T. (Eds.), Network Analysis – Methodological Foundations, LECT NOTES COMPUT SC, 2832, Springer-Verlag, New York, 2005

[61] Padgett J.F., Ansell C.K., Robust action and the rise of the medici, AM J SOCIOL, 1993, 98, 1259-1319
[62] Palla G., Derényi I., Farkas I., Vicsek T., Uncovering the overlapping community structure of complex networks in nature and society, NATURE, 2005, 435, 814-818
[63] Przulj N., Corneil D.G., Jurisica I., Efficient estimation of graphlet frequency distributions in protein¸protein interaction networks, BIOINFORMATICS, 2006, 22, 974-980
[64] Scott J., Social Network Analysis, 2nd edition, reprinted edition, SAGE Publications, London, 2003
[65] Stauffer D., Aharony A., Introduction to Percolation Theory, CRC, Boca Raton, USA, 1994
[66] Tarkoma S., Overlay Networks: Toward Information Networking, Auerbach Publications, 2010
[67] Wasserman S., Faust K., Social Network Analysis – Methods and Applications, revised, reprinted edition, Cambridge University Press, Cambridge, 1999
[68] Watts D.J., Strogatz S.H., Collective dynamics of 'small-world' networks, NATURE, 1998, 393, 440-442
[69] Yook S.-H., Jeong H., Barabási A.-L., Modeling the internet's large-scale topology, P NATL ACAD SCI USA, 2002, 99, 13382-3386
[70] Zachary W.W., An information flow model for conflict and fission in small groups, J ANTHROPOL RES, 1977, 33, 452-473
[71] Zahoránszky L.A., Katona G.Y., Hári P., Málnási-Csizmadia A., Zweig K.A., Zahoránszky-Köhalmi G., Breaking the hierarchy – a new cluster selection mechanism for hierarchical clustering methods, ALGORITHM MOL BIOL, 2009, 4, 12
[72] Zweig K., Kaufmann M., A systematic approach to the one-mode projection of bipartite graphs, SOCIAL NETWORK ANALYSIS AND MINING, (in press)
[73] Zweig K.A., How to forget the second side of the story: A new method for the one-mode projection of bipartite graphs, Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining ASONAM 2010 (9-11 August 2010, Odense, Denmark), IEEE Computer Society, 2010, 200-207