Extremal Dependencies and Rank Correlations in Power Law Networks

Yana Volkovich¹, Nelly Litvak¹, and Bert Zwart²

¹ University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands {y.volkovich,n.litvak}@ewi.utwente.nl
² CWI, Science Park Amsterdam, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands bert.zwart@cwi.nl

Abstract. We analyze dependencies in complex networks characterized by power laws (Web sample, Wikipedia sample and a preferential attachment graph) using statistical techniques from the extreme value theory and the theory of multivariate regular variation. To the best of our knowledge, this is the first attempt to apply this well developed methodology to comprehensive graph data. The new insights this yields are striking: the three above-mentioned data sets are shown to have a totally different dependence structure between graph parameters, such as in-degree and PageRank. Based on the proposed approach, we suggest a new measure for rank correlations. Unlike most known methods, this measure is especially sensitive to rank permutations for top-ranked nodes. Using the new correlation measure, we demonstrate that the PageRank ranking is not sensitive to moderate changes in the damping factor.

Keywords: Extremal dependencies, Statistical analysis, Power laws, PageRank, Web, Wikipedia, Preferential attachment.

1 Introduction

What do we know about structure of complex networks? There is a vast literature on the subject but we are still far from complete understanding. One of the common points is the presence of power laws. In simple words, a power law with exponent $\alpha > 0$ means that a *tail probability*, i.e. the probability of obtaining a value greater than x, is roughly proportional to $x^{-\alpha}$. The standard example of a power law is a Pareto distribution

$$\mathbb{P}(X > x) = x_0^{\alpha} x^{-\alpha}, \ x \ge x_0,$$

where x_0 is the positive minimum possible value of X. The power law random variables are *heavy-tailed* [20], which means that the probability of obtaining extremely high values is relatively large. For instance, for such random variables, the moments of order α or higher are infinite. For excellent surveys on history, properties, modeling, and mining of power laws, and their role in complex networks we refer to e.g. [4,12,15,16,17].

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 1642–1653, 2009.

[©] ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2009

A natural mathematical formalism for analyzing power laws is provided by the theory of regular variation. This theory has been developed in the context of analysis of extremes [5], financial time series [14], and traffic in communication networks [19]. By definition, the random variable X is regularly varying with index α , if

$$\mathbb{P}(X > x) = x^{-\alpha} L(x), \ x > 0, \tag{1}$$

where L(x) is a slowly varying function, that is, for x > 0, $L(tx)/L(t) \to 1$ as $t \to \infty$, for instance, L(x) may be equal to a constant or $\log(x)$. Clearly, a power law can be modeled as an instance of regular variation.

The World Wide Web is an important example of a network with a complex graph structure characterized by power laws. The question of measuring correlations in the Web graph has led to many controversial results. Most notably, there is no agreement in the literature on the dependence between in-degree and Page-Rank of a Web page [6,9]. In this respect, Chakrabarti and Faloutsos [4] confirm that measuring correlation in power law data is tricky because the important large values do not appear very often, and thus, the coefficient of correlation might give a wrong impression about the dependencies in the tails. This merely confirms the common knowledge in the extreme value theory community [5,20] that the correlation coefficient is an uninformative dependence measure in heavytailed data. The correlation is a 'crude summary' of dependencies that is most informative for jointly normal random variables. It is a common and simple technique but it is not subtle enough to distinguish between the dependencies in large and in small values.

In this work we propose to employ the extreme value theory [2] and the theory of regular variation [20] that provide a range of statistical procedures designed to deal with multivariate data of which the marginal distributions exhibit power laws. In particular, this paper points out that this body of statistical theory contains a well-developed notion of dependence. This notion called *extremal dependence* is characterized by *spectral measure*, which seems to be much more suitable for the power law data than standard correlation measures. To the best of our knowledge, the proposed methods have never been applied to comprehensive graph data.

In this paper, we compute the spectral measures for in-degrees, out-degrees and PageRank scores in three large data sets: an EU-2005 Web sample, a Wikipedia sample and a Growing Network graph based on the preferential attachment model by Albert and Barabási [1]. Our experimental results reveal a dramatically different correlation structure in the three data sets. For instance, the results for in-degree and PageRank in Wikipedia strongly suggest an independence between these two parameters. Similar analysis for the Web graph reveals a non-trivial dependence structure. Finally, a preferential attachment graph shows a very strong dependence between in-degree and PageRank.

The analysis of extremal dependence leads us to propose a new rank correlation measure which is particularly plausible for power law data. The measure has the appealing property that it is especially sensitive to rank permutations for top-ranked nodes. Using the new correlation measure, we demonstrate that the PageRank ranking is not sensitive to moderate changes in the damping factor.

The paper is organized as follows. The data sets are described in detail in Section 2. Some background on extremal dependencies and the definition of spectral measure are given in Section 3. Further, in Section 4 we describe the estimator of the spectral measure. The experimental results are presented in Section 5. Based on the proposed approach, a new rank correlation measure is presented in Section 6, and a discussion is provided in Section 7.

2 Data Sets

We chose three data sets that represent different network structures. As the Web sample, we used the EU-2005 data set with 862.664 nodes and 19.235.140 links, that was collected by The Laboratory for Web Algorithmics $(LAW)^1$, the Università degli studi di Milano [3]. We also performed experiments on the Wikipedia (En) graph, that contains 4.881.983 nodes and 42.062.836 links. Finally, we simulated a Growing Network by using preferential attachment rule for 90% of new links [1]. The graph consists of 10.000 nodes with constant outdegree d = 8. In Figure 1 we show the cumulative log-log plots for in-degrees, out-degrees and PageRank scores in all data sets. The PageRank scores in the network of n nodes are computed according to the classical definition [18]:

$$PR(i) = c \sum_{j \to i} \frac{PR(j)}{d_j} + \frac{c}{n} \sum_{j \in \mathcal{D}} PR(j) + \frac{1 - c}{n}, \quad i = 1, \dots, n$$
(2)

where PR(i) is the PageRank of page $i, i = 1, ..., n; d_j$ is the number of outgoing links of page j; the sum is taken over all pages j that link to page $i; \mathcal{D}$ is a set of nodes without outgoing links; and c is the damping factor, which is equal 0.5 and 0.85 in our case. It is well-known that the linear system (2) has a unique solution for any $c \in (0, 1)$. The PageRank scores PR(i), i = 1, ..., n constitute a probability distribution on the set of n nodes. Throughout the paper however we use the scale-free PageRank scores nPR(i). In this setting, it is easier to compare the probabilistic properties of PageRank and in- and out-degree, that are also scale-free.

The log-log plots for Figure 1 resemble the signature straight lines indicating power laws. However, we suggest to combine several techniques in order to establish the presence of heavy tails and to evaluate the power law exponent. In [21] we used QQ plots, Hill and altHill plots as well as Pickands plots [20] to confirm that the in-degree and PageRank follow power laws with similar exponents for all three data sets. We also showed that the out-degree can be modeled reasonably well as a power law with exponent around 2.5-3.

Although all plots in Figure 1 look alike, it does not imply that the three networks have identical structure. In Section 5 we rigorously examine the dependencies between the network parameters.

¹ http://law.dsi.unimi.it; Accessed in January 2007.



Fig. 1. Cumulative log-log plots for in/(out)-degree, PageRank (c=0.5) and PageRank (c=0.85)

3 Extremal Dependencies and Spectral Measure

In this section we briefly present a background on extremal dependencies and spectral measure. Our presentation is based on Chapter 8 of Beirlant et al. [2], where an interested reader can find an in-depth treatment of this subject in the general framework of the extreme value theory.

Let X and Y be two possibly dependent non-negative random variables with complementary distribution functions

$$\overline{F}_X(x) = \mathbb{P}(X > x), \quad \overline{F}_Y(y) = \mathbb{P}(Y > y), \quad x, y > 0.$$

Denote

$$X_* = \frac{1}{\overline{F}_X(X)}, \quad Y_* = \frac{1}{\overline{F}_Y(Y)}.$$

It is easy to see that the distribution of X_* and Y_* is Pareto with $\alpha = 1$. Now, we apply the polar coordinate transformation with respect to L_2 norm

$$(R, \Theta) = \operatorname{Polar}(X_*, Y_*),$$

where $\operatorname{Polar}(x,y) = \left(\sqrt{x^2 + y^2}, \operatorname{arctan}(y/x)\right)$. Clearly, R_t is positive and Θ takes values in $[0, \pi/2]$. Furthermore, there exists a limiting measure S such that

$$\lim_{t \to \infty} t \mathbb{P}(R > t, \Theta \in A) = S(A) \tag{3}$$

for any Borel set $A \subset [0, \pi/2]$ that has a compact closure the boundary of measure zero with respect to S. The measure S is called *angular* or *spectral* measure and is used to describe so-called *extremal dependencies* i.e. the dependencies between extremely large values of X and Y. Furthermore, $S([0, \pi/2])$ is finite, and we use

$$\tilde{S}(A) = S(A)/S([0, \pi/2])$$

as a probability spectral measure, where Θ_* is a random variable that has distribution $\tilde{S}(A)$. We also note that the spectral measure is well defined for any norm $|| \cdot ||$ in \mathbb{R}^2 and can be extended to \mathbb{R}^n .

Unlike the correlation coefficient, the angular measure provides a subtle characterization of the dependencies in the tails of X and Y. Indeed, the limit in (3) as $t \to \infty$ describes the distribution of Θ provided that R is large, which can only happen when X_* or Y_* are large. In other words, the measure S characterizes the limiting behavior of the joint distribution of very high quantiles of X and Y. One may also notice that this approach is closely related to copulas that are commonly used e.g. in financial mathematics to describe the joint distribution of a random vector [2]. As a result, we expect to obtain different distributions of \tilde{S} for vectors with different dependence structure. In particular, it is easy to check that in case of a complete dependence X = Y we have $\tilde{S}(\{\pi/4\}) = 1$. Furthermore, in case of a complete independence we obtain $\tilde{S}(\{0\}) = 1/2$ and $\tilde{S}(\{\pi/2\}) = 1/2$ since in this case, for large t, the event [R > t] is most likely to occur, roughly, if either $[X_* > t]$ and Y_* is small, or $[Y_* > t]$ and X_* is small. All other distributions of \tilde{S} correspond to intermediate cases between complete dependence.

4 Spectral Measure Estimator

The equation (3) also gives a recipe for an estimator of the spectral measure $S(\cdot)$ [2]. Consider a pair of graph parameters (X, Y), and let X_j and Y_j be observations of X and Y for the corresponding node j. In order to estimate the angular measure, we start by using the rank transformation of (X, Y), leading to $\{(r_j^x, r_j^y), 1 \le j \le n\}$, where r_j^x is the descending rank of X_j in (X_1, \ldots, X_n) and r_j^y is the descending rank of Y_j in (Y_1, \ldots, Y_n) . The motivation for this rank transformation is that $(r_j^x/n, r_j^y/n)$ is a natural estimator of $(\overline{F}_X(X_j), \overline{F}_Y(Y_j))$.

Next, we choose k = 1, ..., n and apply the polar coordinate transform as follows

$$\operatorname{Polar}\left(\frac{k}{r_{j}^{x}}, \frac{k}{r_{j}^{y}}\right) = (R_{j,k}, \Theta_{j,k}).$$

$$\tag{4}$$



Fig. 2. Starica plot for in-degree and PageRank (c = 0.85)

Now, taking t = n/k in (3) and interpreting $R_{i,k}$ as a realization of kR/n, we obtain the following estimator for $\tilde{S}(A)$:

$$\frac{\sum_{i=1}^{n} \mathbf{1}[R_{i,k} > 1, \Theta_{i,k} \in A]}{\sum_{i=1}^{n} \mathbf{1}[R_{i,k} > 1]},$$
(5)

where $\mathbf{1}[\cdot]$ is an indicator function. As an outcome, we make a plot for a cumulative distribution function of Θ given by the above estimator. More details can be found in Chapters 8 and 9 of [2], and Chapter 9 of [20].

It was proved in [2,20] that the empirical measure converges to a proper distribution on $[0, \pi/2]$ as $n, k \to \infty$, $k/n \to 0$. That is, ideally, we need to consider only a relatively small part of a large data set. In practice the problem remains: how to choose a suitable value of k? In the case of bivariate regular variation, this can be determined by making a *Starica* plot. This technique helps to determine where the bivariate power law behavior actually 'starts'. To make the Starica plot, we consider radii $R_{1,k}, \ldots, R_{n,k}$ from (4) and rank them in descending order $R_{(1)} \geq \ldots \geq R_{(n)}$. To get Starica plot we graph

$$\left\{ \left(\frac{R_{(j)}}{R_{(k)}}, \frac{R_{(j)}}{R_{(k)}} \cdot \frac{j}{k}\right), \ 1 \le j \le n \right\}, \text{ or } \left\{ \left(R_{(j)}, \frac{R_{(j)}j}{\sum_{i=1}^{n} \mathbf{1}_{\{R_{i,k} \ge 1\}}}\right), \ 1 \le j \le n \right\}.$$

The idea is that for suitable k the ratio in the ordinate should be roughly a constant and equal 1 for the values of the abscissa in the neighborhood of 1. The plot looks different for the different parameters k and one can either find a suitable k by trial and error or use numerical algorithms to compute optimal k. A Starica plot for good k will have a region in the right neighborhood of x = 1 where the plot is hugging the y = 1 line. If the line is going steep up at x = 1 then the chosen k is too large. On the other hand, if the graph stabilizes around y = 1 for some x < 1 then it means that k is too small, and we miss some valuable tail data. We refer to Resnick [20] for more details and references.

In Figure 2 we present Starica plots for the pair of in-degree and PageRank (c=0.85). The plots behave nicely in all three data sets, which makes our angular measure more reliable. The Growing Network exhibits an ideal Starica plot (Figure 2(c)). In [21] we provided the plots and the appropriate values of k for the other combinations: in-degree and PageRank (c = 0.5), in-degree and out-degree; and out-degree and PageRank (c = 0.5, c = 0.85).

5 Dependence Measurements on the Data

After defining a suitable k, we compute the pairwise angular measure. In Figure 3 we depict $\theta \in [0, \pi/2]$ against the estimated probability spectral measure $[\theta, \pi/2]$, which, according to (5), equals to the fraction of pages i where the angle $\Theta_{i,k}$ is greater or equal to θ provided that $R_{i,k} > 1$.

The results are striking. Let us first look at Figures 3(a) and (b) that characterize the dependence between in-degree and PageRank. For the Wikipedia data set we observe that about half of the observations are concentrated around 0 whereas another half is close to $\pi/2$. This suggests an independence of the tails of in-degree and PageRank (c=0.85 and c=0.5). That is, in Wikipedia data set an extremely high in-degree almost never implies an extremely high ranking. The picture is completely the opposite for Growing Networks, where the angular measure is entirely concentrated around $\pi/4$ indicating a complete dependence. Thus, in highly centralized preferential attachment graphs, most connected nodes are also most highly ranked.

Finally, the Web graph exhibits a subtle dependence structure that results in angular measure which is almost uniform on $[0, \pi/2]$. This suggest that Page-Rank popularity measure can not be replaced by in-degree without significant disturbance in the ranking (of course, in-degree can not be used as a popularity measure for many other reasons, for instance, because it is easy to spam by creating link farms; we refer to [11] for further discussion of PageRank and other popularity measures).

The picture is different in Figure 3(c) where we depict the angular measure for in-degree and out-degree in the Web and in Wikipedia. In the Web, the inand out-degree tend to be independent which justifies the distinction between hubs and authorities [10]. In Wikipedia the in- and out-degrees are dependent but this dependence is not absolute.



Fig. 3. Cumulative functions for Angular Measures

Finally, the dependence between out-degree and PageRank in the Web and Wikipedia in Figure 3(d) resembles the patterns observed for in-degree and Page-Rank.

6 The Θ Rank Correlation Measure

We start by noting again that the estimator of the angular measure described in Section 4 is based on a rank transformation. This is clearly seen from formula (4) where only rank of the parameters X and Y plays a role. This observation naturally leads to a new measure for rank correlations.

In summary, our idea is as follows. As before, we define r_i^1 and r_i^2 as a ranking order of page *i* in scheme 1 and 2, respectively, where $i = 1 \dots n$. Now we suggest to represent the difference between the two rank positions of *i* by the angle

$$\Theta_i = \arctan(r_i^1/r_i^2).$$

For example, in Figure 4, Θ_i is depicted for a node that has rank 3 in scheme 1 and rank 6 in scheme 2. Note that this is exactly the angle in $[0, \pi/2]$ computed in (4) in order to construct the spectral measure estimator. The value Θ close to



Fig. 4. Rank Correlation

 $\pi/4$ means a relatively small change in ranking. On the other hand, Θ around $\pi/2$ means that the node *i* is much better off with scheme 2, and the value close to 0 says that the node is ranked much higher by scheme 1. Thus, we actually measure the rank difference for node *i* in radians! Having computed Θ_i for every *i* (or for a certain set of highly ranked nodes *i*) we construct a corresponding empirical cumulative distribution function for Θ . As in the previous section, the resulting angular measure can be used to characterize the rank correlations.

We note that we characterize the rank correlation by a measure or a plot rather than a number. Compared to the common rank correlation measures such as Kendall's τ and Spearman's ρ , our proposed measure has an important advantage that it is able to reveal the slightest rank disturbance among highly rank nodes while neglecting even moderate perturbations among lowly ranked nodes. Indeed, Kendall's τ and Spearman's ρ are defined as

$$\tau = 1 - \frac{2d_{\Delta}}{n(n-1)}$$
, and $\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$,

where d_{Δ} is the number of pairs in the symmetric difference of $\{(r_i^1, r_j^1), 1 \leq i < j \leq n\}$ and $\{(r_i^2, r_j^2), 1 \leq i < j \leq n\}$; and $d_i = r_i^1 - r_i^2$ is the difference between two ranks of page *i*. Now, if we consider the nodes ranked $1, \ldots, n$, and swap the ranks 1 and 10, then we get $\tau = 1 - 2 * 18/n(n-1), \rho = 1 - 6 * 162/n(n^2 - 1),$ and for our correlation measure at node 1 we obtain $\Theta_1 = \arctan(1/10) \approx 0.1$, that is close to the *x*-axis, and is a visible deviation from $\pi/4$. On the other hand, swapping the numbers 1001 and 1010 yields the same values of τ and ρ , but $\Theta_{1001} = \arctan(1001/1010) \approx \pi/4$. Thus, the Θ rank correlation measure actually evaluates the rank disturbance visible for users. Certainly, the $\arctan(\cdot)$ function makes our measure symmetric with respect to the schemes 1 and 2.

Naturally, in this framework, it is also possible to compute such angular measure only for the top ranked pages. This can be done along the same lines as in Section 4 as follows. Based on the polar transformation (4) we can separate top ranked pages by considering only points $\{\Theta_{i,k} : R_{i,k} > 1\}$. Here the question



Fig. 5. Cumulative functions for Angular Measures for PageRank (c=0.5) and PageRank (c=0.85)

of choosing k does not arise anymore. Indeed, the technique involving Starica plot was needed to get an idea where the power law behavior 'starts' in order to measure statistical dependency for the heavy-tailed data as in [20]. On the other hand, if we are interested in rank correlations, we may simply pick the k that gives us the top proportion of pages we are interested in. Note that by increasing k we do not change the observed values of Θ , we merely increase their number. As a result, in the angular measure, each observation will simply have less weight. On contrary, decreasing k means 'zooming in' the rank perturbations on the top.

One more advantage of the proposed correlation measure is the fast and easy implementation since for each node *i*, only the fraction r_i^1/r_i^2 has to be computed.

Below we present the example of the proposed rank correlation measure in Growing Networks, Web and Wikipedia. We rank the three data sets by using the definition of PageRank (2), where the damping factor is equal to c = 0.5 and c = 0.85. In Figure 5 we plot cumulative functions for angular measures for k = 100 and the values of k's that have been chosen according to the Starica plots as described in Section 4 (see e.g. Figure 2).

For Growing Network data set we observe the strong correlation between ranking schemes. We can also conclude that in Wikipedia the change in the damping factor affects only about 20% of considered pages, in the top-hundred group as well as in the larger group. For the Web data, the correlation between ranking is not significant for approximately half of the pages. However, for the top pages, the difference in the damping factor mixes up the order of ranking.

7 Discussion

Analysis of dependencies in real-life graph and synthetic data contributes towards a better understanding and modeling of complex graph structures. Clearly, for adequate modeling, it is not sufficient to maintain power laws. For instance, it was already argued in [7] that robustness of Internet power law router graph is in strong disagreement with a preferential attachment model. Likewise, our analysis clearly reveals a striking disagreement of the preferential attachment graph with dependence structure of the Web and Wikipedia. Better models have to be sought and existing models have to be thoroughly analyzed before we can conclude that they adequately reflect important features of complex networks.

The idea of a spectral measure estimator is naturally extended to yield the Θ rank correlation measure. The main idea of this measure is that we characterize the rank correlations by a cumulative distribution of Θ_i 's, where $i = 1, \ldots n$. This way, one can actually see how many pages change their ranks significantly. Such measure is substantially more informative than just one number, that represents the correlation in the whole graph. For instance, Melucci [13] noticed that Kendall's τ tends to grow close to one for large data sets. The author provides an example where Kendall's τ for ranking orders of only a few hundred Web pages becomes almost 1, in spite of the large number of rank perturbations. We remark however that if for some reason having one number is necessary, one can always compute, e.g. the expected deviation of Θ from $\pi/4$.

As mentioned before, the proposed correlation measure is quite harsh with respect to lowly ranked nodes. Indeed, the node ranked 1000 must fall all the way to 2000 to make the same effect as number 1 becoming number 2. We would like to emphasize that such discrepancy is especially suitable for ranking order emerging from a heavy-tailed data, such as PageRank or in-degree. This is because in such data, there is a huge difference between the highest values of the realizations, cf. [8].

Acknowledgements

This work is supported by NWO Meervoud grant no. 632.002.401. Part of this research has been funded by the Dutch BSIK/BRICKS project. This article is also the result of joint research in the 3TU Centre of Competence NIRICT (Netherlands Institute for Research on ICT) within the Federation of Three Universities of Technology in The Netherlands.

References

- Albert, R., Barabási, A.L.: Emergence of scaling in random networks. Science 286, 509–512 (1999)
- 2. Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: Statistics of Extremes: Theory and Applications. Wiley, Chichester (2004)
- Boldi, P., Vigna, S.: The WebGraph framework I: Compression techniques. In: Proc. of the Thirteenth International World Wide Web Conference (WWW 2004), pp. 595–601 (2004)
- 4. Chakrabarti, D., Faloutsos, C.: Graph mining: Laws, generators, and algorithms. ACM Comput. Surv. 38(1), 2 (2006)
- de Haan, L., de Ronde, J.: Sea and wind: multivariate extremes at work. Extremes 1(1), 7–45 (1998)
- Donato, D., Laura, L., Leonardi, S., Millozi, S.: Large scale properties of the Webgraph. Eur. Phys. J. 38, 239–243 (2004)
- Doyle, J.C., Alderson, D.L., Li, L., Low, S., Roughan, M., Shalunov, S., Tanaka, R., Willinger, W.: The robust yet fragile nature of the Internet. PNAS 102(41), 14497–14502 (2005)
- 8. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events. Springer, Heidelberg (1997)
- Fortunato, S., Boguñá, M., Flammini, A., Menczer, F.: Approximating PageRank from in-degree. In: Aiello, W., Broder, A., Janssen, J., Milios, E.E. (eds.) WAW 2006. LNCS, vol. 4936, pp. 59–71. Springer, Heidelberg (2008)
- Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. JACM 46(5), 604–632 (1999)
- 11. Langville, A.N., Meyer, C.D.: Google's PageRank and beyond: the science of search engine rankings. Princeton University Press, Princeton (2006)
- Li, L., Alderson, D.L., Doyle, J.C., Willinger, W.: Towards a theory of scalefree graphs: definition, properties, and implications. Internet Math. 2(4), 431–523 (2005)
- Melucci, M.: On rank correlation in information retrieval evaluation. SIGIR Forum 41(1), 18–33 (2007)
- Mikosch, T.: Modelling dependence and tails in financial time series. In: Symposium in Honour of Ole E. Barndorff-Nielsen (Aarhus, 2000). Memoirs, vol. 16, pp. 61–73. Univ. Aarhus, Aarhus (2000)
- 15. Mitzenmacher, M.: A brief history of generative models for power law and lognormal distributions. Internet Math. 1(2), 226–251 (2004)
- Newman, M.E.J.: The structure and function of complex networks. SIAM Rev. 45(2), 167–256 (2003)
- Newman, M.E.J.: Power laws, Pareto distributions and Zipf's law. Contemp. Phys. 46, 323–351 (2005)
- Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: Bringing order to the Web. Technical report, Stanford Digital Library Technologies Project (1998)
- Park, K., Willinger, W.: Self-similar network traffic and performance evaluation. Wiley, Chichester (2000)
- 20. Resnick, S.I.: Heavy-tail Phenomena. Springer, New York (2007)
- Volkovich, Y., Litvak, N., Zwart, B.: A framework for evaluating statistical dependencies and rank correlations in power law graphs. Memorandum 1868, University of Twente, Enschede (2008)