Physica A 388 (2009) 2221–2226

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# Lobby index in networks

A. Korn [a], A. Schubert [b], A. Telcs [c,*]

[a] Department of Telecommunications and Media Informatics, University of Technology and Economics Budapest, 1117 Budapest, Magyar tudósok körútja 2, Hungary

[b] Institute for Research Organisation, H-1245 Budapest, P.O.Box 994, Hungary

[c] Department of Computer Science and Information Theory, University of Technology and Economics Budapest, 1117 Budapest, Magyar tudósok körútja 2, Hungary

ARTICLE INFO

*Article history:*
Received 10 November 2008
Received in revised form 28 January 2009
Available online 25 February 2009

*Keywords:*
Scale-free networks
Centrality measures

ABSTRACT

We propose a new node centrality measure in networks, the lobby index, which is inspired by Hirsch's $h$-index. It is shown that in scale-free networks with exponent $\alpha$ the distribution of the $l$-index has power tail with exponent $\alpha\,(\alpha+1)$. Properties of the $l$-index and extensions are discussed.



Efficient communication means high impact (wide access or high reach) and low cost. This goal is common in communication networks, in society and in biological systems. In the course of time many centrality measures have been proposed to characterize a node's role, position, or influence in a network but none of them capture the efficiency of communication. This paper is intended to fill this gap and propose a new centrality measure, the lobby index.

Hirsch [1] proposed the $h$-index: "*the number of papers with citation number $\geq h$, as a useful index to characterize the scientific output of a researcher*". Barabási et al. [2] devised a very simple network model which has several key properties: most importantly *the degree distribution has a power-law upper tail, the node degrees are independent, and typical nodes are close to each other.* Schubert et al. [3] used the $h$-index as a network indicator, particularly in scale-free networks. This paper is devoted to the characterization of network nodes with an $h$-index type measure.

**Definition 1.** The $l$-index or **lobby index** of a node $x$ is the largest integer $k$ such that $x$ has at least $k$ neighbors with a degree of at least $k$. (See also (1).)

In what follows some properties of the lobby index are investigated; it is shown that in **S**cale **F**ree (SF) networks, with exponent $\alpha$, the distribution of the $l$-index has a fat tail with exponent $\alpha\,(\alpha+1)$. Furthermore the empirical distribution of the $l$-index in generated and real life networks is investigated and some further extensions are discussed.

## 1. Centrality measures

Freeman's prominent paper [4] (1979) pointed out that: "*Over the years, a great many measures of centrality have been proposed. The several measures are often only vaguely related to the intuitive ideas they purport to index, and many are so complex that it is difficult or impossible to discover what, if anything, they are measuring*". It is perhaps worth noting that research in this field dates back to Bavelas [5] (1949).

* Corresponding author. Tel.: +36 1 463 3189.
*E-mail addresses:* korn.andras@tmit.bme.hu (A. Korn), schuba@iif.hu (A. Schubert), telcs@szit.bme.hu (A. Telcs).

0378-4371/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2009.02.013

At the time of Freeman's paper most centrality measures were equivalents or modifications of the three major and widely accepted indexes, the degree (cf. Ref. [6]), closeness (cf. Refs. [7,8]) and betweenness (cf. Ref. [5]) centrality (see also Borgatti [9]).

As time passed, many new centrality measures were proposed. After years of research and application, the above three and eigenvector centrality (a variant of which computer scientists call PageRank [10] and Google uses to rank search results) can be said to have become a standard; the others are not widely used. The historical three and eigenvector centrality are thus the conceptual base for investigating centrality behavior of nodes and full networks.

Notwithstanding Freeman's wise warning the present paper proposes the lobby centrality (index) in the belief that Hirsch's insight into publication activity (which produces the citation network) has an interesting and relevant message to network analysis in general.

*The diplomat's dilemma.* It is clear that a person has strong lobby power, the ability to influence people's opinions, if he or she has many highly connected neighbors. This is exactly the aim of a lobbyist or a diplomat [11]. The diplomat's goal is to have strong influence on the community while keeping the number of his connections (which have a cost) low. If $x$ has a high lobby index, then the $l$-core $L(x)$ (those neighbors which provide the index) has high connectivity (statistically higher than $l(x)$, see (6) and the comment there). In this sense, the $l$-index is closely related to the solution of the diplomat's dilemma.

*Communication networks.* Research of communication networks and network topology is in interaction. Node centrality measures are essential in the study of net mining [12], malware detection [13], in reputation-based peer-to-peer systems [14], delay tolerant networks [15] and others (see Refs. [16,17] and the references therein). We expect that in the case of social and communication networks (some of which are also based on social networks) the lobby-index is located between the bridgeness [18], closeness, eigenvector and betweenness centrality. Based on this intermediate position of the lobby index we expect that it can be a useful aid in developing good defence and immunization strategies for peer-to-peer networks as well as help create more efficient broadcasting schemes in sensor networks and marketing or opinion shaping strategies.

*The distribution of the l-index.* Let us consider scale-free networks and assume that the node degrees are independent. The degree is denoted by $\deg(x)$ for nodes and the $l$-index is defined as follows. Let us consider all $y_i$ neighbors of $x$ so that $\deg(y_1) \geq \deg(y_2) \cdots$; then,

$$l(x) = \max\{k : \deg(y_k) \geq k\}. \tag{1}$$

**Theorem.** *If the vertex degrees are independent and* $\mathbb{P}(\deg(x) \geq k) \approx ck^{-\alpha}$ *for all nodes x, then*

$$\mathbb{P}(l(x) \geq k) \simeq k^{-\alpha(\alpha+1)} \tag{2}$$

*for all nodes x.*[1]

The proof is provided in the Appendix.

*The Hirsch index.* The original Hirsch index is based on a richer model: author $\leftrightarrow$ paper and paper $\leftrightarrow$ citing paper links. Let $x$ be a randomly chosen author of the scientific community under scrutiny and $n = n(x)$ is the number of his/her papers (either in general or within a defined period). Let $y_i$ denote the individual papers (where $i = 1, \ldots n$,) and $c(y_i)$ their citation score (in decreasing order), so that $c(y_1) \geq c(y_2) \geq \cdots \geq c(y_n)$. $h(x)$ is the Hirsch index of $x$ :

$$h(x) = \max\{k : c(y_k) \geq k\}.$$

Assume that the paper productivity has an $\alpha$-fat tail: $G_l^P = \mathbb{P}(n(x) \geq l) \approx cl^{-\alpha}$ and the citation score has a $\beta$-fat tail:

$$G_l^Q = \mathbb{P}(c(y) \geq l) \approx cl^{-\beta}.$$

Along the lines of the argument that led to (2) one can see that $h$ has an $\alpha(\beta+1)$-fat tail [19]:

$$\mathbb{P}(h(x) \geq k) \simeq k^{-\alpha(\beta+1)}. \tag{3}$$

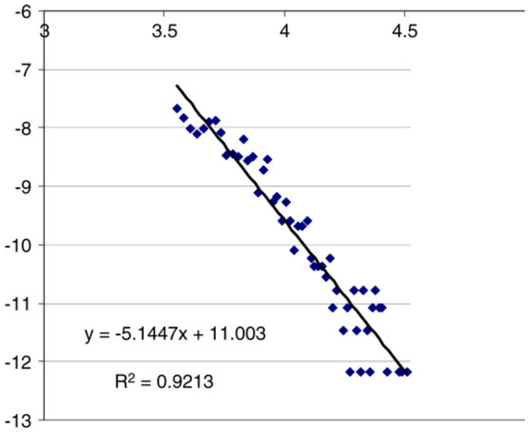
*How good is an l-index of k?* If a node $x$ of degree $n$ has an $l$-index of $k$, Glänzel's [20] observation provides a preliminary assessment of this value:

$$l(x) \approx c \deg(x)^{\frac{1}{\alpha+1}} \tag{4}$$

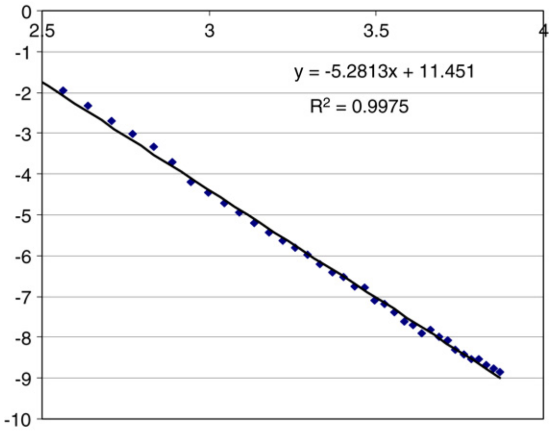
where $\alpha$ is the tail exponent of the degree distribution. Consequently a lobbyist is doing a good job of solving the diplomat's dilemma if $l(x) \gg \deg(x)^{\frac{1}{\alpha+1}}$. On the other hand our result shows that $l(x) \geq k$ means that $x$ belongs to the top $100c_\alpha k^{-\alpha(\alpha+1)}$ percent of lobbyists.

*The lobby gain.* The performance of a lobbyist is indicated by a measure called the lobby gain. The lobby gain shows how the access to the network is multiplied using a link to the $l$-core. Let us use the notation $D_i(y) = \{z : d(x, y) = i\}$ and set $D_2^L(x) = \cup_{y\in L(x)} D_1(y) \setminus [D_1(x) \cup \{x\}]$ then the number of second neighbors reachable via the $l$-core is $\deg_2^L(x) = \left|D_2^L(x)\right|$

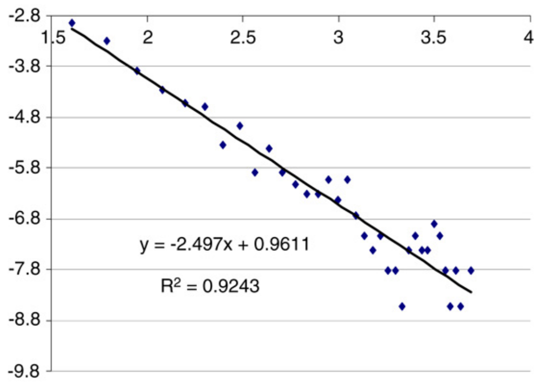
[1] Here and in what follows $a_n \approx b_n$ means that $\frac{a_n}{b_n} \to c$ as $n \to \infty$ and $a_n \simeq b_n$ means that there is a $C > 1$ such that for all $n$, $\frac{1}{C} \leq \frac{\alpha_n}{b_n} \leq C$.



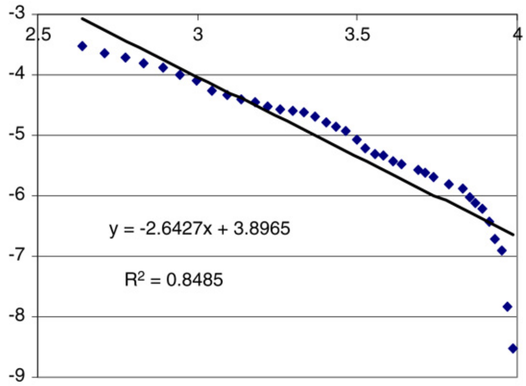

(a) The BA graph.




(b) The GBA graph.




(c) The AS graph.




(d) The IG graph.

**Fig. 1.** The log–log plots for the distribution of $l$-index.

and the lobby gain is defined as

$$\Gamma_l(x) = \frac{\deg_2^L(x)}{l(x)}. \tag{5}$$

The lobby gain $\Gamma_l(x)$ is much larger than one if a typical link to the $l$-core provides a lot of connections to the rest of the network for $x$ via that link. It can be shown (see Ref. [20]) that the number of second neighbors reachable via the $l$-core (with multiplicity) is $l(x)^2$.

*The degree distribution within the l-core.* The influential acquaintances of a given lobbyist follow a fat tail distribution provided the underlying network is SF. In other words if $y \in L(x)$ and $l = k$ then the truncated distribution (by $k$) of the degree distribution of $y$ again follows a fat tail distribution: for $m > k > 0$

$$\mathbb{P}(\deg(y) \geq m | y \in L(x) \text{ and } l = k) \approx c\left(\frac{m}{k}\right)^{-\alpha}. \tag{6}$$

Let us note that this conditional or truncated distribution has a higher expected value than the original one.

## 2. Network examples

The analysis of different networks received particular attention in the last decades. The research goals and tools vary greatly. Here we regress to the roots and consider some "classic" networks and study the distribution of their lobby index.

*Generated scale-free networks.* We have generated 50 20 000-node Barabási (**BA**) networks [2] with 10 new links each step, starting with 10 initial nodes. The degree distribution passed the preliminary test and has $\alpha = 1.96$, i.e. a 1.96-fat tail. As Fig. 1(a) shows, the empirical distribution of the lobby index has $\widehat{\eta} = 5.14$ while the theory predicts $\eta = 5.76$.

We have used the generalized Barabási model (**GBA**) [21] which can provide arbitrary $\alpha > 1$. We generated 50 graphs of 10 000 nodes with the proposed algorithm and obtained networks with $\alpha = 1.9186$ in average, which would imply $\eta = 5.60$; we observed $\widehat{\eta} = 5.28$.

**Table 1**
Spearman rank correlation.

| | $l$ | $cl$ | $bw$ | $ev$ |
|---|---|---|---|---|
| $l$ | 1 | .652 | .768 | .604 |
| $cl$ | *.652* | 1 | .500 | .972 |
| $bw$ | *.768* | **.500** | 1 | .479 |
| $ev$ | *.604* | **.972** | **.479** | 1 |

**Table 2**
The tail exponents of networks.

| | $\alpha$ | $\eta$ | $\widehat{\eta}$ |
|---|---|---|---|
| BA graph | 1.96 | 5.76 | 5.14 |
| Generalized BA | 1.92 | 5.60 | 5.28 |
| AS | 1.61 | 4.21 | 4.14 |
| IG | 1.23 | 2.74 | 2.45 |

Let us remark that the estimate of the tail exponent of fat tail distributions has a sophisticated technique [22] superior to the line fit on the log–log scale. The careful analysis and application of these methods to the $l$-index will be published elsewhere.

*The AS level graph.* The **A**utonomous **S**ystem (**AS**) level of the Internet infrastructure has already been investigated in depth (c.f. Ref. [23] and its bibliography). It turned out that it not only has a scale-free degree distribution but displays the rich club phenomenon as well. High degree nodes are more densely interlinked than expected in a **BA** graph. The standard choice for **AS** a source of sample data is the CAIDA [24] project. We determined the exponent of the tail of the degree distribution and compared it with the exponent of the tail of the empirical distribution of the $l$-index. We found that $\alpha = 1.61$ and $\eta = \alpha\,(\alpha + 1) = 4.21$ and $\widehat{\eta} = 4.14$ is estimated from the empirical distribution of $l$.

*The IG model.* Mondragón et al. [23] proposed a modification of the Barabási network model, the **I**nteractive **G**rowth (**IG**) model to generate scale-free networks which exhibit the rich club behavior. In each iteration, a new node is linked to one or two existing nodes (hosts). In the first case the host node is connected to two additional peer nodes using the preferential attachment scheme while in the latter case only one of the involved (randomly chosen) hosts is connected to a new peer. We implemented this algorithm and again compared the exponents extracted from sample data. In this case the network size was 3000 and the probability of one host was 0.4 and of two hosts was 0.6. Again the log–log fit of the degree distribution tail yielded $\alpha = 1.23$, $\eta = 2.74$ and $\widehat{\eta} = 2.45$ given by the empirical distribution of $l$.

*The place of the lobby index among other centrality measures.* As we already indicated above the lobby index lies somewhere between the closeness centrality ($cl$), betweenness ($bw$) and eigenvector ($ev$) centrality. Strong correlation with degree centrality is out of the question in the light of (4). In order to gain a better picture on the behavior of the lobby index we determined the Spearman correlation between these centrality measures in the AS graph.

The correlations in Table 1 indicate that the $l$-index contains a well-balanced mix of other centrality measures; the $l$-index is slightly closer to the three "classical" centralities than they are to each other (the quadratic mean of the three correlations, in boldface, is 0.638 while the quadratic mean of of the correlations, in italic, with the $l$-index is 0.678). The Kendall correlations of the investigated centralities have been calculated and yielded a very similar picture. For biological networks the Spearman correlation between the closeness and eigenvector centrality is high (c.f. Ref. [25]); high Pearson correlation can be observed in other networks as well (c.f. Ref. [26]). One centrality measure can be used to approximate the other, which is not the case with the $l$-index for the AS graph, but may happen in other types of networks. This will save computation time given the simplicity of the calculation of the $l$-index. Freeman's paper [4] as well as Refs. [27,26] are calls for a further analysis of centrality measures and the $l$-index in different types of networks.

*Conclusion.* A new centrality measure, the $l$-index is proposed and examined. It is shown that the distribution of the $l$-index has $\alpha\,(\alpha + 1)$-fat tail in **SF** networks with exponent $\alpha$. There is a good match between empirical observations (collected in Table 2) and the theoretical result. In this case the aim of the empirical results was not to verify the theory but to emphasize that the investigated networks behave in the expected way with respect to the lobby index.

The lobby index is placed on the map along other centrality measures. Some further extensions and properties are discussed as well: the relation to the diplomat's dilemma is investigated and the lobby index is demonstrated to be a good performance measure for lobbyists.

## Acknowledgments

We thank Tamás Nepusz who kindly provided the calculations of the classical centralities and their correlations in Table 1 for the AS graph and gave same valuable remarks on the manuscript. We also thank Fülöp Bazsó for his useful comments on the draft. Thanks are due to Tommaso Colozza and Erik Bodzsár who gave very valuable numerical calculations and simulation results for this work. The authors were partially supported by the High Speed Network Laboratory of the Budapest University of Technology and Economics. The authors also wish to thank the referees for their helpful comments and suggestions.

## Appendix

In what follows we provide a rigorous derivation of (2).[2] Let us use the notation $l_k = \mathbb{P}\left(l\left(x\right) = k\right)$ for the distribution of the $l$-index, and $G_k = \mathbb{P}\left(\deg\left(x\right) \geq k\right) = 1 - F_k$

$$\begin{aligned} l_k &= \sum_{l=0}^{\infty} \mathbb{P}\left(l\left(x\right) = k, \deg\left(x\right) = k + l\right) \\ &= \sum_{l=0}^{\infty} \mathbb{P}\left(l\left(x\right) = k \mid \deg\left(x\right) = k + l\right) \mathbb{P}\left(\deg\left(x\right) = k + l\right). \end{aligned}$$

Partition of unity and conditional probability is used (Bayes Theorem). As a result we have to investigate what is the probability that a node has $k$ links, each has degree $\geq k$ and $l$ other links with degree not higher than $k$ given that it has $k+l$ links in total. That criteria makes exactly $l\left(x\right) = k$.

First we develop a lower estimate for $l_k$.

$$l_k \geq c_1^k k^{-\alpha k} \sum_{l=0}^{\infty} \left(k+l\right)^{-(\alpha+1)} \binom{k+l}{l} \left(1 - c_1 k^{-\alpha}\right)^l.$$

We estimate $l_k$ using $1 - c_1 k^{-\alpha} \approx \mathrm{e}^{-c_1 k^{-\alpha}}$ and $\binom{k+l}{l} \geq \frac{l^k}{k!}$

$$\begin{aligned} l_k &\geq c \frac{c_1^k k^{-\alpha k}}{k!} \sum_{l=0}^{\infty} l^{k-(\alpha+1)} \mathrm{e}^{-\frac{c_1 l}{k^\alpha}} \\ &\geq c \frac{c_1^k k^{-\alpha k}}{k!} \int_0^{\infty} x^{k-(\alpha+1)} \mathrm{e}^{-\frac{c_1 x}{k^\alpha}} \mathrm{d}x \\ &\geq c c_1^k k^{-\alpha k} \frac{\Gamma\left(k - (\alpha+1) + 1\right)}{k!} \left(\frac{k^\alpha}{c_1}\right)^{k-(\alpha+1)+1} \\ &= c k^{-\alpha^2} \frac{\Gamma\left(k - \alpha\right)}{k!} \\ &\geq c k^{-\alpha^2} \frac{\left(k-\alpha\right)^{k-\alpha-1/2} \mathrm{e}^{k-\alpha+\Theta/12(k-\alpha)}}{k^{k+1/2} \mathrm{e}^{k+\Theta/12(k)}} \\ &\geq c k^{-\alpha(\alpha+1)-1} \end{aligned}$$

where the Stirling formula has been used for $\Gamma\left(k - \alpha\right)$ and $k!$ as well and $0 < \Theta < 1$. The upper estimate works similarly as follows.

$$\begin{aligned} l_k &= \left(G_k\right)^k \sum_{l=0}^{\infty} \left(k+l\right)^{-(\alpha+1)} \binom{k+l}{l} \left(F_{k+1}\right)^l. \\ l_k &\leq c \frac{c_1^k k^{-\alpha k}}{k!} \sum_{l=0}^{\infty} \left(k+l\right)^{k-(\alpha+1)} \mathrm{e}^{-\frac{c_1 l}{(k+1)^\alpha}} \\ &\leq c \frac{c_1^k k^{-\alpha k}}{k!} \sum_{l=0}^{\infty} \left(k+l\right)^{k-(\alpha+1)} \mathrm{e}^{-\frac{c_1 l}{k^\alpha}\left(\frac{k}{k+1}\right)^\alpha} \\ &= c \frac{c_1^k k^{-\alpha k}}{k!} \sum_{l=k}^{\infty} l^{k-(\alpha+1)} \mathrm{e}^{-\frac{c_1 (l-k)}{k^\alpha}\left(\frac{k}{k+1}\right)^\alpha} \\ &= c \frac{c_1^k k^{-\alpha k}}{k!} \sum_{l=k}^{\infty} l^{k-(\alpha+1)} \mathrm{e}^{-\frac{c_1 l}{k^\alpha}\left(\frac{k}{k+1}\right)^\alpha} \\ &\leq c \frac{c_1^k k^{-\alpha k}}{k!} \sum_{l=0}^{\infty} l^{k-(\alpha+1)} \mathrm{e}^{-\frac{c_1 l}{k^\alpha}\left(\frac{k}{k+1}\right)^\alpha} \\ &\leq c \frac{c_1^k k^{-\alpha k}}{k!} \int_0^{\infty} x^{k-(\alpha+1)} \mathrm{e}^{-\frac{\left(\frac{k}{k+1}\right)^\alpha c_1 x}{k^\alpha}} \mathrm{d}x. \end{aligned}$$

[2] Henceforth $c$ will be an arbitrary positive constant unless specified otherwise. Its value may change from occurrence to occurrence.

Introducing a new variable one obtains

$$
\begin{aligned}
l_k &\le c\frac{c_1^k k^{-\alpha k}}{k!}\int_0^\infty x^{k-(\alpha+1)} e^{-\frac{\left(\frac{k}{k+1}\right)^\alpha c_1 x}{k^\alpha}} dx \\
&= c\frac{c_1^k k^{-\alpha k}}{k!}\left[\frac{\left(\frac{k}{k+1}\right)^\alpha c_1}{k^\alpha}\right]^{-k+\alpha}\int_0^\infty y^{k-(\alpha+1)} e^{-y} dy \\
&= cc_1^k k^{-\alpha k}\frac{\Gamma\,(k-(\alpha+1)+1)}{k!}\left(\frac{k^\alpha}{c_1}\right)^{k-\alpha}\left[\left(\frac{k+1}{k}\right)^\alpha\right]^{k-\alpha} \\
&\le ck^{-\alpha^2}\frac{\Gamma\,(k-\alpha)}{k!}\left(1+\frac{1}{k}\right)^k \\
&\le ck^{-\alpha^2}\frac{(k-\alpha)^{k-\alpha-1/2}\, e^{k-\alpha+\Theta/12(k-\alpha)}}{k^{k+1/2} e^{k+\Theta/12(k)}} \\
&\le ck^{-\alpha(\alpha+1)-1}
\end{aligned}
$$

where at the end the Stirling formulas have been used as in the lower estimate.

## References


[1] J.E. Hirsch, PNAS 102 (2005) 16569–16572.
[2] A.L. Barabasi, R. Albert, H. Jeong, Physica A 272 (1999) 173–187.
[3] A. Schubert, Scientometrics 78 (2009) 375–382.
[4] L.C. Freeman, Social Networks 1 (1979) 215–239.
[5] A. Bavelas, Human Organization 7 (1948) 16–30.
[6] M.E. Shaw, Journal of Psychology 38 (1954) 139–149.
[7] A. Bavelas, D. Barrett, Journal of the Acoustical Society of America 22 (1950) 1271–1282.
[8] G. Sabidussi, Psychomatrika 31 (1966) S81–603.
[9] S.P. Borgatti, Social Networks 27 (2005) 55–71.
[10] S. Brin, L. Page, Computer Networks and ISDN Systems 30 (1998) 107–117.
[11] P. Holme, G. Ghoshal, URL http://www.citebase.org/abstract?id=oai:arXiv.org:0805.3909, 2008.
[12] E. Schwarzkopf, D. Heckmann, D. Dengler, A. Kröner, in: R.S. Baker, J.E. Beck, B. Berendt, A. Kröner, E. Menasalvas, S. Weibelzahl (Eds.), Complete On-Line Proceedings of the UM 2007 Workshop on Data Mining for User Modeling DMUM07, Corfu, Greece, 2007, pp. 63–75.
[13] M. Bloem, T. Alpcan, T. Başar, in: Proc. 46th IEEE Conference on Decision and Control, New Orleans, LA, 2007.
[14] C. von der Weth, K. Bhm, in: cec-eee, 2007, pp. 270–277.
[15] E. Yoneki, P. Hui, S. Chan, J. Crowcroft, MSWiM (2007) 225–234.
[16] S. Schmidt, T. Alpcan, S. Albayrak, T. Başar, A. Muller, in: 2nd International Workshop on Critical Information Infrastructures Security, CRITIS'07, Malaga, Spain, 2007.
[17] A.H. Dekker, B. Colbert, Complexity International 12 (2008).
[18] T. Nepusz, A. Petrczi, L. Ngyessy, F. Bazs, Physical Review E 77 (2008).
[19] K. Barcza, A. Telcs, Paretian publication patterns imply paretian. Hirsch index, Scientometrics (in press).
[20] W. Glänzel, Scientometrics 67 (2006) 315–321.
[21] P.L. Krapivsky, S. Redner, F. Leyvraz, Physical Review Letters 85 (2000) 4629–4632.
[22] A. Clauset, C.R. Shalizi, M.E.J. Newman, 2007. Preprint 0706.1062.
[23] S. Zhou, R.J. Mondragón, Physical Review E 70 (2004) 066108.
[24] URL http://www.caida.org/data/active/skitter_aslinks_dataset.xml.
[25] D. Koschützki, F. Schreiber, in: German Conference on Bioinformatics, 2004, pp. 199–206.
[26] C.Y. Lee, Correlations among centrality measures in complex networks, http://www.citebase.org/abstract?id=oai:arxiv.org:physics/0605220, 2006.
[27] B.H. Junker, D. Koschützki, F. Schreiber, BMC Bioinformatics 7 (2006) 219.