

Credit allocation for research institutes

J.-P. WANG¹, Q. GUO¹, K. YANG¹, J.-T. HAN² and J.-G. LIU^{1,2,3(a)}

¹ *Research Center of Complex Systems Science, University of Shanghai for Science and Technology Shanghai 200093, PRC*

² *Data Science and Cloud Service Centre, Shanghai University of Finance and Economics Shanghai 200433, PRC*

³ *Department of Physics, University of Fribourg - CH-1700 Fribourg, Switzerland*

PACS 89.75.Fb – Structures and organization in complex systems

PACS 05.10.-a – Computational methods in statistical physics and nonlinear dynamics

PACS 02.20.-a – Group theory

Abstract – It is a challenging work to assess research performance of multiple institutes. Considering that it is unfair to average the credit to the institutes which is in the different order from a paper, in this paper, we present a credit allocation method (CAM) with a weighted order coefficient for multiple institutes. The results for the APS dataset with 18987 institutes show that top-ranked institutes obtained by the CAM method correspond to well-known universities or research labs with high reputation in physics. Moreover, we evaluate the performance of the CAM method when citation links are added or rewired randomly quantified by the Kendall's Tau and Jaccard index. The experimental results indicate that the CAM method has better performance in robustness compared with the total number of citations (*TC*) method and Shen's method. Finally, we give the first 20 Chinese universities in physics obtained by the CAM method. However, this method is valid for any other branch of sciences, not just for physics. The proposed method also provides universities and policy makers an effective tool to quantify and balance the academic performance of university.

Introduction. – Credit allocation of multiple institutes is a challenging task owing to the universality of collaboration [1–6]. Figure 1 shows that the percentage of research papers published by collaborating institutes has increased dramatically since 1960, suggesting that cooperation among institutes has become more and more frequent. Assessment of scientific impact is significant for researchers and research institutes, especially in the process of recruitment and promotion, award of grants, offering tenure, granting, etc. [7–9].

The significance of credit allocation triggers a variety of approaches [10–13], including the impact factor of journals [14]; the total number of citations [15]; the total number of papers [16]; the *h*-index reflects on publications and citations [17] and the *g*-index is an alternative for the *h*-index, based on the distribution of citations [18], etc. But, because of the inaccuracy of citation data, Moed [19] found that on average, the number of non-matching references was about 7% of citation matched, and the citation index depended on the role of journal articles in the

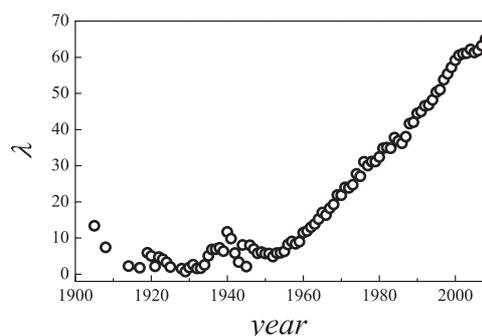


Fig. 1: λ denotes the percentage of research papers published by co-institutes. This plot exhibits the change of λ as the publishing year. The result suggests that cooperation among institutes is more and more frequent.

different fields. Also, these metrics are just for individual's research output, while the scientific impact is determined by the scientific community [20]. Van Raan [21] pointed out that ranking of research institutes by these methods was not reasonable.

^(a)E-mail: liujg004@ustc.edu.cn

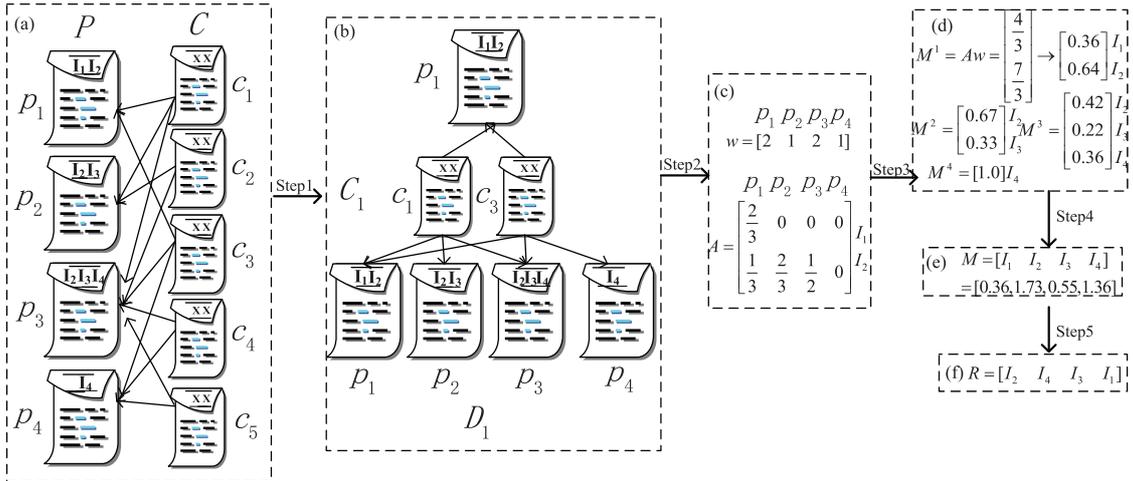


Fig. 2: (Color online) Schematic illustration of the CAM method. (a) The original citation network, cited paper set $P = \{p_1, p_2, p_3, p_4\}$, citing paper set $C = \{c_1, c_2, c_3, c_4, c_5\}$. (b) Step 1, randomly selected a paper in P , such as the paper p_1 with two institutes I_1 and I_2 . After its citing paper set $C_1 = \{c_1, c_3\}$, and paper set $D_1 = \{p_1, p_2, p_3, p_4\}$ are established. (c) Step 2, calculating the citation count in which each paper in D_1 is cited by papers in C_1 , the weight vector $w = [2, 1, 2, 1]$ is obtained. Using the function of credit allocation $a_{I_i, h} = 2 * (k - r + 1) / ((k + 1) * k)$ to I_i institute from each paper in D_1 , e.g., $a_{I_1, 1} = 2 * (2 - 1 + 1) / ((2 + 1) * 2) = 2/3$, $a_{I_1, 2} = 0$, $a_{I_1, 3} = 0$, $a_{I_1, 4} = 0$. Similarly, $a_{I_2, 1} = 1/3$, $a_{I_2, 2} = 2/3$, $a_{I_2, 3} = 1/2$, $a_{I_2, 4} = 0$. (d) Using the $m_{I_i, j} = \sum_h a_{I_i, h} w_{j, h}$ to each paper in D_1 , then constructed by normalizing $m_{I_i, j}$ by column, e.g., $m_{I_1, 1} = 4/3 / (4/3 + 7/3) \rightarrow 0.36$, $m_{I_2, 1} = 0.64$. (e) Doing a similar calculation for paper p_2, p_3, p_4 , then cumulating the score of $m_{I_i, j}$ with the same institute, we obtained $M = [0.36, 1.71, 0.55, 1.33]$ corresponding to I_1, I_2, I_3, I_4 , respectively. (f) Finally, we acquired the institute ranking R by descending order, i.e., $[I_2, I_4, I_3, I_1]$.

Graph-ranking methods [22,23] are often applied to solve the collective problem [24–27]. Wang *et al.* [28] found that the paper level graph-ranking methods were more reasonable for allocation of scientific credit. Shen *et al.* [29] proposed an average credit allocation method based on citation relationship for multi-authors. However, in a paper with multiple authors, their contributions are hardly equal and not all of them should get full credit, resulting in the problem of citation bubble [30].

Inspired by the above problems, a relatively fair credit allocation method is necessary. In this paper, we present a credit allocation method (CAM) to allocate the credit of multiple research institutes based on a weighed order coefficient related by the order of the institutes in a paper. The process firstly calculates the citation count of each paper, then allocates the credit based on the order of each institute in the paper, i.e., the more top institute order is, the more credit it gets, which is different from the method by Zhang [30] in which the first author and the last author share the same credit. By analyzing the academic performance of the world university using the APS data set, the results show that the top-ranked institutes obtained by the CAM method correspond to well-known universities or research labs with long history and high reputation in physics. We test on the real data set with artificial spammers, including random adding edges and random rewiring edges [31], the experimental results indicate that the CAM method has higher performance in robustness compared with the number of citations (*TC*) method and Shen’s method (the one proposed in [29], noted by Shen’s method).

The remainder of this paper is organized as follows. The second section presents the model and methods. The third section describes the data, the versatility of the CAM method in empirical network and tolerance of noisy data. The last section discusses the results and suggests some potential directions for future research.

Model and methods. – The citation relationships can be described by a directed network, where the cited papers are denoted by set $P = \{p_1, p_2, \dots, p_m\}$, the citing papers are denoted by set $C = \{c_1, c_2, \dots, c_n\}$, and the links are formed by two sets, the direction of which is from citing papers to cited papers, the links are indicated by $E = \{e_1, e_2, \dots, e_l\}$. The institutes set $I = \{I_1, I_2, \dots, I_g\}$.

Credit allocation method (CAM). The CAM method can be calculated as follows (processes as shown in fig. 2):

- i) We randomly select a paper p_j , which is produced by a institutes, including the I_i institute, and form the paper set $C_j = \{c_1, c_2, \dots, c_b\}$ citing p_j . Next we identify paper set $D_j = \{p_1, \dots, p_c\}$, every paper in this set has at least one common institute with paper p_j and cited by the paper set C_j .
- ii) We calculate the citation count $w_{j, h}$ of the h -th paper in D_j , defined as the number of times that the h -th paper in D_j is cited by papers in C_j . Then calculate a credit sharing matrix A , whose element $a_{I_i, h}$ denotes the credit amount of I_i institute from the h -th paper in D_j . Assuming that I_i is at the r -th position in the h -th paper with k institutes, $a_{I_i, h} = 2 * (k - r + 1) / ((k + 1) * k)$.

Table 1: Basic statistics of dataset.

Journal	Papers	Citations	Period
<i>Physical Review (Series I)</i>	292	668	1893–1912
<i>Physical Review</i>	44084	590665	1913–1969
<i>Physical Review A</i>	51945	418196	1970–2009
<i>Physical Review B</i>	134530	1191515	1970–2009
<i>Physical Review C</i>	29081	202312	1970–2009
<i>Physical Review D</i>	54641	526930	1970–2009
<i>Physical Review E</i>	34703	1507974	1993–2009
<i>Physical Review Letters</i>	90376	1507974	1958–2009
<i>Review of Modern Physics</i>	2486	115697	1929–2009
<i>Physical Review Special Topics - Accelerators and Beams</i>	1078	2457	1998–2009
<i>Physical Review Special Topics CPhysics Education Research</i>	1	0	2005–2009
Total	443217	4710547	1893–2009

- iii) $m_{I_i,j}$, the credit of institute I_i is obtained from paper p_j [29]:

$$m_{I_i,j} = \sum_h a_{I_i,h} w_{j,h}, \quad (1)$$

then normalized to $m_{I_i,j}$.

- iv) For each paper in the network, we execute from step i) to step iii), then cumulate the $m_{I_i,j}$ with the I_i institute, *i.e.*, $m_{I_i} = \sum_{j=1}^m m_{I_i,j}$, finally the credit of all institutes \mathbf{M} is established.
- v) By descending order for \mathbf{M} , we get the institute ranking \mathbf{R} .

We validate our method by comparing with Shen’s method and the number of citation (TC) method index which is commonly as used for research evaluation [32,33].

The total number of citation (TC). For an institute, TC_i denotes the total citation of the institute I_i [34]:

$$TC_i = \sum_{j=1}^{|\Phi_i|} \Psi(q_j(i)), \quad (2)$$

where Φ_i denotes the paper set which contains I_i , $q_j(i)$ denotes the j -th paper with institute i in Φ , $\Psi(q_j(i))$ denotes the number of papers citing paper $q_j(i)$, $|\cdot|$ denotes the size of the set.

Measurements. We put in the artificial disturbance by randomly adding citation and randomly rewiring citation. Kendall’s Tau [35,36] and Jaccard index [37] are introduced in this paper to measure the performance. The higher Kendall’s Tau index is, the better robustness the corresponding algorithm has. We also discuss how robustness is influenced by the different length L of the ranking list, then utilize the Jaccard index to evaluate it. A brief introduction of each measure is shown.

1) *Kendall’s Tau.* For an institute I_i , we denote x_i as the institute i ranking position before the artificial disturbance

is added and y_i as its ranking position after the artificial disturbance is added:

$$\tau = \frac{2}{N(N-1)} \sum_{i < j} \text{sgn}[(x_i - x_j)(y_i - y_j)], \quad (3)$$

where N is the total number of institutes in the citation relationships, $\text{sgn}(x)$ is a piecewise function such that when $x > 0$, $\text{sgn}(x) = +1$; $x < 0$, $\text{sgn}(x) = -1$; when $x = 0$, $\text{sgn}(x) = 0$. τ measures the correlation between two ranking lists, whose value is in the range $[-1, 1]$ and a larger τ corresponds to a better performance.

2) *Jaccard index.* We denote U_L and V_L as two sorted lists of credit score in the same length L of lists before and after the artificial disturbance is added, $S \in (0, 1)$. A larger S corresponds to a better performance:

$$S = \frac{|\Gamma(U_L) \cap \Gamma(V_L)|}{|\Gamma(U_L) \cup \Gamma(V_L)|}. \quad (4)$$

Experimental results. –

Data description. The American Physical Society (APS) dataset consist of all papers published by journals of the American Physical Society between 1893 and 2009. For each paper the dataset includes DOI, title, date of publication (day, month, year), author names and affiliations of each author. To investigate the scientific allocation problem of multi-institutes in a paper, we select papers with more than one institute, divide institutional level into the university and company level by Wildcard and truncation. Besides, the retrieved records with confusing names (*e.g.*, same abbreviations, different institutes) are manually inspected. This filter gives us a final set of 18987 institutes, 443217 papers, and 4710547 citations (table 1).

Versatility of the CAM method in empirical networks. The ranking list in physics obtained by the CAM method, Shen’s method and (TC) method is shown in table 2.

Table 2: The university ranking according to three algorithms in partial result.

Institute	CAM	Shen's method	(<i>TC</i>) method
Univ. of California	1	1(+0)	1(+0)
Bell Lab.	2	2(+0)	11(+9)
Max Planck Institute	3	3(+0)	2(-1)
MIT	4	4(+0)	5(+1)
Univ. of Chicago	5	5(+0)	4(-1)
Univ. of Illinois	6	6(+0)	6(+0)
Stanford Univ.	7	8(+1)	8(+1)
Brook. Nat. Lab.	8	7(-1)	7(-1)
Univ. of Tokyo	9	9(+0)	3(-6)
Harvard Univ.	10	10(+0)	12(+2)
Cornell Univ.	11	11(+0)	19(+8)
Princeton Univ.	12	12(+0)	15(+3)
Russian Academy of Sciences (RAS)	19	37(+18)	48(+29)
Yale Univ.	20	20(+0)	27(+7)
Columbia Univ.	22	23(+1)	47(+25)
Istituto Nazionale di Fisica Nucleare (INFN)	52	40(-12)	10(-42)
Brown Univ.	59	62(+3)	91(+32)

Top-ranked institutes by the CAM method correspond to well-known universities or research labs with long history of excellence in physics. Note that Bell Lab, whose scientists and engineers have made seminal scientific discoveries, won eight Nobel Prizes. Harvard University, Yale University, Princeton University, Columbia University, Brown University and Cornell University are the ivy league colleges, which gain their higher positions in our CAM algorithm. The (*TC*) method reflects the impact of a scientific paper, and the total citation of publication are equal to each institute in a paper/ institutes. This measurement does not differentiate the importance of the cited papers: a citation coming from an obscure paper has the same weight as one from a ground breaking, highly cited work [38]. For Shen's method, the rule of average allocation makes the authors who make more contribution suffer a loss of credit, while exaggerating the credit of authors who make less contribution. An obvious difference among CAM method, Shen's method and (*TC*) method lies in the formulation, where the citation relationship and the weighted order coefficient play important roles in regulating credit flows, making the CAM method more reasonable.

Tolerance of noisy data. Citation network data may be unreliable. Evans *et al.* [39] found that according to empirical studies, there were 31% of the papers cited incorrectly, and even 10% of the citations were wrong. Todd *et al.* [40] found that one in four citations in marine biology papers was inappropriate. Haussmann [41] found that at least 19% of the citations in physical geography did not provide clear support for the statements they were meant to support. Liang *et al.* [42] found that scientists had referencing misbehavior and behavior of self-copying of references. Therefore, the tolerance of the algorithm against

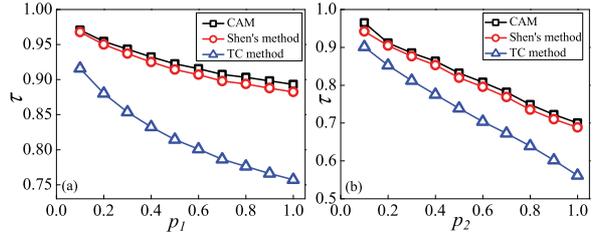


Fig. 3: (Color online) Kendall's Tau τ vs. adding link rate p_1 , rewiring link rate p_2 for the CAM method, Shen's method and (*TC*) method. Panel (a) exhibits Kendall's Tau τ vs. p_1 , and panel (b) reports Kendall's Tau τ vs. p_2 . At the tolerance of noisy data cases, Kendall's Tau τ of the CAM method in adding link rate p_1 and rewiring link rate p_2 are much better than the ones of Shen's method and the (*TC*) method. The p_1 and p_2 varies from 0.1 to 1.0, respectively.

spurious links is an appropriate metric [43]. In this letter, we measure the performance of effectiveness and robustness among the CAM method, Shen's method and (*TC*) method in rankings by Kendall's Tau, τ and the Jaccard index S when citation links are added or rewired randomly, which is very common in real citation networks [31,44].

As shown in fig. 3, Kendall's Tau τ for the CAM method is between 0.893 and 0.970 for $p_1 \in [0.1, 1]$ and between 0.700 and 0.965 for $p_2 \in [0.1, 1]$, indicating that the ranking lists generated by the CAM method before and after the artificial disturbance is added are highly identical to each other. While Kendall's Tau τ for the (*TC*) method is between 0.757 and 0.92 for $p_1 \in [0.1, 1]$ and between 0.561 and 0.901 for $p_2 \in [0.1, 1]$. And for Shen's method, Kendall's Tau τ is between 0.882 and 0.967 for $p_1 \in [0.1, 1]$ and between 0.688 and 0.942 for $p_2 \in [0.1, 1]$. In comparison, the CAM method performs better than the

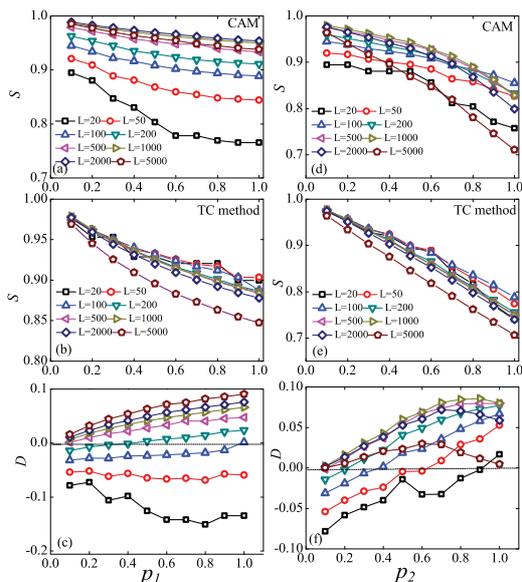


Fig. 4: (Color online) The Jaccard index S vs. adding link rate p_1 , rewiring link rate p_2 and different top L for the CAM method and (TC) method. D denotes the different of S_L between CAM minus (TC) method. Panels (a)–(c) exhibit Jaccard index S , D vs. adding link rate p_1 and different top L , and panels (d)–(f) report the Jaccard index S , D vs. rewiring link rate p_2 and different top L . The p_1 and p_2 varies from 0.1 to 1.0, respectively. L ranges from 20 to 5000.

(TC) method and Shen’s method since the CAM method takes into account the detailed citation pattern and the order of the institutes in a paper. As in fig. 4, the Jaccard index S vs. adding link rate p_1 , rewiring link rate p_2 and different top L for the CAM method and (TC) method. Figures 4(a) and (d) show that with the CAM method, the smaller p_1 and the larger L , the higher the value of Jaccard index S . Moreover, figs. 4(b) and (e) show that with the (TC) method, the smaller p_1 and the smaller L , the higher the value of Jaccard index S . In the plot of (c) and (f), the parts of the D greater than 0 suggest that the robustness of the CAM method is better than that of the (TC) method.

Table 3 shows the top 20 Chinese universities by the CAM method in 2009. According to the CAM method, the University of Science and Technology of China, Nanjing University, Peking University, Tsinghua University have ranking positions 1, 2, 3, 4, respectively, in the physics field.

Conclusions and discussions. – In this letter, we presented a credit allocation algorithm for multiple institutes, namely, credit allocation method (CAM). We assumed that the information about the scientific credit allocation is encoded in the citation pattern of papers. And the credit allocation of each institute is determined by the order of the institute in a paper. So we proposed a weight order coefficient according to which the higher the institute order was, the more credit it got. We use

Table 3: List of the first 20 universities according to the CAM method among the Chinese universities in physics.

Institute	Ranking
University of Science and Technology of China	1
Nanjing University	2
Peking University	3
Tsinghua University	4
Fudan University	5
Beijing Normal University	6
Zhejiang University	7
Shanghai Jiaotong University	8
Shandong University	9
Huazhong University of Science and Technology	10
Nankai University	11
Jilin University	12
Wuhan University	13
Xi’an Jiaotong University	14
Southeast University	15
Sichuan University	16
Sun Yat Sen University	17
Harbin Institute of Technology	18
Tianjin University	19
Central South University	20

the APS dataset to examine the institute ranking. Result shows that the top ranking institutes obtained by the CAM method correspond to well-known universities or research labs with high reputation in physics.

To validate our method, we further investigate that tolerance of ranking against spurious citation links, *i.e.*, random adding links and random rewiring links affect the robustness of the institute academic performance compared with Shen’s method and the number of citations (TC) method, applying Kendall’s Tau τ and Jaccard index S . The results show that Kendall’s Tau τ of the CAM method is over 0.9, 0.7, in the cases of randomly adding edge and randomly rewiring edge, respectively, while Kendall’s Tau τ of the CAM method is over 0.75, 0.56 in the cases of randomly adding edge and randomly rewiring edge, respectively. In comparison, the CAM method has better tolerance of ranking against spurious links. We also discuss how to affect the performance about the top L ranking using the Jaccard index S . The results show that either larger L or smaller proportion of adding edge results in a better robustness of the CAM method. Due to the limitation of the APS dataset, our results are emphasized in physics. However, our method is valid in any other branch of sciences, not just for physics.

In spite of our making some attempts and achieving good results, there are still shortages. For instance, the number of citations collected by a paper strongly depends on the paper’s age [45], and we did not consider this factor

in our method. And institutes with long history may easily get more citations, researchers could consider the institute's date, etc.

* * *

We acknowledge the American Physical Society Research Group for providing us the American Physical Society data. We thank JIAN-HONG LIN, MENG-JIE LI, LU DAI, XIAO-LU LIU and all others at Research Center of Complex Systems Science. This work is partially supported by the National Natural Science Foundation of China (Grant Nos. 71171136, 61374177, 71371125), Research supported by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, Supported by Shuguang Program Project of Shanghai Educational Committee (Grant No. 14SG42). J-GL is supported by the Sino Swiss Science and Technology Cooperation (No. 09-032016).

REFERENCES

- [1] JONES B. F., WUCHTY S. and UZZI B., *Science*, **322** (2008) 1259.
- [2] KATZ J. S. and HICKS D., *Scientometrics*, **40** (1997) 541.
- [3] PRESSER S., *Collab. Quality Res.*, **10** (1980) 95.
- [4] BEAVER D. B., *Scientometrics*, **60** (2004) 399.
- [5] KENNEDY D., *Science*, **301** (2003) 733.
- [6] ALLEN L., BRAND A., SCOTT J., ALTMAN M. and HLAVA M., *EPL*, **81** (2008) 580004.
- [7] VAN YPERSELE J. P., *Scientometrics*, **85** (2010) 243.
- [8] AGUILLO I. F., BAR-ILAN J., LEVENE M. and ORTEGA J. L., *Scientometrics*, **85** (2010) 243.
- [9] REN Z. M., ZENG A., CHEN D. B., LIAO H. and LIU J. G., *EPL*, **106** (2014) 48005.
- [10] WU Q., *J. Am. Soc. Inf. Sci. Technol.*, **61** (2010) 609.
- [11] VINKLER P., *J. Inf. Sci.*, **35** (2009) 602.
- [12] KREIMAN G., *Front. Comput. Neurosci.*, **5** (2011) 48.
- [13] SINATRA R., WANG D., DEVILLE P., SONG C. M. and BARABÁSI A. L., *Science*, **354** (2016) aaf5239.
- [14] GARFIELD E., *Science*, **178** (1972) 471.
- [15] PETERSEN A. M., WANG F. and STANLEY H. E., *Phys. Rev. E*, **81** (2010) 036114.
- [16] NEWMAN M. E. J., *Proc. Natl. Acad. Sci. U.S.A.*, **101** (2004) 5200.
- [17] HIRSCH J. E., *Proc. Natl. Acad. Sci. U.S.A.*, **102** (2005) 16569.
- [18] EGGHE L., *Scientometrics*, **69** (2006) 131.
- [19] MOED H. F., *Nature*, **415** (2002) 731.
- [20] NEYLON C., WU S., *PLoS Biol.*, **7** (2009) e1000242.
- [21] VAN RAAN A. F. J., *Scientometrics*, **62** (2005) 133.
- [22] CHEN P., XIE H., MASLOV S. and REDNER S., *J. Inf.*, **1** (2007) 8.
- [23] KLOSIK D. F. and BORNHOLDT S., *PLoS ONE*, **9** (2014) e113184.
- [24] HOU L., LIU J. G., PAN X., SONG W. J. and LI X. D., *Int. J. Mod. Phys. C*, **25** (2014) 1450002.
- [25] WANG J. P., GUO Q., YANG G. Y. and LIU J. G., *Physica A*, **428** (2015) 250.
- [26] LIU X. L., GUO Q., HOU L., CHENG C. and LIU J. G., *Physica A*, **436** (2015) 629.
- [27] LI D., GUO Q., HAN J. T. and LIU J. G., *Phys. Lett. A*, **380** (2016) 3460.
- [28] WANG H., SHEN H. W. and CHENG X. Q., *Scientometrics*, **109** (2016) 827.
- [29] SHEN H. W., BARABÁSI A. L., *Proc. Natl. Acad. Sci. U.S.A.*, **111** (2014) 12325.
- [30] ZHANG C. T., *EMBO Rep.*, **10** (2009) 416.
- [31] LÜ L., ZHANG Y. C., YEUNG C. H. and ZHOU T., *PLoS ONE*, **6** (2011) e21202.
- [32] GLÄNZEL W., SCHUBERT A. and CZERWON H. J., *Scientometrics*, **45** (1999) 185.
- [33] LIN C. S., HUANG M. H. and CHEN D. Z., *J. Inf.*, **7** (2013) 611.
- [34] GARFIELD E., *Essays Inf. Sci.*, **1** (1973) 406.
- [35] KENDALL M. G., *Biometrika*, **30** (1938) 81.
- [36] LIU J. G., LIN J. H., GUO Q. and ZHOU T., *Sci. Rep.*, **6** (2016) 21380.
- [37] JACCARD P., *Bull. Soc. Vaud. Sci. Nat.*, **37** (1901) 547.
- [38] MASLOV S. and REDNER S., *J. Neurosci.*, **28** (2008) 11103.
- [39] EVANS J. R., NADJARI H. I. and BURCHELL S. A., *JAMA*, **263** (1990) 1353.
- [40] TODD P. A., GUEST J. R. and BURCHELL S. A., *Mar. Ecol. Prog. Ser.*, **408** (2010) 299.
- [41] HAUSSMANN N. S., MCINTYRE T. and BUMBY A. J., *Prog. Phys. Geogr.*, **37** (2013) 543.
- [42] LIANG L., ZHONG Z. and ROUSSEAU R., *Scientometrics*, **101** (2014) 1973.
- [43] GUIMERÁ R. and SALES-PARDO M., *Proc. Natl. Acad. Sci. U.S.A.*, **106** (2009) 22073.
- [44] LARSEN P. and VON INS M., *Scientometrics*, **84** (2010) 575.
- [45] PRICE D. J., *Science*, **149** (1965) 510.