鲁索与中国科学计量学的发展

Ronald Rousseau and the Development of Scientometrics in China

第八届科学计量学与大学评价国际研讨会论文集

Proceedings of the 8th International Conference on Scientometrics and University Evaluation

金伟良 张晓林◎主编

科学出版社

索与国科学计量学的发展

Ronald Rousseau and the Development of Scientometrics in China

第八届科学计量学与大学评价国际研讨会论文集

Proceedings of the 8th International Conference on Scientometrics and University Evaluation

金伟良 张晓林◎主编

科学出版社
北京

图书在版编目(CIP)数据

鲁索与中国科学计量学的发展：第八届科学计量学与大学评价国际研讨会论文集/金伟良，张晓林主编．—北京：科学出版社，2014.12

ISBN 978-7-03-042761-8

Ⅰ.①鲁… Ⅱ.①金…②张… Ⅲ.①科学计量学－国际学术会议－文集 Ⅳ.①G301-53

中国版本图书馆 CIP 数据核字（2014）第 291880 号

责任编辑：朱丽娜 张翠霞/责任校对：张凤琴
责任印制：赵 博/整体设计：楠竹文化
编辑部电话：010-64033934
E-mail：fuyan@mail. sciencep. com

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www. sciencep. com
骏杰印刷有限公司 印刷
科学出版社发行 各地新华书店经销
*
2015 年 1 月第 一 版 开本：720×1000 1/16
2015 年 1 月第一次印刷 印张：17 3/4
字数：340 000
定价：78.00 元
（如有印装质量问题，我社负责调换）

本书编委会

主　编　金伟良　张晓林

副主编　杨立英　朱苏永

编委会（按姓氏汉语拼音排序）

丁洁兰　金伟良　林承亮　马建华
史双青　汤建民　徐晓艺　杨国梁
杨立英　岳　婷　张国昌　张晓林
诸葛洋　朱苏永

学术顾问　武夷山　Ronald Rousseau　蒋国华

主办　中国科学学与科技政策研究会科学计量学与信息计量学专业委员会

承办　浙江大学宁波理工学院
中国科学院文献情报中心

赞助　爱思唯尔出版集团
汤森路透科技集团

序　言

武夷山　杨立英

自20世纪90年代以来，我国科学计量学研究与科研量化评价发展迅速，利用定量分析方法开展大学评估工作的国际影响也在不断扩大，这要归功于中国科学计量学研究工作者的艰辛努力和极富成效的国际交流。在与我国科学计量学界密切交往的诸多国际同仁之中，有一位非常杰出的友好人士，他就是国际著名科学计量学家鲁索教授（Prof. Dr. Ronald Rousseau，现任国际科学计量学与信息计量学学会会长，比利时鲁汶大学教授，科学计量学领域的最高荣誉——普赖斯奖获得者）。

鲁索教授自1998年首次访华以来的16年间，为促进中外科学计量学界的学术交流、培养中国科学计量学研究力量作出了杰出贡献。他到我国各地的学术访问次数近30次；他亦分别被中国科学院文献情报中心、河南师范大学、大连理工大学、浙江大学、上海大学、南京农业大学、武汉大学等聘为客座教授或荣誉教授；截至2013年10月他与国内众多科学计量学者合著发表的论文已达62篇。为表达中国同行对鲁索教授的敬意，感谢他不遗余力为推动中国科学计量学研究与应用所作出的卓越贡献，2013年年初，经国内诸多同行郑重提议，中国科学学与科技政策研究会科学计量学与信息计量学专业委员会决定于2013年11月7～8日召集“第八届科学计量学与大学评价国际研讨会暨鲁索教授与中国学者合作研讨会”。

研讨会在浙江省宁波市宁波理工学院举行。会议吸引了全国各地及比利时、荷兰、瑞典、英国、西班牙等国家的95名中外专家与会，其中有三位外国专家是国际科学计量学界的学术最高奖——普赖斯奖获得者，鲁索教授本人亲自到会并发言。研讨会紧密围绕国内外科学计量学研究的最新进展、大学评估与科研量化评价、中国科学计量学研究的发展历程（特别是中国学者与鲁索教授的合作经历）等主题进行了学术交流与研讨。

鲁索教授报告了他关于测度网络节点的桥梁作用的最新成果——Q指标；国际科学计量学与信息计量学学会秘书长、普赖斯奖得主、比利时鲁汶大学Wolfgang Glänzel教授作了《关于应用高级的文献计量学指标和引文表现等级对大学及研究机构进行评价研究》的报告，报告中提出了一个基于“引文表现等级”的新的评价方法，以便为不同层级的评价研究提供一个“无预设参数”；普赖斯奖得主、荷兰阿姆斯特丹大学Loet Leydesdorff教授做了《基于三螺旋模型的创新系统协同效应的测度

指标》的报告，介绍了他与很多学者（包括中国学者）合作开展的大量实证研究。

本次会议最突出的特色之一是纪念鲁索教授初次访华暨与中国学者合作 15 周年。与鲁索教授有密切合作关系的众多中国学者在这一主题的发言中抒发了感谢、感激、感慨之情。与已故赵红州教授共同开创了中国科学计量学研究先河的蒋国华研究员作了《中国科学计量学界的伟大朋友》的演讲；金碧辉研究员在《我眼中的鲁索教授》报告中回顾了两人长达 15 年的学术合作历程；梁立明教授以 Web of Science 库为数据来源，分析了鲁索教授的学术研究“节律”；叶鹰教授用文献计量学研究方法分析了鲁索教授和国内学者合作发表论文的被引特征；清华同方中国知网万锦堃教授在《鲁索教授与清华》报告中，通过大量的图片展现了鲁索教授对中国知网文献计量学研究的支持和帮助；汤森路透科技集团中国部科学计量学研究顾问岳卫平博士也通过大量珍贵的照片生动地回顾了她和鲁索教授 7 次会面时的情景。鲁索教授在其每份电子邮件的最后都附有一句话“In scientific affairs one can never be too generous”（在科学事务中不存在过分慷慨的问题）。他本人就实践着这一理念，不断给科学计量学界同行提供慷慨的咨询与建议。我们确信，正如中国医学界不会忘记白求恩一样，中国科学计量学界将永远铭记鲁索教授对我们的慷慨帮助。

研讨会为中国学者提供了与国际高水平同行进行学术交流的平台，充分表达了中国科学计量学界对鲁索教授的诚挚感谢。会议的成功举办离不开许多单位与个人的鼎力支持。在此，我们隆重感谢浙江大学宁波理工学院、中国科学院文献情报中心承办会议；感谢爱思唯尔出版集团、汤森路透公司对会议的赞助；感谢浙江省第十届政协副主席盛昌黎女士到会并致辞；感谢宁波大学理工学院金伟良院长为会议提供的诸项便利条件和周到服务；感谢中国科学院文献情报中心张晓林馆长百忙中赴会并发言；感谢中国科学计量学研究前辈蒋国华教授在会议筹办、组织过程中的特殊贡献；感谢国家行政学院史朝教授为会议承办所作的大量协调工作。

在同行评议的基础上，编委会遴选出较优秀的会议论文 22 篇结集正式出版。这些论文在一定程度上反映了当前科学计量学领域主要的研究问题，我们希望论文集的出版能够为中国科学计量学者的研究工作提供借鉴和参考。

国际科学计量学与情报计量学学会是 1994 年正式成立的，迄今已经 20 年了。20 岁，是青春勃发的年龄。我们有幸成为科学计量学研究事业的一员，光荣地与国际科学计量学与情报计量学学会一起成长。我们洒下了辛勤的汗水，也体会到了摘取研究果实的愉悦。我们热望更多的年轻人加入这支队伍，让“我们的队伍越走越长”。

CONTENTS | 目录

三、科学计量指标研究

四、科学计量与大学评估

五、科学计量学在中国

一、科学计量理论与方法研究

1-1 Gauging the Bridging Function of Nodes in a Network: The Gefura Measure

Ronald Rousseau[1], **Raf Guns**[2], **Liu Yuxian**[3]

Abstract

In this article the authors review the study of so-called Q-measures in networks. Q-measures are related to betweenness centrality but are defined in case the network is subdivided into subgroups. They are indicators gauging the bridging, brokerage or gatekeeper function of a node. Two definitions, a basic and a structural Q-measure, are proposed in the case that there are more than two subnetworks and the difference between these two approaches is illustrated. As the term Q-measure is not optimal (there exists other Qs) we propose the term gefura measure (meaning bridge measure) and the corresponding symbol Γ as a better term for this network measure.

Keywords: centrality measures; bridging function; subnetworks; gefura measure

1 Networks

It is stating obviously when pointing out that graphs or networks (these two words will be considered as synonyms) are everywhere. Indeed, road maps represent the network of cities and highways, and similarly we have other transportation networks such as the worldwide air transport network (Barrat et al., 2004; Guimerà

① Institute for Education and Information Sciences, IBW, University of Antwerp, Venusstraat 35, Antwerp, B-2000 (Belgium) & KU Leuven, Leuven B-3000 (Belgium). ronald. rousseau@ua. ac. be.

② Institute for Education and Information Sciences, IBW, University of Antwerp, Venusstraat 35, Antwerp, B-2000 (Belgium). raf. guns@ua. ac. be.

③ Library of Tongji University, Tongji University, Siping Street 1239, Shanghai, 200092 (China). yxliu@lib. tongji. edu. cn.

et al. , 2005) or the shipping and harbors network. Nowadays the Internet is probably the best known network. This network consists actually of a worldwide assemblage of computer networks consisting of local, regional and global academic, business, government, private and public subnetworks (Wikipedia, 2013). Since decades sociologists study friendship and other social networks (Wasserman and Faust, 1994; Otte and Rousseau, 2002), while nowadays the study of gene and disease networks is a hot topic. Indeed, following Goh et al. (2007)—cited already more than 700 times in the Web of Science—medical researchers came to the conclusion that a disease is rarely the consequence of problems with one single gene. Only a study of the intercellular network can lead to advances in the study and ultimately identifying drug targets and biomarkers for complex diseases (Barabási et al. , 2011).

Within the information sciences colleagues study article citation graphs, journal citation graphs, collaboration networks, co-word networks, author co-citation networks and many other co-occurrence networks. Maps of science are constructed based on the complete Web of Knowledge or Scopus. Moreover, a whole new subfield related to the Internet, namely webometrics, has emerged within the field of informetrics.

Generally four types of networks may be distinguished: undirected unweighted networks, directed unweighted networks, undirected weighted networks and directed weighted networks. Directed networks may further be subdivided into acyclic and general directed networks. Examples within the information sciences are given in Table 1.

Table 1 Examples of different types of networks in informetrics

Types	Undirected	Directed
Unweighted	Author collaboration network, not taking frequencies into account	Article citation network (acyclic, except special cases); journal citation network, not taking frequencies into account
Weighted	Author collaboration network, including the frequency of collaboration	Article citation network, including weights related to frequency or importance of use (acyclic); general journal citation network, including citation frequencies

Networks can be characterized by several different measures. These can be subdivided into two main groups: measures related to the network as a whole, and measures related to nodes or links (separately). The simplest global measure is probably the number of nodes. Density, another global measure, is an indicator for the general level of connectedness of the graph. If every node is directly connected to every other node, we have a complete graph. The density of a graph is defined as the number of links (or arcs) divided by the number of links in a complete graph with the same number of nodes. This indicator as well as those that follow will be defined only for simple, i. e. , undirected unweighted networks, unless stated otherwise.

The next group of network measures, namely centrality measures, consists of node indicators. The most important ones are: degree centrality, closeness centrality, betweenness centrality and eigenvector centrality. Degree centrality of a node is defined as the number of ties that a node has. Closeness centrality of a node is equal to one divided by the total distance (using geodesics in the graph) of this node from all other nodes. Betweenness centrality may be defined loosely as the number of shortest paths that pass through a given node. More precisely this notion is defined as follows (Anthonisse, 1971; Freeman, 1977).

For an undirected network with n nodes the (normalized) betweenness centrality of node a, denoted as $C_B(a)$ is defined as:

$$C_B(a) = \frac{2}{(n-1)(n-2)} \sum_{g,h} \frac{p_{g,h}(a)}{p_{g,h}} \tag{1}$$

In this formula, $p_{g,h}$ denotes the number of shortest paths connecting nodes g and h; $p_{g,h}$ (a) denotes the number of these shortest paths passing through a, where a is not one of the endpoints.

The fourth centrality measure can best be introduced using the matrix representation of a graph. For simplicity we consider only simple graphs. Assume that a graph has n nodes (also called vertices). Then the adjacency matrix $\mathbf{A}$ is a square (n, n) -matrix, such that element $a_{ij} = 1$ if vertex i is connected to vertex j and $a_{ij} = 0$ otherwise (including the diagonal elements a_{jj}). As the graph is undirected a_{ij} is always equal to a_{ji} so that the matrix $\mathbf{A}$ is symmetric. As the matrix $\mathbf{A}$ is symmetric it can be shown that all its eigenvalues are real (Lay, 2003). Recall that an eigenvalue of a matrix is a number λ (in general this may be a complex number) for which there exists a vector $\mathbf{X} = (x_1, x_2, \cdots, x_n)$ such that $\mathbf{A}.\mathbf{X} = \lambda\mathbf{X}$ or stated otherwise:

$$\sum_{j=1}^{n} a_{ij}x_j = \lambda x_i, i = 1, \cdots, n \tag{2}$$

In the case of an adjacency matrix it can be shown that all components of the eigenvector can be chosen to be positive. Then eigenvector centrality of node i is defined as:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} a_{ij}x_j \tag{3}$$

where λ is the largest eigenvalue of matrix $\mathbf{A}$. We see that the eigenvector centrality of a vertex is proportional to the sum of the eigenvector centralities of the vertices to which it is connected. Although this description sounds like a circular definition, equation (3) and the underlying mathematical theory (the Perron-Frobenius theorem) show that it is not and that it leads to a well-defined notion (Meyer, 2000). Density and degree centrality can also be expressed using the adjacency matrix. Density is equal to $\frac{\sum_{i,j=1}^{n} a_{ij}}{n(n-1)}$ and the degree centrality of node j is $\sum_{k=1}^{n} a_{jk}$.

Degree centrality of a node is related to activity; closeness centrality of a node relates to efficiency; a high value for the betweenness centrality of a node reflects a high control on the flow of information by this node; finally a high eigenvector centrality refers to a node that is highly influential. All four centrality measures can be said to be related to leadership.

2 Q-measures for Two Groups: Definition and an Example

In 2004, Flom, Friedman, Strauss and Neaigus, four researchers from New York (USA) introduced a new sociometric network measure, denoted as Q, for individual actors as well as for whole networks (Flom et al., 2004). As this measure captures the idea of bridges between two groups, the higher an actor's Q-value, the more this actor behaves as a bridge/broker between the two groups. Hence using Q-measures leads to an answer of the question: Which nodes are most important in bridging activities between groups in a network? Figure 1 illustrates the type of network Flom et al. (2004) had in mind.

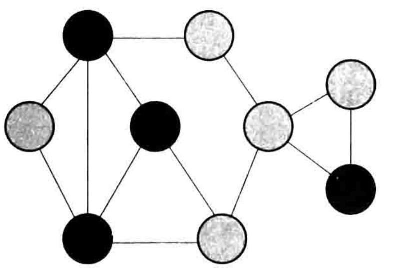

Figure 1 Simple network subdivided into two groups

We note that existing measures such as betweenness centrality did not make a distinction between nodes belonging to different groups. This is the main motivation for the introduction of the Q-measure. An example: consider a network consisting of all universities of two countries (the two groups) linked if they (= at least one researcher in each university) collaborate in a given topic using Q-measures, answers the question: Which universities are the main facilitators of international collaboration? Q-measures are based on geodesics (shortest distances) between nodes.

The defining formulae

Assume that there are T actors (nodes) in the network. Group $\boldsymbol{A}$ contains m nodes, while the other group, denoted as $\boldsymbol{B}$, contains n nodes, hence $T=m+n$. If actor x belongs to group $\boldsymbol{A}$, and assuming for simplicity that actor x is a_m, then the Q-measure for this actor, $Q(x)$, is defined as shown below. In this formula $p_{a,b}$ and $p_{a,b}(x)$ are defined in the definition of betweenness centrality in formula (1).

$$Q(x) = \frac{1}{(m-1)n}\left(\sum_{i=1}^{m-1}\sum_{j=1}^{n}\frac{p_{a_ib_j(x)}}{p_{a_ib_j}}\right) \tag{4}$$

Flom et al. (2004) also introduced a Q-measure for the whole network, denoted as Q_{net}, as the normalized average difference between the most central node (in the Q-sense), denoted as Q^*, and all other nodes. This is:

$$Q_{\text{net}} = \frac{\sum_{i=1}^{m}\left(Q^* - Q(a_i)\right) + \sum_{j=1}^{n}\left(Q^* - Q(b_j)\right)}{T-1} \tag{5}$$

An example: consider Figure 2 consisting of two groups $\boldsymbol{A} = \{a_1, a_2, a_3\}$ and $\boldsymbol{B} = \{b_1, b_2\}$.

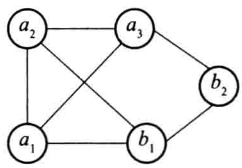

Figure 2 Example network for the calculation of a Q-measure

Table 2 shows all nodes that lie on a shortest path (sometimes there are two shortest paths between nodes situated in different groups). When there is no such node this is denoted by "—".

Table 2 Example paths for the nodes

Nodes	a_1	a_2	a_3
b_1	—	—	a_1/b_2
b_2	b_1/a_3	b_1/a_3	—

$$Q(a_1) = \frac{1}{4}\left(0+\frac{1}{2}+0+0\right)=\frac{1}{8};\ Q(a_3) = \frac{1}{4}\left(0+0+\frac{1}{2}+\frac{1}{2}\right)=\frac{1}{4}$$

$$Q(b_1) = \frac{1}{3}\left(\frac{1}{2}+\frac{1}{2}+0\right)=\frac{1}{3};\ Q(b_2) = \frac{1}{3}\left(0+0+\frac{1}{2}\right)=\frac{1}{6}$$

Finally $Q(a_2) = 0$. In group $\boldsymbol{A}$, a_3 plays a more important role than a_1, while a_2 plays no role at all as a bridge. In group $\boldsymbol{B}$ all nodes play a bridging role, but b_1's role is more important.

3 Literature Review Related to Q-measures for Two Groups

Q-measures were introduced in social network theory by Flom et al. in 2004. The idea was picked up by Rousseau and presented at the 68th ASIST Conference (Rousseau, 2005). Calculations for theoretical examples (line graphs, star graphs) as well as two co-author networks based on real data were presented. A more elaborated example presented first at the 9th International Conference on Science & Technology Indicators (2006, Leuven, Belgium) was published (Chen and Rousseau, 2008). Their result indicated that Cambridge University, Manchester University, Technische Universität Berlin, the Max Planck Institute, Stuttgart University and Forschungszentrum Karlsruhe play the most important roles as bridges between England and Germany (at least within the *Journal of Fluid Mechanics*). It was concluded that having a high degree centrality and being a key node are important factors explaining the ranking of nodes in a network according to Q-value. Moreover, it was found that institutes with a high Q-value have, on average, a higher production than those with a lower Q-value. Moreover, after a presentation by Rousseau in Dalian the idea of using Q-measures was worked out—using a dedicated computer program—by Zhang, Yin and Pang (2009) for the COLLNET network, divided in male and female researchers, as an example.

4 Q-measures for More Than Two Groups

The next step taken in the study of Q-measures was the extension of its definition to more than two groups. This, however, turned out not to be trivial. Depending on the relative importance one gives to groups or to nodes we came up with two definitions (Guns and Rousseau, 2009).

4.1 Definition 1: the Global Structural Q-measure

Consider a network subdivided into S non-overlapping groups. Each group is denoted as G_i ($i=1,\cdots,S$) and contains m_i members. Then the global Q-measure of node a is denoted as $Q_G^S(a)$ (the notation is explained further on) and is defined as:

$$Q_G^S(a) = \frac{2}{S(S-1)} \sum_{k,l} \left(\frac{1}{\mathrm{TP}_{k,l}} \sum_{\substack{g \in G_k \\ h \in G_l}} \frac{p_{g,h}(a)}{p_{g,h}} \right) \tag{6}$$

where $p_{g,h}$ and $p_{g,h}(a)$ are used as defined earlier. The symbol $TP_{k,l}$ refers to the number of possible combinations of elements belonging to groups G_k and G_l . Hence, $TP_{k,l} = \#(G_k \setminus \{a\}) \cdot \#(G_l \setminus \{a\})$, where "#" refers to the number of elements in the set between brackets. The term "global" is used because we will further decompose this global measure into a local and an external one. Observe that in this definition two kinds of normalization are applied.

Yet, a global Q-measure when several groups are present in the network can also be defined differently.

4.2 Definition 2: the Global Basic Q-measure

In this alternative definition, denoted as $Q_G^B(a)$, only one normalization is applied.

$$Q_G^B(a) = \frac{1}{M}\sum_{\substack{g,h\in V \\ \mathrm{group}(g)\neq \mathrm{group}(h)}} \frac{p_{g,h}(a)}{p_{g,h}} \tag{7}$$

where the symbol M is defined as:

$$M = \sum_{k,l} \#(G_k \setminus \{a\}) \cdot \#(G_l \setminus \{a\}) \tag{8}$$

The sum in formula (8) is calculated over all pairs of different groups and the notation group (g) refers to the group to which node g belongs.

When there are only two groups these two definitions (the basic one and the structural one) coincide with equation (4).

4.3 The Difference Between These Two Definitions

We illustrate the different behavior of basic and structural Q-measures by the following example (Figure 3). In this example we use the following convention: nodes a_1, a_2, …form a group, nodes b_1, b_2, …form another group, and so on.

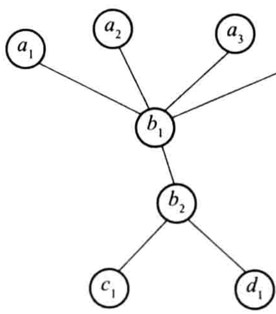

Figure 3 Example: four groups, illustrating the difference between structural and basic Global Q-measures

In the example shown in Figure 3, we have four groups: group A with nodes