Scientometrics (2011) 87:133–147
DOI 10.1007/s11192-010-0332-3

Q-measures and betweenness centrality in a collaboration network: a case study of the field of informetrics

Raf Guns · Yu Xian Liu · Dilruba Mahbuba

Received: 22 July 2010 / Published online: 1 January 2011

Abstract We study global and local Q-measures, as well as betweenness centrality, as indicators of international collaboration in research. After a brief review of their definitions, we introduce the concepts of external and internal inter-group geodesics. These concepts are applied to a collaboration network of 1129 researchers from different countries, which is based on publications in bibliometrics, informetrics, webometrics, and scientometrics (BIWS in short) from the period 1990–2009. It is thus illustrated how international collaboration (among authors from different countries) in BIWS is carried out. Our results suggest that average scores for local Q-measures are typically higher, indicating a relatively low degree of international collaboration in BIWS. The dominating form of international collaboration is bilateral, whereas multilateral collaboration is relatively rare in the field of BIWS. We also identify and visualize the most important global and local actors. Dividing the entire period in four 5-year periods, it is found that most international collaboration in the field has happened in the last time slice (2005–2009). A comparison of the different time slices reveals the non-linear growth of the indicators studied and the international expansion of the field.

Electronic supplementary material The online version of this article (doi:10.1007/s11192-010-0332-3) contains supplementary material, which is available to authorized users.

R. Guns · Y. X. Liu · D. Mahbuba
University of Antwerp, IBW, Venusstraat 35, City Campus, B-2000 Antwerpen, Belgium
e-mail: raf.guns@ua.ac.be

Y. X. Liu (✉)
Electronic Reading Room, Library of Tongji University, Siping Street 1239,
Shanghai 200092, People's Republic of China
e-mail: yxliu@lib.tongji.edu.cn

D. Mahbuba
Asian University for Women, 20/A M.M. Ali Road, Chittagong 4000, Bangladesh
e-mail: dilruba.mahbuba@auw.edu.bd

Keywords Co-author collaborative network · Globalization · Global Q-measure · Local Q-measure · Betweenness · External inter-group geodesic · Internal inter-group geodesic · Evolvement of network

Introduction

During the past few decades, the study of graphs and networks has become important in an increasing number of scientific disciplines, such as physics, biology, computer science, economics, and sociology. The branch of sociology that studies networks is known as social network analysis or SNA (Wasserman and Faust 1994). Within SNA several centrality measures have been developed that indicate how central or important a specific actor is in the network. Recent research has also explored their applicability in bibliometrics and informetrics (e.g., Leydesdorff 2009; Otte and Rousseau 2002; Yan and Ding 2009).

Centrality is a multifaceted notion, and many centrality measures have been devised in SNA and related fields. In this paper, we study three centrality measures that are based on geodesics (shortest paths) between nodes: global Q-measures, local Q-measures, and betweenness centrality. Informally, these measures characterize different aspects of 'bridgeness', the way in which a node in a network forms a 'bridge' between other parts of the network.

We present the results of their application to a co-authorship network of authors with one or more publications on bibliometrics, informetrics, webometrics, and scientometrics (BIWS) during the period 1990–2009. In this network, the nodes represent authors. Two authors are linked if they have co-authored at least one publication. Each node belongs to a group, corresponding to the author's country. Thus, authors from the same country are in the same group.

The remainder of the article is structured as follows. In the next section, we introduce the three measures examined here and provide their formal definitions. In sect. 3, it is discussed how the data for the case study have been collected and processed. After that, we look at the results and discuss them. The last section contains the conclusions.

Definition of Q-measures and betweenness centrality

We now assume that we have a network $G = (V, E)$, consisting of a set V of nodes or vertices and a set E of links or edges between them.

Betweenness centrality is a sophisticated measure that characterizes the importance of a given node for establishing short pathways between other nodes (Freeman 1977). Mathematically, the betweenness centrality of node a can be expressed as

$$C_B(a) = \sum_{g,h \in V} \frac{p_{g,h}(a)}{p_{g,h}} \tag{1}$$

where $p_{g,h}$ is the number of geodesics between nodes g and h and $p_{g,h}(a)$ is the number of geodesics between nodes g and h that pass through a. Normalization leads to a number between 0 and 1. For an undirected network with n nodes:

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

It has become one of the standard centrality measures in social network analysis, along with degree centrality, closeness centrality and eigenvector centrality. Many variants of betweenness centrality have since been proposed, such as group betweenness centrality (Everett and Borgatti 1999) and edge betweenness centrality (Girvan and Newman 2002).

Depending on the network, it may be the case that each node belongs to a larger group. For instance, authors belong to a department, a university or a country; articles belong to a journal; journals belong to a publisher or a scientific discipline. Flom et al. (2004) introduced a new indicator, called Q-measures, for the brokerage role of nodes between two groups in a connected, undirected, unweighted network. Q-measures are another variant of betweenness centrality where only shortest paths between nodes from the different groups are taken into account. We will refer to shortest paths between nodes from different groups as *inter-group geodesics*. Furthermore, we will distinguish (for a given node a) between external and internal inter-group geodesics. If a belongs to group A, then any geodesic $\gamma_{b,c}(a)$where $b \in A$ and $c \in V \backslash A$ (or vice versa) is an *internal inter-group geodesic* of a. A geodesic $\gamma_{b,c}(a)$where b, $c \in V \backslash A$ is an *external inter-group geodesic* of a.

If the network consists of two groups G (consisting of m nodes) and H (consisting of n nodes), then the Q of node a is defined as follows.

$$Q(a) = \frac{1}{TP} \sum_{\substack{g \in G \\ h \in H}} \frac{p_{g,h}(a)}{p_{g,h}} \tag{3}$$

Here, TP denotes the total number of possible combinations of nodes from the two groups. If $a \in G$, then $TP = (m-1) \cdot n$ and if $a \in H$, then $TP = (n-1) \cdot m$.

Q-measures have subsequently been studied and applied in (Rousseau 2005; Chen and Rousseau 2008; Rousseau and Zhang 2008; Guns and Rousseau 2009). Rousseau and Zhang (2008) introduced Q-measures for networks with weighted links. Guns and Rousseau (2009) expanded the definition to networks with any finite number of groups and showed that in this case one can define both a global and a local variant. Assume that there are S groups ($2 \leq S < +\infty$) $G_1, \ldots, G_S$ and that the number of nodes in group G_i is denoted as m_i, then the global Q-measure of node a is defined as:

$$Q_G(a) = \frac{1}{C} \sum_{k,l} \left(\frac{1}{TP_{k,l}} \sum_{\substack{g \in G_k \\ h \in G_l}} \frac{p_{g,h}(a)}{p_{g,h}} \right) \tag{4}$$

where $C = \binom{S}{2} = \frac{S(S-1)}{2}$ is the number of ways to choose two different groups and $\frac{1}{TP_{k,l}} \sum_{\substack{g \in G_k \\ h \in G_l}} \frac{p_{g,h}(a)}{p_{g,h}}$, denoted as $Q_{k,l}(a)$, is the binary Q-measure of a with respect to groups G_k and G_l. In the context of Q-measures for several groups such a Q-measure is called a partial Q-measure. $TP_{k,l}$ is the number of possible node combinations from groups G_k and G_l. This is $m_k \cdot m_l$ if a does not belong to G_k or G_l. It is $(m_k-1) \cdot m_l$ if a belongs to G_k and $m_k \cdot (m_l-1)$ if a belongs to G_l. If m_k-1 or m_l-1 is zero the corresponding TP is taken to be 1 (or any finite number). The exact value plays no role as $p_{g,h}(a) = 0$ in this case. The global Q-measure of node a always satisfies the inequality $0 \leq Q_G(a) \leq 1$.

The local Q-measure for actor a belonging to group G_k with m_k nodes is defined as:

$$Q_L(a) = \frac{1}{S-1}\sum_{l\neq k}\left(\frac{1}{TP_{k,l}}\sum_{\substack{g\in G_k\\ h\in G_l}}\frac{p_{g,h}(a)}{p_{g,h}}\right) = \frac{1}{S-1}\sum_{l\neq k}\left(\frac{1}{(m_k-1)\cdot m_l}\sum_{\substack{g\in G_k\\ h\in G_l}}\frac{p_{g,h}(a)}{p_{g,h}}\right) \quad (5)$$

Note that eqn (5) uses partial Q-measures as well, but the cases where a does not belong to G_k or G_l are not used here. Just as for the global Q-measure, the local Q-measure of node a satisfies the inequality $0 \le Q_L(a) \le 1$.

Q-measures are concerned with the question to what extent a node forms part of inter-group geodesics. Especially when there are many groups in a network, it is possible that a node forms part not only of internal inter-group geodesics but external ones as well. Local Q-measures just consider the internal inter-group geodesics, whereas global Q-measures consider both the internal and external inter-group geodesics.

Remark. What happens if the network is unconnected? It was originally noted by Freeman (1977) and recently emphasized by Brandes (2008) that the definition of betweenness centrality is applicable in this case as well. If g and h belong to different components, there exists no (shortest) path from g to h and therefore $p_{g,h}(a) = p_{g,h} = 0$. As this implies a division by zero, 0/0 is set equal to zero, by convention. Given this convention, one can determine the betweenness centrality, local or global Q-measure for any node in a network, even if it is an isolate (i.e., unconnected to any other node): in that case, $C_B(a) = Q_G(a) = Q_L(a) = 0$. Note that, in general, this favors the nodes in the larger components: other things being equal, if a belongs to a small component, $p_{g,h}(a)$ will more often be equal to 0, simply because g and/or h belong to another component, leading to lower values for each of these centrality measures. It accords with 'common sense' that central nodes in a large component are more important than central nodes in a small component.

Data and methods

In order to find out how global Q-measures, local Q-measures and betweenness centrality behave practically, we have applied them to a collaboration network of researchers in bibliometrics, informetrics, webometrics and scientometrics (BIWS in short). In this section, we explain how the data were gathered and cleansed, and how the measures were determined.

Data gathering

Data were collected from Thomson Reuters's Web of Science (WoS) on August 26, 2009 using the following query, restricted to the time period 1990–2009:

TS = (informetric*OR bibliometric*OR scientometric*OR webometric*) OR
SO = (Scientometrics OR Journal of Informetrics) OR CF = Scientometric*

This query yielded 4457 unique records. The advantage of searching by topic (TS) rather than by publication outlet (SO) is that we can also include articles on these topics that appeared in journals outside the field of information science. We should, however, stress that if these keywords are not used the article cannot be found with these queries. Therefore, articles published in journals that are specifically focussed on BIWS (*Scientometrics* and, more recently, *Journal of Informetrics*) are included as well. Finally, the last part of the query adds all articles from the ISSI conference proceedings. Note that this part of the query is specific enough not to include other conferences,[1] but broad enough to include all ISSI conferences in WoS for the given time period. The ISSI conferences are collected in the WoS under several different names,—sometimes, the same conference can have several names (either intended or unintended, e.g., spelling errors). Of course, we do not claim the current data set to be 100% exhaustive for the time period considered.

Data preprocessing

All data were pre-processed using the Network Workbench Tool or NWB (Börner 2009), freely available from http://nwb.slis.indiana.edu. Several steps have been taken to ensure a high-quality data set. First, duplicate records were removed, such that each publication counts exactly once. Then we proceeded to extract a co-author network. Most pre-processing is performed on this co-authorship network.

Q-measures are calculated on a network where each node is assigned to exactly one group. In the context of a collaboration network, this presupposes three conditions:

1. Each author is represented by exactly one node;
2. Each publication (or at least each multi-authored publication) is assigned to the correct authors, such that links accurately reflect collaboration ties;
3. Each author is correctly assigned to a group (in our case, a country).

The first two conditions are related to the process of author name disambiguation (Smalheiser and Torvik 2009). Indeed, problems of homonymy and synonymy can severely complicate the fulfillment of both conditions. We now discuss some characteristics we observed in this regard.

In our data, we found that homonymy was most likely to occur in Chinese names, mainly due to some common family names. For instance, there were three different people with the name 'Zhang, J' in our data. If the given name is available in WoS, it can help to disambiguate between authors. In cases where both given name and family name are the same, one needs to check other features (see further).

Synonymy is quite common in our data: many authors are listed more than once with different variations of their names. In most cases, the synonymous names are similar, often differing by only one or two characters. For example, András Schubert is present as 'Schubert, A', 'Schubert, AP' and 'Schuber, A'. We used the Duplicate Node Detector of NWB to detect these cases. Manual inspection was used to weed out authors that were incorrectly flagged by NWB as duplicates. In this article, we use (one of) the name variant(s) as it occurs in the Web of Science; this implies that diacritics, such as umlauts or accents, are not preserved (e.g., 'Glanzel' rather than 'Glänzel').

Some synonymous names are less similar and therefore missed by the Duplicate Node Detector. Possible apparent reasons are:

- Differences in indexation, e.g., 'Vandesompel, H', 'De Sompel, HV', and 'Sompel, HVD';
- Name changes and name variations. Researchers from Spanish-speaking countries typically have two surnames, and a common variation involves using only one or both of these (e.g., 'Anegon, FD' versus 'de Moya-Anegon, F', and 'Moya-Anegon, FD');

---

[1] There is one exception: the 'Second Berlin Workshop on Scientometrics and Informatics/Collaboration in Science and in Technology' is also included. Since this conference also deals with scientometrics, we do not consider this a downside.

- Errors in the database, e.g., ‘Briggs, MB’, and ‘Eriggs, ME’ (the latter apparently being due to an OCR error).

Other features that can be used to determine synonymy and homonymy include:

- Co-authorship. Co-authors are always different persons, which is especially useful to distinguish between collaborating family members. On the other hand, authors can often be recognized on the basis of their co-authors (Kang et al. 2009);
- Institutional affiliation and country;
- E-mail address;
- Cited and citing authors (authors tend to self-cite);
- Publication outlet (authors tend to publish in a limited range of journals);
- Frequency (erroneous recordings of a name typically occur only once).

We have tried to use a combination of the above features to accurately determine the nodes and links of the network (condition 1 and 2). On some occasions, it proved easier to check the original articles or the publication list on the authors’ web pages.

The third condition pertains to the assignment of an author to the right group,—in our case, to the right country. We should first clarify what exactly we mean by an author’s country. Indeed, one author may have several affiliations, belonging to different countries. This mainly happens over time: an internationally mobile author is first active in country *a*, a few years later in country *b* and so on. An author may also hold several affiliations in different countries at the same time. In this situation we assign the author to the country where he/she has published most papers.

After making sure that the three conditions were fulfilled, we removed all isolate nodes from the network. These are authors that do not have any co-authors in the network. They are not interesting in this context, because for each isolated node $Q_L = Q_G = C_B = 0$.

The resulting co-authorship network contains 4703 nodes and 8410 links. This network consists of 757 connected components, the largest of which contains 1129 nodes. This is significantly larger than the second largest component, which contains only 81 nodes. The giant component includes most, but not all, researchers whose primary field is informetrics, whereas most of the small components consist of researchers from other fields, with only one or a few publications related to informetrics. For this reason (and for practical reasons), we decided to limit the current study to the 1129 authors in the giant component. These authors are from 46 different countries.

Calculation of Q-measures and betweenness centrality

The data were exported from NWB for use in the NetworkX software (http://networkx.lanl.gov). NetworkX does not provide a finished program,—rather, it is a ‘toolbox’ that can be expanded for specific network analysis tasks. While the calculation of betweenness centrality is part of NetworkX, we built upon it for Q-measures calculation. Our implementation uses a breadth-first search to determine all geodesics between two nodes *g* and *h*. Based on these geodesics, we can determine for a given node *a* the fraction $p_{g,h}(a)/p_{g,h}$. Doing this for all relevant node pairs yields the global and local Q-measure. It should be noted that this approach is much less efficient than the algorithm of Brandes (2001, 2008) for betweenness centrality and related measures, but for the purposes of the current study, this limitation does not pose real problems.

Table 1 Number of articles and authors in four consecutive time periods

Period	Articles	Authors	Collaborating authors	% collaborating authors	Studied authors
1990–1994	619	663	496	74.8	211
1995–1999	843	905	687	75.9	284
2000–2004	1059	1310	1150	87.8	451
2005–2009	1936	2776	2531	91.2	691

Time-sliced data

One of the goals of our research was to explore how Q-measures and betweenness centrality may evolve over time. We therefore sliced the data of the entire period 1990–2009 into four periods of each five years: 1990–1994, 1995–1999, 2000–2004, and 2005–2009. The raw numbers for these periods are summarized in Table 1; they reflect the strong growth of the field of BIWS over the last decade (Egghe 2005). The number of articles has more than tripled, and the number of authors in the last period is more than four times higher than in the first period. Moreover, there is a clear trend towards more collaboration, as can be seen from the fourth and fifth columns of Table 1.

The same preprocessing steps as described earlier were applied to the time-sliced data. We decided to focus on the 1129 authors that are studied for the entire period, mainly because the time periods were too short for any sufficiently large components to form. Of course, many of these authors have only published in one or two of the time periods. The number of studied authors for each time period can be found in the last column of Table 1.

Results and discussion

Q-measures and betweenness centrality in the entire period 1990–2009

First, we examine the results as calculated for the entire time period (1990–2009). All results are obtained for the network formed by the giant component, consisting of 1129 nodes. We find high correlation (Pearson's $r = 0.93$) between Q_G and C_B, whereas correlations with Q_L are lower (for Q_G, $r = 0.59$ and for C_B, $r = 0.50$). This is because Q_L is based on a subset of the geodesics necessary for the calculation of Q_G, which is itself based on a subset of all geodesics, which are necessary for calculating C_B.

The role of individual researchers in international collaboration, as reflected by centrality measures

In this section, we determine which individual researchers play a significant role in international collaboration, as reflected by the three centrality measures under consideration. There are 335 researchers out of 1,129 in the giant component for whom at least one of the two Q-measure variants is larger than 0. These researchers are direct or indirect facilitators of international collaboration. Table 2 shows the top ten authors as ranked by Q_G.

The ranking according to betweenness (not shown) is very similar to the one by Q_G. The ranking according to the local Q-measures, on the other hand, shows a significantly

Table 2 Top ten authors ranked by global Q-measure

Rank	Name	C_B	Q_G	Q_L	Country
1	Glanzel, W	0.3538	0.3415	0.4495	Belgium
2	Kretschmer, H	0.2617	0.2425	0.2955	Germany
3	Rousseau, R	0.1823	0.1971	0.2611	Belgium
4	Thelwall, M	0.1930	0.1689	0.3139	England
5	Lewison, G	0.0846	0.1535	0.3884	England
6	Leta, J	0.1064	0.1337	0.9929	Brazil
7	Scharnhorst, A	0.1464	0.1298	0.0728	Netherlands
8	Borner, K	0.1383	0.1266	0.3320	USA
9	Leydesdorff, L	0.0834	0.1174	0.2191	Netherlands
10	Klavans, R	0.0928	0.1172	0.2117	USA

different pattern, revealing the importance of a given author to his own country's participation in international collaboration. An interesting example is Tomas Cahlik (Czech Republic), the only author whose $Q_L = 1$. In other words, all geodesics between the Czech Republic and other countries pass through him. Moreover, Cahlik is not part of any external inter-group geodesic. This implies that his $Q_G = 2/S$ (= 2/46). The top ten authors as ranked by their local Q-measure are shown in Table 3.

It should be stressed that ranking according to Q_L may be deceiving. For instance, there are only two authors from the Czech Republic, which places Cahlik's top spot in a somewhat different light. Indeed, the more authors there are from a given country, the harder it is for any of these authors to achieve a high Q_L value. The same is true for Q_G, but to a much more limited extent. Meaningful comparisons on the basis of Q_L are thus only possible within one country (or between authors from countries with a comparable number of authors). Table 3 should be interpreted with this caveat in mind. Local Q-measures may, however, give one some reference information when looking for collaborators in a specific region.

It turns out that there exist remarkable size differences between the three indicators studied. Table 4 summarizes some descriptive statistics. Note that we use the third quartile rather than the median, since the median for each of these measures equals zero. The difference between C_B and Q_G turns out to be small, but Q_L is much higher than the other two measures. Informally, the difference between Q_G and Q_L mainly depends on how often a node belongs to external inter-group geodesics. In our case each of these statistics is much smaller for Q_G than for Q_L, implying that most nodes mostly form bridges for nodes from their own group. If homophily (McPherson et al. 2001) occurs in a network (i.e., nodes prefer to link to nodes from the same group), the average local Q-measure for this network and these groups will typically be larger than the average global Q-measure. It seems reasonable to expect authors to primarily collaborate with colleagues from the same country—i.e., homophily with regard to country—which explains why in our case study the local Q-measure statistics far exceed their global counterparts. If research happens in larger and/or more international teams, global Q-measures will typically become larger. According to our data, it seems that international collaboration in the field of BIWS is not very intense; collaboration is mostly local and bilateral. There are 42 researchers whose Q_G is larger than Q_L. By definition these nodes are part of at least one external inter-group geodesic between nodes from other countries.

Table 3 Top ten authors ranked by local Q-measure

Rank	Name	C_B	Q_G	Q_L	Country
1	Cahlik, T	0.0018	0.0435	1.0000	Czech Republic
2	Leta, J	0.1064	0.1337	0.9929	Brazil
3	Horlesberger, M	0.0123	0.0380	0.8750	Austria
4	Schiebel, E	0.0106	0.0326	0.7500	Austria
5	Falagas, ME	0.0141	0.0317	0.7143	Greece
6	Kim, H	0.0159	0.0301	0.6923	South Korea
7	Fischer, AL	0.0229	0.0378	0.6676	Scotland
8	Schubert, A	0.0728	0.0683	0.6026	Hungary
9	Kongthon, A	0.0280	0.0276	0.5716	Thailand
10	Etemad, S	0.0071	0.0248	0.5714	Iran

Table 4 Comparison of the three measures over all nodes

	Q_G	Q_L	C_B
Sum	5.7783	31.4258	5.3561
Average	0.0051	0.0278	0.0047
Third quartile	0.0002	0.0018	0.0006

It rarely occurs that the local Q-measure of a node is 0 while the global Q-measure is greater than 0. No cases can be found in the network of the entire period, but some rare examples do occur in the time-sliced networks. In that case, an author is part of at least one external inter-group geodesic, but not of any internal inter-group geodesics.

$Q_G(a)$ and $Q_L(a)$ being equal to 0 does not necessarily mean that there are no links between node a and nodes from other groups, only that a is not part of an inter-group geodesic. It is possible that $Q_G(a) = Q_L(a) = 0$ but $C_B(a) > 0$. This is the case for 46 researchers. It is very likely that these researchers have only collaborated with compatriots. Although the betweenness centrality of these nodes is generally quite low, this indicates that these nodes play a bridging role within their own country even if they are not involved in international collaboration.

The role of countries in international collaboration

In the current case study, the groups represent countries. We are interested in the question which countries play a major role in international collaboration. For each country, we count the number of authors, the number of authors whose $Q_G > 0$ and the sum of each of the three centrality measures. The results are shown in Table 5, restricted to the ten countries with the largest summed Q_G.

Of the 46 countries in the data, there are 13 countries whose summed centrality measures equal zero. The authors from the other 33 countries take on at least some bridging role in international collaboration. In general, the United States clearly have the advantage in terms of number of authors and centrality. Remarkably, the network contains almost as many researchers from Spain and Spain's summed Q_L even exceeds the USA's. Its global Q-measure is however much lower. The specific role of China and its ties to other countries are discussed in more detail by Guns and Liu (2010).

Table 5 Country domination in international collaboration

Country	Authors	Authors with $Q_G > 0$	C_B	Q_G	Q_L
USA	186	54	0.7669	0.7514	1.7457
Belgium	51	17	0.6336	0.6440	1.1775
Netherlands	53	25	0.6682	0.6061	1.3377
England	82	19	0.4554	0.4908	1.1014
France	95	31	0.2998	0.4443	1.4626
India	63	20	0.3246	0.3800	1.5404
Germany	46	16	0.3170	0.2719	0.8850
Spain	150	40	0.6428	0.2533	1.8965
Brazil	34	13	0.1680	0.1999	1.6172
China	67	17	0.1764	0.1808	0.9243

Network visualization

Figure 1 shows the network as visualized by Pajek, using the Kamada-Kawai layout algorithm. Nodes with different color and/or shape belong to different countries, although (for practical reasons) the 31 ‘smallest’ countries (with the smallest amount of authors) all have light-gray triangles for nodes. The online supplementary material contains a full-color version. This visual representation confirms that most nodes tend to group together with other nodes of the same country. It is also easy to see that some countries are as a group more central than others (cf. Everett and Borgatti 1999). For instance, Belgium (black diamonds, in the centre) is fairly central, while Chile (on the outer right) is mostly on the outside of the network. This can obviously affect a single node’s global Q-measure (and betweenness), because it is easier for a Belgian author to form part of many (external inter-group) geodesics than it is for a Chilean one.

The visualization also reveals some interesting characteristics that would be much harder to determine from a set of indicators. It is, for instance, much easier to see how coherent a specific group is, that is, how well its nodes are connected to one another. One can observe that there are roughly two Spanish (dark circles) communities: one (top central) is mainly connected to foreign colleagues via Aguillo, IF and Bordons, M, whereas the other is positioned in the top right, with Torres-Salinas, D and De Moya-Anegon, F as main international bridges. Another interesting example is The Netherlands (light diamonds), which also consists of two clearly separate communities: the Leiden CWTS community appears in the middle, slightly to the right, whereas another, much smaller community is positioned on the right. The latter contains researchers like Leydesdorff, L and Van Den Besselaar, P. China (white diamonds, slightly below the middle) seems fairly coherent.

The same network with the same layout is also shown in Fig. 2, where the nodes are sized according to their Q-measure. Vertical size reflects Q_L, while horizontal size reflects Q_G. Some important nodes have been manually labeled. Again we see that, due to homophily, the local Q-measure is in most cases much higher than the global one. Moreover, no single country can be said to really dominate the network: there appear to be typically no more than one or two major ‘bridge authors’ per country.

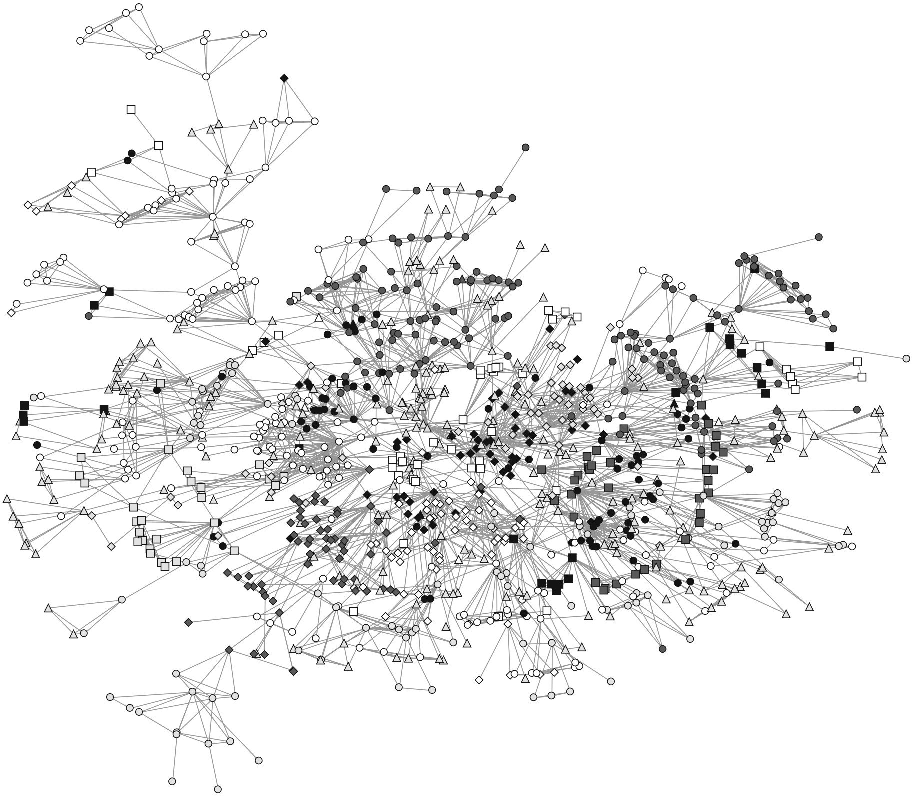

Fig. 1 Giant component of collaboration network in entire period 1990–2009. Larger groups are: Belgium (*black diamonds*), China (*white diamonds*), France (*light circles*), England (*black circles*), India (*dark diamonds*), Netherlands (*light diamonds*), Spain (*dark circles*), USA (*white circles*). See supplementary material for a full-color version

Evolutionary characteristics of international collaboration in BIWS

The authors in the largest component of the entire period constitute the main body of researchers in BIWS. As previously explained, we created consecutive time-sliced networks consisting of the same authors (if they published during a given time slice) for the periods 1990–1994, 1995–1999, 2000–2004, and 2005–2009. Subsequently, we calculated the same set of centrality measures for each time-sliced network, to see how the collaboration network evolved over time. Table 6 lists some general characteristics of each time slice. Over time, the network becomes both larger and denser: the number of nodes increases, as does the number of links per node.

We calculated the Pearson correlation coefficient for each centrality measure between the entire period and the time slices. As could be expected, correlations for the later periods are much better, since these networks are much closer to the network for the entire period. However, the measures do not grow linearly. For 748 nodes, the centrality measures equal zero across all time-sliced networks and the network of the entire period. For the remaining 381 nodes that have at least one centrality measure larger than 0, we found that there are 77 nodes whose centrality measures equal zero across all time-sliced networks but not in the

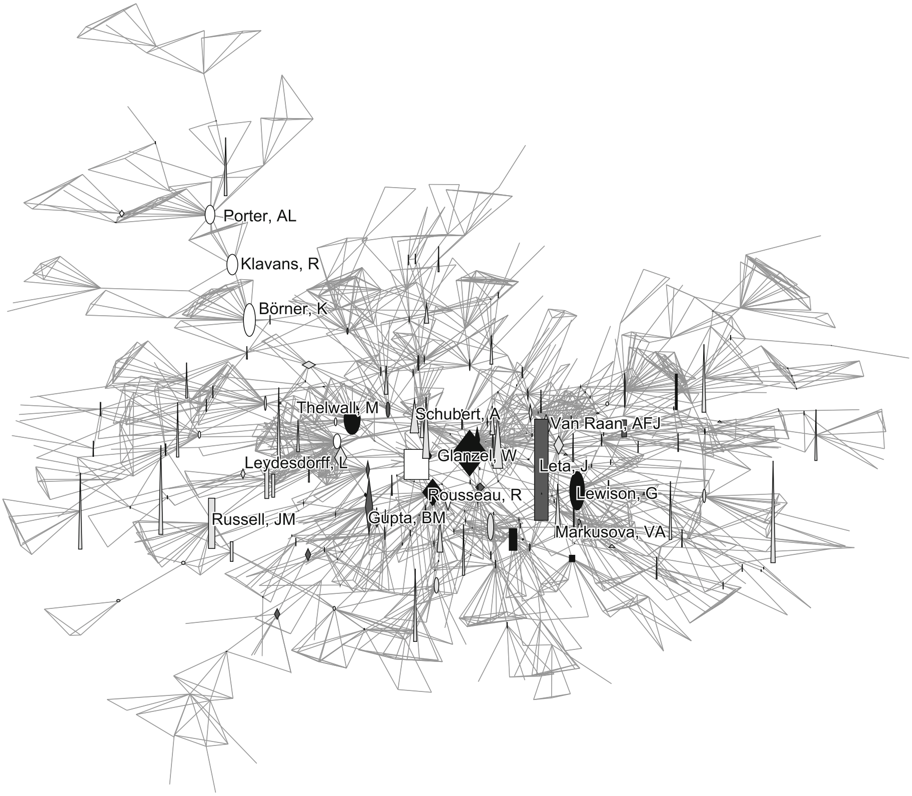

Fig. 2 Giant component of collaboration network in entire period 1990–2009. Vertical node dimension is scaled with Q_L, horizontal node dimension is scaled with Q_G. Some prominent nodes have been labeled. Colors and shapes are the same as in Fig. 1

Table 6 Characteristics of BIWS collaborative network in different time slices

	Nodes	Links	Nodes with $Q_G > 0$	Total C_B	Total Q_G	Total Q_L
1990–1994	211	294	32	0.0520	0.0373	0.2847
1995–1999	284	407	53	0.0692	0.0385	0.3590
2000–2004	451	752	72	0.3285	0.3959	1.9702
2005–2009	691	1460	177	2.5964	1.5995	9.8844

network for the entire period. These researchers form part of geodesics that are only present when we look at a longer time period: they began collaborating with different people over the course of time. Some examples include Burrell, QL, Hartley, J, and Laudel, G. The reverse is also possible: the centrality measures of Van Vijk, E (the Netherlands) are larger than 0 in the period 1995–1999, but all his centrality measures for the entire period are zero. This is caused by later collaborations of his collaborator Nederhof, leading to new (shorter) geodesics in the network for the entire period.

Authors who have some degree of centrality both in every time slice and in the entire period may be regarded as the main facilitators of international collaboration in the field of

Table 7 Central authors in all time slices and in the entire period

Name	Rank according to Q_G
Glanzel, W	1
Kretschmer, H	2
Rousseau, R	3
Lewison, G	5
Leydesdorff, L	9
Tijssen, R	14
Moed, H	15
Schubert, A	22
Markusova, VA	24
Rao, IKR	34
Braun, T	46
Bordons, M	91
Van Raan, AFJ	101
Garg, KC	187

BIWS in 1990–2009. Most of them (64%) are also ranked in the top 25, or even in the top 10 (36%, see Table 2). We list these authors and their rank in Table 7.

Some newer and/or younger authors do not occur in the older time slices, but have still gained high Q_G values. These ‘new talents’, taken from the 20 researchers with the highest Q_G, are listed in Table 8, along with the year they first appeared in the network and the time period when they got their first $C_B > 0$ and $Q_G > 0$. Remarkably, Klavans, R occupies the tenth spot of Table 2, but does not have $C_B > 0$ or $Q_G > 0$ in any time-sliced network. Closer scrutiny of Klavans’s publication record reveals that in 2001 he published a paper with colleagues from Israel and South Africa. Then, no papers of his are found in our data set until 2005, when he published several articles in collaboration with Boyack, K and Borner, K, both of which are more central in the time-sliced network. This illustrates how time-sliced data may complement and nuance the rankings found for the entire period.

To determine how these fluctuations in centrality affect the relative position of countries, we sum up the Q_G scores per country. The result can be seen in the pie charts in the online material. There is a clear tendency towards more countries getting involved in international collaboration. Some countries (Belgium, France, England, Germany, Hungary, India, the Netherlands, Russia, Spain, and USA) are present in every time slice.

Table 8 Central authors in entire period but not in all time slices

Name	Country	Entered network in	First C_B >0 in	First Q_G >0 in
Thelwall, M	England	2001	2000–2004	2000–2004
Leta, J	Brazil	1996	1995–1999	2000–2004
Borner, K	USA	2003	2005–2009	2005–2009
Klavans, R	USA	2001	None	None
Meyer, M	England	1998	2000–2004	2000–2004
Kundra, R	India	1996	2000–2004	2000–2004
Vaughan, L	Canada	2002	2000–2004	2000–2004
Davis, M	Australia	1999	2000–2004	2000–2004

The first period (1990–1994) is relatively evenly divided among six countries, with the Netherlands and USA as the main players. Note, though, that for this period we only have little data (cf. Table 6). There is a significant change in the next period (1995–1999), when the Netherlands clearly dominate the international BIWS scene, accounting for almost 50% of the sum of all global Q-measure. According to any of the measures used, it absolutely dominates the international scene. The later two periods show a tendency towards a more even distribution between countries. Moreover, from 2000 onwards many ‘new’ countries appear to engage in international collaboration.

It is remarkable that the USA ranks relatively low in the last three time-sliced networks. When all the data are merged into one network, however, the global Q-measure of the USA exceeds that of all other countries.

China’s part has grown significantly throughout these four time slices. In 1990–1994, there are no Chinese BIWS researchers engaged in international collaboration. Their relative share grows in the following two time slices and then slightly decreases in the last one,—although in absolute terms, the last period is stagnating rather than decreasing. One possible factor contributing to the growth of China in the period 2000–2004 is the nineth ISSI conference held in China, which significantly promoted Chinese BIWS research.

Conclusions

We have studied global and local Q-measure, as well as betweenness centrality, and applied these measures to a real-world collaboration network of 1129 BIWS researchers from the period 1990–2009. Our findings suggest that in this kind of network one will typically find higher average scores for the local Q-measure. The much higher scores for local Q-measure than for the other two measures indicate that most collaboration in BIWS is local or bilateral. Nevertheless, some authors have clearly taken on a role of international bridge in BIWS research collaboration.

Dividing the entire period in four 5-year periods we have found that international collaboration (and, indeed, collaboration in general) in BIWS has been steadily on the rise throughout the past two decades. During the last period (2005–2009), many more researchers and countries have engaged in international collaboration. In general, our study illustrates how these measures can be used for large-scale practical applications. It also paints a picture of BIWS as field that is both expanding and intensifying its (international) collaboration.

Acknowledgments A preliminary version of the current paper was prepared on the occasion of prof. Ronald Rousseau’s 60th birthday. We are very grateful to prof. Rousseau for his insightful comments and advice, which have helped to significantly improve the paper.

References

Börner, K. (2009). Rete-Netzwerk-Red: Analyzing and visualizing scholarly networks using the Network Workbench Tool. In B. Larsen & J. Leta (Eds.), *Proceedings of ISSI 2009—the 12th international conference on scientometrics and informetrics* (pp. 619–630).

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology, 25*(2), 163–177.

Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Social Networks, 30*(2), 136–145.

Chen, L. X., & Rousseau, R. (2008). Q-measures for binary divided networks: Bridges between German and English institutes in publications of the *Journal of Fluid Mechanics*. *Scientometrics, 74*, 57–69.

Egghe, L. (2005). Expansion of the field of informetrics: Origins and consequences. *Information Processing and Management, 41*(6), 1311–1316.

Everett, M. G., & Borgatti, S. P. (1999). The centrality of groups and classes. *Journal of Mathematical Sociology, 23*, 181–202.

Flom, P. L., Friedman, S. R., Strauss, S., & Neaigus, A. (2004). A new measure of linkage between two sub-networks. *Connections, 26*(1), 62–70.

Freeman, L. C. (1977). A set of measures of centrality based upon betweenness. *Sociometry, 40*, 35–41.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *PNAS, 99*, 7821–7826.

Guns, R., & Liu, Y. X. (2010). Scientometric research in China in the context of international collaboration. *Geomatics and Information Science of Wuhan University, 35*, 112–115. (ICSUE 2010 special).

Guns, R. & Rousseau, R. (2009). Gauging the bridging function of nodes in a network: Q-measures for networks with a finite number of subgroups. In B. Larsen & J. Leta (Eds.), *Proceedings of ISSI 2009—the 12th international conference on scientometrics and informetrics* (pp. 131–142).

Kang, I. S., et al. (2009). On co-authorship for author disambiguation. *Information Processing and Management, 45*(1), 84–97.

Leydesdorff, L. (2009). How are new citation-based journal indicators adding to the bibliometric toolbox? *Journal of the American Society for Information Science and Technology, 60*(7), 1327–1336.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415–444.

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science, 28*(6), 441–453.

Rousseau, R. (2005). Q-measures for binary divided networks: An investigation within the field of informetrics. In *Proceedings of the 68th ASIST conference* (Vol. 42, pp. 675–696).

Rousseau, R., & Zhang, L. (2008). Betweenness centrality and Q-measures in directed valued networks. *Scientometrics, 75*, 575–590.

Smalheiser, N. R., & Torvik, V. I. (2009). Author name disambiguation. *Annual Review of Information Science and Technology, 43*, 287–313.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University: Structural Analysis in the Social Sciences.

Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology, 60*(10), 2107–2118.