Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Unnormalized and normalized forms of gefura measures in directed and undirected networks*

Raf GUNS[1], Ronald ROUSSEAU[‡1,2]

([1]*Institute for Education and Information Sciences, University of Antwerp, Venusstraat 35, Antwerp B-2000, Belgium*)

([2]*Department of Mathematics, KU Leuven, Celestijnenlaan 200B, Leuven B-3001, Belgium*)

E-mail: raf.guns@uantwerpen.be; ronald.rousseau@kuleuven.be; ronald.rousseau@uantwerpen.be

Received Dec. 9, 2014; Revision accepted Mar. 12, 2015; Crosschecked Mar. 13, 2015

Abstract: In some networks nodes belong to predefined groups (e.g., authors belong to institutions). Common network centrality measures do not take this structure into account. Gefura measures are designed as indicators of a node's brokerage role between such groups. They are defined as variants of betweenness centrality and consider to what extent a node belongs to shortest paths between nodes from different groups. In this article we make the following new contributions to their study: (1) We systematically study unnormalized gefura measures and show that, next to the 'structural' normalization that has hitherto been applied, a 'basic' normalization procedure is possible. While the former normalizes at the level of groups, the latter normalizes at the level of nodes. (2) Treating undirected networks as equivalent to symmetric directed networks, we expand the definition of gefura measures to the directed case. (3) It is shown how Brandes' algorithm for betweenness centrality can be adjusted to cover these cases.

Key words: Networks subdivided in groups, Partitions, Gefura measures, Q-measures, Brokerage role, Directed and undirected networks, Brandes' algorithm

doi:10.1631/FITEE.1400425 **Document code:** A **CLC number:** TP393; G350

1 Introduction

Networks are abundant. Indeed, road maps represent the network of cities and highways, and similarly we have other transportation networks such as the worldwide air transport network, metabolic networks (Barrat *et al.*, 2004; Guimerà and Amaral, 2005; Guimerà *et al.*, 2005), and the shipping and harbor network. The network that is probably best known, the Internet, consists of a worldwide assemblage of local, regional, and global academic, business, government, private, and public computer networks. Many research topics across all disciplines are nowadays studied from a network perspective. Large-scale analyses of so-called complex networks reveal that the same structural features, such as skewed degree distributions and local clustering, can emerge in different fields (Christensen and Albert, 2007). This underlines the importance of network studies.

The field of informetrics is no exception to this trend (e.g., Otte and Rousseau (2002) and Ding (2011)). Maps of science are constructed based on the complete Web of Knowledge, Scopus, etc. Topics such as collaboration, diffusion, and citation have been studied frequently from the perspective of social network analysis. Moreover, a whole new subfield related to the Internet and the World Wide Web, namely webmetrics, has emerged within informetrics.

In some networks, nodes belong to predefined groups. For instance, a network of friendships in school may have pupils as nodes and classes as

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 71173154)

ORCID: Raf GUNS, http://orcid.org/0000-0003-3129-0330; Ronald ROUSSEAU, http://orcid.org/0000-0002-3252-2538

groups. Likewise, a citation network may have articles as nodes and journals as groups. Previous research has introduced so-called Q-measures (Flom *et al.*, 2004), or the newer and preferred term, gefura measures, as indicators of a node's brokerage role between groups. In this article we make the following new contributions to the study on gefura measures:

1. Analogous to the treatment of betweenness centrality by Brandes (2001; 2008), we start the discussion by considering gefura measures in directed networks. Undirected networks are then equivalent to symmetric directed networks. That is, each undirected link $\{a, b\}$ is equivalent to two directed links (a, b) and (b, a). Fig. 1 contains an example.

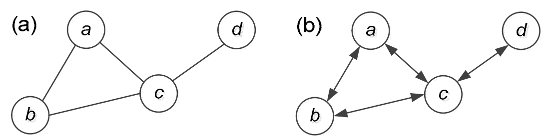

Fig. 1 Undirected network (a) treated as equivalent to a directed network with bidirectional links (b)

2. We systematically study unnormalized gefura measures and show that, next to the 'structural' normalization that has hitherto been applied, a 'basic' normalization procedure is possible.

3. We show that 'structural' normalization pays more attention to the group level, whereas 'basic' normalization pays more attention to the level of nodes.

4. Building on the work of Brandes (2008), an efficient algorithm is introduced to calculate unnormalized or basic gefura measures in both directed and undirected networks.

The main aim of this paper is to provide a general theoretical framework for studying the bridging role of nodes in networks with predefined groups. This article is an extensive elaboration of some ideas presented in Rousseau *et al.* (2015). We apply a slightly other notation than in our previous articles on this topic. This new notation corresponds better with the case of directed networks.

2 Background

2.1 Networks and centrality

We assume that we have a directed network $\mathcal{N}=(V, E)$, consisting of a set of nodes or vertices (V) and a set of links, arcs, or edges (E). Each link $(u, v) \in E$ is a connection from u to v ($u, v \in V$). The number of shortest paths or geodesics from node g to h (in that order) is denoted as $p_{g,h}$. The number of geodesics from g to h that pass through a node a ($a \neq g$, $a \neq h$, so a is not an endpoint) is denoted as $p_{g,h}(a)$. Of course, in an undirected network, $p_{g,h}=p_{h,g}$, and similarly $p_{g,h}(a)=p_{h,g}(a)$, while this is usually not the case in a directed network.

Networks can be characterized by several different measures. Centrality measures are indicators that characterize the importance of individual nodes. The most important ones are degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality, referred to as rank prestige by Wasserman and Faust (1994). We focus on betweenness centrality because of its importance to the following discussion. Betweenness centrality characterizes a node's control over the geodesic information flow through the network. The betweenness centrality of node a, denoted as $C_{\mathrm{B}}(a)$, is defined as

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

where we assume that g, h, and a are three different nodes. From now on we will use the notation $g \neq h \neq a$, which should be understood as $g \neq h$, $g \neq a$, and $h \neq a$. By convention, we set 0/0=0. In this way, the formula of betweenness centrality and subsequent formulae in this paper can also be applied to unconnected networks (Freeman, 1977). This convention is equivalent to considering only those node pairs (g, h) where h is reachable from g.

Since we treat undirected networks as symmetric directed networks, each path in an undirected network is counted twice. For instance, one would normally claim that the undirected network in Fig. 1 contains one geodesic between b and d (b-c-d), whereas the symmetric directed interpretation yields two geodesics (b-c-d and d-c-b). Hence, when applying Eq. (1) to an undirected network, one should divide the outcome by two.

The maximum value of $C_{\mathrm{B}}(a)$ in Eq. (1) depends on the size of the network. If one wants to compare values between nodes from different networks, normalization to values between 0 and 1 can be applied. For a network with N nodes, this becomes

$$C_{\mathrm{B}}^{N}(a)=\frac{1}{(N-1)(N-2)}\sum_{g,h\in V}\frac{p_{g,h}(a)}{p_{g,h}}. \qquad (2)$$

The highest value reached by normalized betweenness centrality is by definition the value one. It is obtained by the center of a star network.

2.2 Brokerage between groups in a network

In some cases, a network's nodes may belong to different groups. We assume that node groups are known (which clearly separates our work from, for instance, research on community finding algorithms). Brokerage can informally be understood as the extent to which a node facilitates information exchange between other nodes, especially nodes that belong to different groups. Brokerage between disjoint groups in networks has previously been studied by Gould and Fernandez (1989), who studied subnetworks of the form *a*-*x*-*b*, where *a*, *x*, and *b* are nodes. Depending on the question whether or not these nodes belong to the same or different groups, they distinguished between five different brokerage roles for *x*.

Gould and Fernandez (1989) considered only situations where the 'outer' nodes *a* and *b* are directly linked to the bridging node *x*. It is, however, well known that information in a network may travel through several intermediary nodes. Hence, one can also imagine indicators of brokerage that account for situations where the bridging node is not necessarily directly linked to the outer nodes. This is the case for the gefura measures studied in this paper.

When it comes to brokerage, bridging, or gatekeepership, Burt's theory of social capital and structural holes (Burt, 2004) should be mentioned. Structural holes in a social network are described as disconnected or poorly connected areas between otherwise densely connected groups of people. Brokers play the role of connectors between two or more poorly connected groups (Fig. 2) and are rewarded for this by an increase in social capital. Burt's theory applies only to social networks and does not start from predefined groups; instead, groups are determined by the network structure.

A third line of research pertaining to brokerage was initiated by Flom *et al*. (2004) and later elaborated in work of the authors and colleagues. This research has introduced a 'family' of measures, which are different from the work of Gould and Fernandez and/or Burt in several ways:

1. No assumptions on the type of network need to be made (connected or disconnected, directed or undirected).
2. The measures are agnostic to setting; i.e., they can also be applied outside a social network context.
3. The measures quantify a node's bridging role even if it is not directly linked to the outer nodes.
4. Groups are well-defined from the outset and brokers belong, in principle, to one of these groups (although we do not exclude the case that a broker is a singleton group on its own).

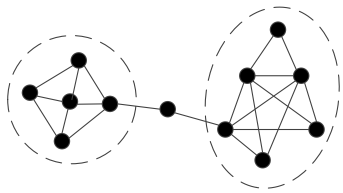

Fig. 2 Example of brokerage according to Burt

Flom *et al.* (2004) began this line of research by introducing 'Q-measures' as an indicator of brokerage in an undirected network where nodes belong to one of two groups (e.g., males and females). Q-measures are essentially a variant of betweenness centrality that considers only geodesics between nodes from different groups. Assume that there are T actors (nodes) in the network. Group G contains m nodes, while the other group, denoted as H, contains n nodes; hence, $T=m+n$. If actor a belongs to group G, and assuming for simplicity that actor a is g_m, then the Q-measure for this actor, $Q(a)$, is defined as

$$Q(a)=\frac{1}{(m-1)n}\sum_{i=1}^{m-1}\sum_{j=1}^{n}\frac{p_{g_i,h_j}(a)}{p_{g_i,h_j}}. \qquad (3)$$

Here, p_{g_i,h_j} denotes the number of shortest paths from the ith to the jth node, while $p_{g_i,h_j}(a)$ is the number of shortest paths from the ith to the jth node that pass through node a.

Using a slightly simpler notation, the same formula can be expressed as follows:

$$Q(a)=\frac{1}{(m-1)n}\sum_{\substack{g\in G\\ h\in H}}\frac{p_{g,h}(a)}{p_{g,h}}, \quad g\neq h\neq a. \qquad (4)$$

Q-measures were applied in Rousseau (2005), Chen and Rousseau (2008), and Zhang *et al.* (2009). Whereas the initial definition applied to connected, unweighted, and undirected networks, Rousseau and Zhang (2008) considered Q-measures in the context of weighted directed networks and proposed that in this case the definition should be based on flow betweenness centrality (Freeman *et al.*, 1991) rather than shortest-path betweenness centrality as defined in Eq. (1).

The main restriction up to this point was that the definition allowed for only two groups, even though many real-world networks have three or more node groups. This restriction was removed by Guns and Rousseau (2009), who generalized the Q-measure definition to any finite number of groups and introduced the distinction between 'global' Q-measures (brokerage between all different groups) and 'local' Q-measures (brokerage between one's own group and the other groups). The new definitions were applied to international co-authorship networks (Guns and Liu, 2010; Guns *et al.*, 2011). Rousseau *et al*. (2013; 2014) focused on mathematical properties of Q-measures, including a convex decomposition, relations with betweenness centrality, and a characterization of nodes with a normalized Q-measure equal to one (the highest possible value). In Liu *et al.* (2013) the authors used a binary tree as a model for a hierarchical structure. All nodes situated at the same height were assumed to form a group. Q-measure values were determined for each node in a binary tree of arbitrary height.

We note that networks for which node sets consist of disjoint subsets can be considered as a special type of multilayer networks, as described in Boccaletti *et al.* (2014).

A note on naming: Colleagues working in social network theory and marketing observed that the term 'Q-measure' is non-descriptive and, moreover, is used in other contexts, such as Tobin's Q in marketing (Brainard and Tobin, 1968), Q-analysis as a mathematical technique to study and analyze structures (Atkin, 1972), and Q-measure as an information retrieval metric in a graded relevance context (Sakai, 2007). The best known Q in network theory is probably Newman and Girvan (2004)'s Q denoting modularity in a network. So, indeed, the use of the symbol Q and the term Q-measure to study brokerage is not optimal at all.

As these measures gauge the bridging role of nodes, the term 'gefura measure' (after old Greek γεφυρα), meaning bridge measure, might be a more descriptive term with universal appeal. From now on the authors intend to use the symbol Γ (capital gamma: the first letter of gefura) instead of Q.

3 Unnormalized gefura measures in directed networks

Most of the definitions in previous studies of gefura or Q-measures are normalized to the interval [0, 1]. This is, for instance, the case in Eq. (3), where the division by $(m-1)n$ ensures a maximum value of one. Guns and Rousseau (2009) noted that "one may imagine circumstances, e.g., real-world transportation networks, where normalization, leading to a relative measure, is not optimal. Then no division […] is performed. This yields an absolute measure of 'bridgeness'." More specifically, normalization makes it possible to compare gefura measures across networks of different sizes. If this is not the purpose, then normalizing brings no real advantage. We therefore start the discussion by considering gefura measures without normalization, the most straightforward forms.

In practice, many networks are directed; i.e., their links have an inherent direction. Examples include citation networks, telephone networks (*a* calls *b*), and any linear order or any hierarchy. Guns and Rousseau (2009) wrote that "all formulae for Q-measures—the original Q-measures for two groups as well as Q-measures for any finite number of groups—can be applied to directed networks as well." Nevertheless, gefura measures (or Q-measures) in directed networks have so far been studied explicitly only by Rousseau and Zhang (2008), who proposed a definition for Γ-measures in directed, weighted networks (although they pay more attention to the question 'how to deal with link weights'). This section and the following ones therefore assume that we are working in the setting of directed networks, unless explicitly stated otherwise. The undirected case can easily be derived by treating the undirected network as a symmetric directed network, in which each undirected link is replaced by two links between the same node pair but pointing in opposite directions.

If we have a network with three or more groups, the basic procedure remains the same: the gefura measure of node a is calculated by determining the proportion of geodesics between nodes from different groups that pass through a. In this case, one might pose the question exactly which groups are taken into account. That is, one does not necessarily need to consider geodesics between nodes from all groups but could restrict oneself to geodesics between nodes from a subset of groups. While specific situations may warrant including or excluding specific groups, we think that three situations are potentially relevant to any network divided into more than two groups (Rousseau *et al.*, 2013):

1. Global: all group pairs are taken into account.

2. Local: all pairs between a node's own group and other groups are taken into account.

3. External: all group pairs, except those involving a node's own group, are taken into account.

Before moving on to the definitions, we introduce some notations and assumptions. Consider a network subdivided into S non-overlapping groups ($1<S<+\infty$). Each group is denoted as G_i (i=1, 2, …, S) and contains m_i members. It will henceforth be assumed that node a, the node for which we want to calculate a gefura measure, belongs to group G_d ($1\leq d\leq S$). The group to which a given node, say g, belongs, is denoted as group(g).

The global gefura measure is then defined as follows:

$$\Gamma_{\mathrm{G}}(a)=\sum_{\substack{g,h\in V\\ \mathrm{group}(g)\neq\mathrm{group}(h)}}\frac{p_{g,h}(a)}{p_{g,h}}. \tag{5}$$

The only difference between this formula and unnormalized betweenness centrality is the restriction that g and h belong to different groups.

For local gefura measures we essentially have one 'special' group (a node's own group, here called G_d), and consider shortest paths between this group and all the other groups. In a directed network we can distinguish between outward geodesics that originate from G_d and end in another group and inward geodesics that originate from another group and end in G_d (Fig. 3). If a_1 is the node under study, the thick line indicates an outward geodesic from a_3 to b_3, while the dashed line indicates an inward geodesic from b_3 to a_3. Node a_1 is part of the outward geodesic, but not of the inward one. This is an important distinction. For instance, in an author citation network with countries as groups, it makes a profound difference whether one forms an outward bridge (citing authors from other countries) or an inward one (receiving international citations). Hence, we propose that one can distinguish between three cases: inward, outward, and both. We will denote these as $\Gamma_{\mathrm{L(i)}}$, $\Gamma_{\mathrm{L(o)}}$, and $\Gamma_{\mathrm{L(b)}}$, respectively. They are defined as follows:

$$\Gamma_{\mathrm{L(i)}}(a)=\sum_{\substack{g\in G_d\\ h\notin G_d}}\frac{p_{h,g}(a)}{p_{h,g}}, \tag{6}$$

$$\Gamma_{\mathrm{L(o)}}(a)=\sum_{\substack{g\in G_d\\ h\notin G_d}}\frac{p_{g,h}(a)}{p_{g,h}}, \tag{7}$$

$$\begin{aligned}\Gamma_{\mathrm{L(b)}}(a)&=\sum_{\substack{g\in G_d\\ h\notin G_d}}\left(\frac{p_{g,h}(a)}{p_{g,h}}+\frac{p_{h,g}(a)}{p_{h,g}}\right)\\&=\Gamma_{\mathrm{L(i)}}(a)+\Gamma_{\mathrm{L(o)}}(a).\end{aligned} \tag{8}$$

Here, node g always belongs to the same group as a. In an undirected network the distinction between inward and outward geodesics does not hold and $\Gamma_{\mathrm{L(i)}}(a)=\Gamma_{\mathrm{L(o)}}(a)$. For undirected networks, we will simply use $\Gamma_{\mathrm{L}}(a)$ for $\Gamma_{\mathrm{L(b)}}(a)$.

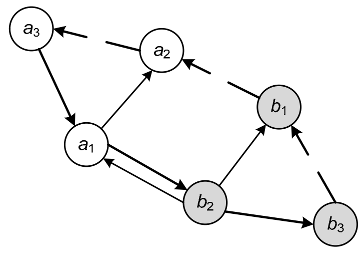

Fig. 3 Inward and outward shortest paths in a directed network

There are two node groups, $\{a_1, a_2, a_3\}$ and $\{b_1, b_2, b_3\}$

If there are only two groups (S=2), then $\Gamma_{\mathrm{L}}(a)=\Gamma_{\mathrm{G}}(a)$.

Finally, the external gefura measure is defined as

$$\Gamma_{\mathrm{E}}(a)=\sum_{j\neq k\neq d}\sum_{g\in G_j,h\in G_k}\frac{p_{g,h}(a)}{p_{g,h}}. \tag{9}$$

Note that the first summation considers all possible j and k (both different from d). Hence, a given group once plays the role of G_j and once the role of G_k. If S=2, the external gefura measure is equal to zero.

One can easily see that there is no overlap between the group pairs on which $\Gamma_{L(b)}$ and Γ_E are based. Furthermore, the group pairs for Γ_G are a disjoint union of those used to calculate $\Gamma_{L(b)}$ and Γ_E. More specifically, the relationship between global, local, and external gefura measures is as follows:

$$\Gamma_G=\Gamma_{L(b)}+\Gamma_E. \tag{10}$$

4 Normalized gefura measures in directed networks

The next step is to define the same measures, global, local, and external gefura, with normalization. Depending on the relative importance one gives to groups or nodes we come up with two definitions.

4.1 Structural normalization

The structural global gefura measure, denoted as Γ_G^S, is defined as follows (Guns and Rousseau, 2009):

$$\begin{aligned}\Gamma_G^S(a)&=\frac{1}{S(S-1)}\sum_{k,l}\left(\frac{1}{P_{k,l}}\sum_{\substack{g\in G_k\\ h\in G_l}}\frac{p_{g,h}(a)}{p_{g,h}}\right)\\&=\frac{1}{S(S-1)}\sum_{k,l}\Gamma_{k,l}(a).\end{aligned} \tag{11}$$

The symbol $P_{k,l}$ refers to the number of possible combinations of elements belonging to groups G_k and G_l. Hence, $P_{k,l}=|G_k\backslash\{a\}|\cdot|G_l\backslash\{a\}|$, where $|\cdot|$ refers to the number of elements in the set. The part between $|\cdot|$ is essentially a normalized gefura measure for two groups G_k and G_l (see Eq. (4)); this can be called a partial gefura measure and is denoted as $\Gamma_{k,l}$ (note that, in general, $\Gamma_{k,l}(a)\neq\Gamma_{l,k}(a)$). Two normalizations are applied here: first, an 'inner' normalization in the partial gefura measure, and second, an 'outer' normalization for the number of groups. In the special case $P_{k,l}$=0, which occurs if k or l equals d and $G_d=\{a\}$, we apply the rule 0/0=0.

Likewise, one can define a structural local gefura measure. Again, we distinguish between inward, outward, and both:

$$\Gamma_{L(i)}^S(a)=\frac{1}{S-1}\sum_{l\neq d}\left(\frac{1}{P_{d,l}}\sum_{\substack{g\in G_d\\ h\in G_l}}\frac{p_{h,g}(a)}{p_{h,g}}\right), \tag{12}$$

$$\Gamma_{L(o)}^S(a)=\frac{1}{S-1}\sum_{l\neq d}\left(\frac{1}{P_{d,l}}\sum_{\substack{g\in G_d\\ h\in G_l}}\frac{p_{g,h}(a)}{p_{g,h}}\right), \tag{13}$$

$$\begin{aligned}\Gamma_{L(b)}^S(a)&=\frac{1}{2(S-1)}\sum_{l\neq d}\left(\frac{1}{P_{d,l}}\sum_{\substack{g\in G_d\\ h\in G_l}}\left(\frac{p_{g,h}(a)}{p_{g,h}}+\frac{p_{h,g}(a)}{p_{h,g}}\right)\right)\\&=\frac{\Gamma_{L(i)}^S(a)+\Gamma_{L(o)}^S(a)}{2}.\end{aligned} \tag{14}$$

In the case of an undirected network, we have (Guns and Rousseau, 2009)

$$\begin{aligned}\Gamma_L^S(a)&=\frac{1}{2(S-1)}\sum_{l\neq d}\left(\frac{1}{P_{d,l}}\sum_{\substack{g\in G_d\\ h\in G_l}}\frac{p_{g,h}(a)}{p_{g,h}}\right)\\&=\frac{1}{2(S-1)}\sum_{l\neq d}\left(\Gamma_{d,l}+\Gamma_{l,d}\right)\\&=\frac{1}{S-1}\sum_{l\neq d}\Gamma_{d,l}.\end{aligned} \tag{15}$$

Also, note that here the global and local measures coincide when S=2. Indeed,

$$\begin{aligned}\Gamma_G^2(a)&=\frac{1}{2(2-1)}\left(\Gamma_{d,l}(a)+\Gamma_{l,d}(a)\right)\\&=\frac{2}{2}\Gamma_{d,l}(a)=\Gamma_L^2(a).\end{aligned}$$

Finally, we define the structural external gefura measure Γ_E^S (Rousseau *et al.*, 2013):

$$\Gamma_E^S(a)=\frac{1}{(S-1)(S-2)}\sum_{\substack{k,l\\ k\neq l\neq d}}\Gamma_{k,l}(a). \tag{16}$$

Rousseau *et al.* (2013; 2014) proved the following convex decomposition of the global gefura measure:

$$\Gamma_{\mathrm{G}}^{\mathrm{S}}(a)=\frac{2}{S}\Gamma_{\mathrm{L}}^{\mathrm{S}}(a)+\frac{S-2}{S}\Gamma_{\mathrm{E}}^{\mathrm{S}}(a). \quad (17)$$

This relation still holds for a directed network.

4.2 Basic normalization

We have seen how structural normalization involves an 'inner' and an 'outer' normalization. An alternative starts from the unnormalized definitions and uses just one normalization step. As this procedure is somewhat simpler, we call it basic normalization.

The basic global gefura measure $\Gamma_{\mathrm{G}}^{\mathrm{B}}$ is defined as follows:

$$\Gamma_{\mathrm{G}}^{\mathrm{B}}(a)=\frac{1}{M_a}\sum_{\substack{g\in G_k\\ h\in G_l\\ k\neq l}}\left(\frac{p_{g,h}(a)}{p_{g,h}}\right), \quad (18)$$

where the symbol M_a is defined as

$$M_a=\sum_{k,l}|G_k\setminus\{a\}|\cdot|G_l\setminus\{a\}|.$$

Note that this is simply the unnormalized global gefura measure (5) divided by a maximum, to obtain a value between zero and one.

The basic normalization of local gefura measures also applies to the inward, outward, and both cases:

$$\Gamma_{\mathrm{L(i)}}^{\mathrm{B}}(a)=\frac{1}{(m_d-1)(N-m_d)}\sum_{\substack{g\in G_d\\ h\notin G_d}}\frac{p_{h,g}(a)}{p_{h,g}}, \quad (19)$$

$$\Gamma_{\mathrm{L(o)}}^{\mathrm{B}}(a)=\frac{1}{(m_d-1)(N-m_d)}\sum_{\substack{g\in G_d\\ h\notin G_d}}\frac{p_{g,h}(a)}{p_{g,h}}, \quad (20)$$

$$\Gamma_{\mathrm{L(b)}}^{\mathrm{B}}(a)=\frac{1}{2(m_d-1)(N-m_d)}\sum_{\substack{g\in G_d\\ h\notin G_d}}\left(\frac{p_{h,g}(a)}{p_{h,g}}+\frac{p_{g,h}(a)}{p_{g,h}}\right)$$
$$=\frac{\Gamma_{\mathrm{L(i)}}^{\mathrm{B}}(a)+\Gamma_{\mathrm{L(o)}}^{\mathrm{B}}(a)}{2}. \quad (21)$$

In the case of undirected networks, this can be simplified to

$$\Gamma_{\mathrm{L}}^{\mathrm{B}}(a)=\frac{1}{2(m_d-1)(N-m_d)}\sum_{\substack{g\in G_d\\ h\notin G_d}}\left(\frac{p_{h,g}(a)}{p_{h,g}}+\frac{p_{g,h}(a)}{p_{g,h}}\right)$$
$$=\frac{1}{(m_d-1)(N-m_d)}\sum_{\substack{g\in G_d\\ h\notin G_d}}\frac{p_{g,h}(a)}{p_{g,h}}. \quad (22)$$

In all these cases it might be necessary to apply the rule 0/0=0. The basic external gefura measure $\Gamma_{\mathrm{G}}^{\mathrm{B}}$ is defined as

$$\Gamma_{\mathrm{E}}^{\mathrm{B}}(a)=\frac{1}{M_{a,0}}\sum_{\substack{g\in G_k\\ h\in G_l\\ k\neq l\neq d}}\frac{p_{g,h}(a)}{p_{g,h}}, \quad (23)$$

where $M_{a,0}=\sum_{k\neq l\neq d}m_km_l$. Note that each product of the form m_km_l occurs also as m_lm_k in this sum.

Similar to the structural gefura measures, a convex decomposition of the global measure can be obtained:

$$\Gamma_{\mathrm{G}}^{\mathrm{B}}(a)=X\Gamma_{\mathrm{L}}^{\mathrm{B}}(a)+Y\Gamma_{\mathrm{E}}^{\mathrm{B}}(a), \quad (24)$$

where

$$X=\frac{(m_d-1)(N-m_d)}{(m_d-1)(N-m_d)+\sum\limits_{\substack{k,l\\ k\neq l\neq d}}m_km_l}, \quad (25)$$

$$Y=\frac{\sum\limits_{k\neq l\neq d}m_km_l}{(m_d-1)(N-m_d)+\sum\limits_{\substack{k,l\\ k\neq l\neq d}}m_km_l}=\frac{M_{a,0}}{M_a}. \quad (26)$$

4.3 Comparison of structural and basic gefura measures

The difference between structural and basic gefura measures can be clarified by rewriting the former to be more similar to the latter. We will use the structural global gefura measure here, but a similar reasoning applies to the local and external measures. For simplicity, we will discuss this within the setting of an undirected network, but the reasoning is similar for directed networks. Eq. (11) can be written as follows:

$$\Gamma_{\mathrm{G}}^{\mathrm{S}}(a)=\frac{1}{M_a}\sum_{\substack{j,k\\ j\neq k}}\sum_{g\in G_j,h\in G_k}w_{j,k,a}\frac{p_{g,h}(a)}{p_{g,h}}, \quad (27)$$

where

$$w_{j,k,a}=\frac{M_a}{S(S-1)}\frac{1}{|G_j\setminus\{a\}|\cdot|G_k\setminus\{a\}|}. \quad (28)$$

Whereas the basic normalization or the unnormalized form treats all geodesics the same, the above equations show that structural normalization is a sort of weighted form of basic normalization. More specifically, the ratio $p_{g,h}(a)/p_{g,h}$ is multiplied with a factor $w_{j,k,a}$, such that more weight is given to geodesics between smaller groups. The result is that structural normalization treats a small group as equal to a large one, whereas basic normalization gives equal weight to each geodesic regardless of the group size.

A few examples may help to clarify this difference. Consider example network A in Fig. 4, where nodes are grouped according to their initial letter: a_1 and a_2 form one group; b_1, b_2, and b_3 form another; and so on. It can easily be seen that only nodes a_1, b_1, and c_1 have Γ>0. We find that $\Gamma_G^S(a_1)=\Gamma_G^S(b_1)=\Gamma_G^S(c_1)=2/3$, whereas $\Gamma_G^B(a_1)=0.37<\Gamma_G^B(b_1)=0.6<\Gamma_G^B(c_1)=0.71$. The values of the structural gefura measure are the same, because the three nodes bridge between the same number of groups. However, c_1's basic gefura measure is the highest because this node bridges between more node pairs than the other two, and likewise, b_1 obtains a higher basic gefura value than a_1. This illustrates how structural normalization normalizes at the group level, while basic normalization works at the node level.

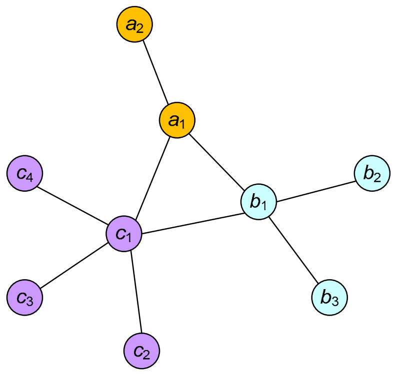

Fig. 4 Example network *A*

Using the same underlying network but a different group partitioning (Fig. 5), we have $\Gamma_G^S(c_1)>\Gamma_G^S(b_1)=\Gamma_G^S(a_1)$ and $\Gamma_G^B(a_1)<\Gamma_G^B(b_1)<\Gamma_G^B(c_1)$. The difference is that c_1 is now also a bridge between the a- and d-group, and between the b- and d-group.

In conclusion, structural normalization attaches more importance to the group level (answering the question: for how many group pairs is the node a broker?), whereas basic normalization attaches more importance to the node level (answering the question: for how many pairs of nodes from different groups is the node a broker?).

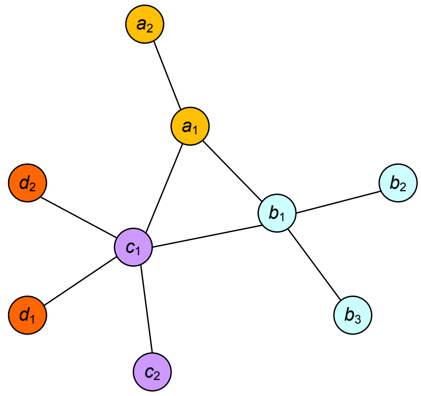

Fig. 5 Example network *B*

5 An adapation of the Brandes algorithm

Brandes (2001) introduced a fast algorithm to determine betweenness centrality. It works in $O(nm)$ time for unweighted networks and $O(n+m)$ space, where n denotes the number of nodes and m the number of links. In a follow-up article, Brandes (2008) showed how the basic algorithm can be adapted to calculate many variants of betweenness centrality, including gefura measures for two groups (with the same asymptotic time complexity). Essentially the same algorithm can be used to calculate (unnormalized or basic) global gefura measures. Structural gefura measures cannot be determined this way and we currently lack an efficient algorithm to determine them. This is a practical disadvantage of structural gefura.

We reuse the notation of Brandes (2008). The algorithm considers each node once; this node is referred to as the 'source'. Here, S is a stack of nodes, Pred[v] is a list of predecessors on the shortest path from the source to $v \in V$, $\sigma[v]$ is the number of geodesics from the source to v, and $\delta[v]$ is the dependency of the source on v. That is, if we denote the source by

g, then $\delta[v]=\sum_{h\in V}(p_{g,h}(v)/p_{g,h})$. The crucial accumulation step then becomes:

Accumulation:
{
for $v\in V$ **do** $\delta[v]\leftarrow 0$
  while S not empty **do**
    pop $w\leftarrow S$
    if group(s)≠group(w) **then** $i\leftarrow 1$ **else** $i\leftarrow 0$
    for $v\in$Pred[w] **do**

$$\delta[v]\leftarrow\delta[v]+\frac{\sigma[v]}{\sigma[w]}\left(i+\delta[w]\right)$$

    if $w\neq s$ **then** $\Gamma_G(w)\leftarrow\Gamma_G(w)+\delta[w]$
}

With a small adaptation, the algorithm can be used as follows to determine local gefura:

Accumulation:
{
for $v\in V$ **do** $\delta[v]\leftarrow 0$
  while S not empty **do**
    pop $w\leftarrow S$
    if group(s)≠group(w) **then** $i\leftarrow 1$ **else** $i\leftarrow 0$
    for $v\in$Pred[w] **do**

$$\delta[v]\leftarrow\delta[v]+\frac{\sigma[v]}{\sigma[w]}\left(i+\delta[w]\right)$$

    if $w\neq s$ **and** group(s)=group(w) **then**
      $\Gamma_L(w)\leftarrow\Gamma_L(w)+\delta[w]$
}

The above algorithms apply to both undirected and directed networks. Here, undirected networks are treated like symmetric directed networks as well. The algorithm for local gefura yields the inward local gefura measure when applied to a directed network. One can obtain the outward local gefura by reversing the direction of all links prior to using the algorithm.

An implementation of this algorithm in Python can be found at https://github.com/rafguns/gefura.

6 Conclusions

In this contribution we studied partitioned networks, i.e., networks subdivided into non-overlapping, non-empty subgroups, which together form the complete node set of the network. Measures specially adapted to this structure used to be called Q-measures, but we explained why this was not a good terminological choice, preferring the name 'gefura measures' instead. These measures were designed as indicators of a node's brokerage role between such groups. We started by a systematic study of unnormalized gefura measures in directed networks and showed that, next to the 'structural' normalization that has hitherto been applied, a 'basic' normalization procedure is possible. While the former normalizes at the level of groups, the latter normalizes at the level of nodes. The undirected case follows straightforwardly by treating undirected networks as symmetric directed networks. Finally, Brandes' algorithm for betweenness centrality was adjusted to cover most of these cases. This led to the open problem of adapting this algorithm to calculate gefura measures in their structural normalization form. We observed that the above-mentioned partition refers to nodes. Links may connect nodes belonging to the same group or may connect nodes belonging to different nodes.

We are convinced that gefura measures constitute a useful tool for social scientists, informetricians, and colleagues studying complex networks, in the case the studied networks are partitioned into subgroups.

References

Atkin, R.H., 1972. From cohomology in physics to *q*-connectivity in social science. *Int. J. Man-Mach. Stud.*, **4**(2):139-167. [doi:10.1016/S0020-7373(72)80029-4]

Barrat, A., Barthélemy, M., Pastor-Satorras, R., *et al.*, 2004. The architecture of complex weighted networks. *PNAS*, **101**(11):3747-3752. [doi:10.1073/pnas.0400087101]

Boccaletti, S., Bianconi, G., Criado, R., *et al.*, 2014. The structure and dynamics of multilayer networks. *Phys. Rep.*, **544**(1):1-122. [doi:10.1016/j.physrep.2014.07.001]

Brainard, W.C., Tobin, J., 1968. Econometric models: their problems and usefulness. Pitfalls in financial model building. *Amer. Econ. Rev.*, **58**(2):99-122.

Brandes, U., 2001. A faster algorithm for betweenness centrality. *J. Math. Sociol.*, **25**(2):163-177. [doi:10.1080/0022250X.2001.9990249]

Brandes, U., 2008. On variants of shortest-path betweenness centrality and their generic computation. *Soc. Netw.*, **30**(2):136-145. [doi:10.1016/j.socnet.2007.11.001]

Burt, R.S., 2004. Structural holes and good ideas. *Amer. J. Sociol.*, **110**(2):349-396. [doi:10.1086/421787]

Chen, L.X., Rousseau, R., 2008. Q-measures for binary divided networks: bridges between German and English institutes in publications of the *J. Fluid Mech. Scientometr.*, **74**(1):57-69. [doi:10.1007/s11192-008-0103-6]

Christensen, C., Albert, R., 2007. Using graph concepts to understand the organization of complex systems. *Int. J. Bifurc. Chaos*, **17**(7):2201-2214. [doi:10.1142/S0218127 40701835X]

Ding, Y., 2011. Scientific collaboration and endorsement: network analysis of coauthorship and citation networks. *J. Inform.*, **5**(1):187-203. [doi:10.1016/j.joi.2010.10.008]

Flom, P.L., Friedman, S.R., Strauss, S., *et al.*, 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 on betweenness. *Sociometry*, **40**(1):35-41.

Freeman, L.C., Borgatti, S.P., White, D.R., 1991. Centrality in valued graphs: a measure of betweenness based on network flow. *Soc. Netw.*, **13**(2):141-154. [doi:10.1016/ 0378-8733(91)90017-N]

Gould, R.V., Fernandez, R.M., 1989. Structures of mediation: a formal approach to brokerage in transaction networks. *Sociol. Method.*, **19**:89-126. [doi:10.2307/270949]

Guimerà, R., Amaral, L.A.N., 2005. Functional cartography of complex metabolic networks. *Nature*, **433**:895-900. [doi:10.1038/nature03288]

Guimerà, R., Mossa, S., Turtschi, A., *et al.*, 2005. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *PNAS*, **102**(22):7794-7799. [doi:10.1073/pnas.040799 4102]

Guns, R., Liu, Y.X., 2010. Scientometric research in China in context of international collaboration. Proc. 6th Int. Conf. on Scientometrics and University Evaluation, p.112-115.

Guns, R., Rousseau, R., 2009. Gauging the bridging function of nodes in a network: Q-measures for networks with a finite number of subgroups. Proc. 12th ISSI, p.131-142.

Guns, R., Liu, Y.X., Mahbuba, D., 2011. Q-measures and betweenness centrality in a collaboration network: a case study of the field of informetrics. *Scientometrics*, **87**(1):133-147. [doi:10.1007/s11192-010-0332-3]

Liu, Y.X., Guns, R., Rousseau, R., 2013. A binary tree as a basic model for studying hierarchies using Q-measures. *SRELS J. Inform. Manag.*, **50**(5):521-528.

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. *Phys. Rev. E*, **69**: 026113.1-026113.15. [doi:10.1103/PhysRevE.69.026113]

Otte, E., Rousseau, R., 2002. Social network analysis: a powerful strategy, also for the information sciences. *J. Inform. Sci.*, **28**(6):441-453. [doi:10.1177/0165551502028 00601]

Rousseau, R., 2005. Q-measures for binary divided networks: an investigation within the field of informetrics. *Proc. Amer. Soc. Inform. Sci. Technol.*, **42**(1):675-696.

Rousseau, R., Zhang, L., 2008. Betweenness centrality and Q-measures in directed valued networks. *Scientometrics*, **75**(3):575-590. [doi:10.1007/s11192-007-1772-2]

Rousseau, R., Liu, Y.X., Guns, R., 2013. Mathematical properties of Q-measures. *J. Inform.*, **7**(3):737-745. [doi:10. 1016/j.joi.2013.06.002]

Rousseau, R., Liu, Y.X., Guns, R., 2014. An addendum and correction to "Mathematical properties of Q-measures" (vol. 7, issue 3, pp.737-745). *J. Inform.*, **8**(3):486-490. [doi:10.1016/j.joi.2014.01.004]

Rousseau, R., Guns, R., Liu, Y.X., 2015. Gauging the bridging function of nodes in a network: the gefura measure. Proc. 8th Int. Conf. on Scientometrics and University Evaluation, in press.

Sakai, T., 2007. On the reliability of information retrieval metrics based on graded relevance. *Inform. Process. Manag.*, **43**(2):531-548. [doi:10.1016/j.ipm.2006.07.020]

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, UK.

Zhang, W.L., Yin, L.C., Pang, J., 2009. The application of Q-measure to gender study in cooperation network. *Sci. Technol. Progr. Pol.*, **26**(15):100-103 (in Chinese).