Detecting global bridges in networks

Journal of Complex Networks Advance Access published October 29, 2015

Journal of Complex Networks (2015) Page 1 of 11
doi:10.1093/comnet/cnv022

Detecting global bridges in networks

Pablo Jensen[†]
IXXI, Institut Rhonalpin des Systèmes Complexes, ENS Lyon, Lyon, France and Laboratoire de Physique, UMR 5672, ENS Lyon, 69364 Lyon, France
[†]Corresponding author. Email: pablo.jensen@ens-lyon.fr

Matteo Morini and Márton Karsai
IXXI, Institut Rhonalpin des Systèmes Complexes, ENS Lyon, Lyon, France and LIP, INRIA, UMR 5668, ENS de Lyon, 69364 Lyon, France

Tommaso Venturini
Médialab, Sciences Po, Paris, France

Alessandro Vespignani
MoBS, Northeastern University, Boston, MA 02115, USA and ISI Foundation, Turin 10133, Italy

Mathieu Jacomy
Médialab, Sciences Po, Paris, France

Jean-Philippe Cointet
Université Paris-Est, SenS-IFRIS, Paris, France

Pierre Mercklé
Centre Max Weber, UMR 5283, ENS Lyon, 69364 Lyon, France

and

Eric Fleury
IXXI, Institut Rhonalpin des Systèmes Complexes, ENS Lyon, Lyon, France and LIP, INRIA, UMR 5668, ENS de Lyon, 69364 Lyon, France

Edited by: Ernesto Estrada

[Received on 1 June 2015; accepted on 25 August 2015]

The identification of nodes occupying important positions in a network structure is crucial for the understanding of the associated real-world system. Usually, betweenness centrality (BC) is used to evaluate a node capacity to connect different graph regions. However, we argue here that this measure is not adapted for that task, as it gives equal weight to 'local' centres (i.e. nodes of high-degree central to a single region) and to 'global' bridges, which connect different communities. This distinction is important as the roles of such nodes are different in terms of the local and global organization of the network structure. In this paper, we propose a decomposition of BC into two terms, one highlighting the local contributions and the other the global ones. We call the latter *bridgeness* centrality and show that it is capable to specifically spot out global bridges. In addition, we introduce an effective algorithmic implementation of this measure

and demonstrate its capability to identify global bridges in air transportation and scientific collaboration networks.

Keywords: centrality measures; betweenness centrality; bridgeness centrality.

1. Introduction

Although the history of graphs as scientific objects begins with Euler's [1] famous walk across Königsberg bridges, the notion of 'bridge' has rarely been tackled by network theorists.[1] Among the few articles that took bridges seriously, the most famous is probably Mark Granovetter's paper on The Strength of Weak Ties [2]. Despite the huge influence of this paper, few works have remarked that its most original insights concern precisely the notion of 'bridge' in social networks. Granovetter suggested that there might be a fundamental functional difference between strong and weak ties. While strong ties promote homogeneous and isolated communities, weak ties foster heterogeneity and crossbreeding. Or, to use the old tönnesian cliché, strong ties generate Gemeinshaft, while weak ties generates Gesellshaft [3]. Although Granovetter does realize that bridging is the phenomenon he is looking after, two major difficulties prevented him from a direct operationalization of such concept: 'We have had neither the theory nor the measurement and sampling techniques to move sociometry from the usual small-group level to that of larger structures' (ibidem, p. 1360). Let us start from 'the measurement and sampling techniques'. In order to compute the bridging force of a given node or link, one needs to be able to draw a sufficiently comprehensive graph of the system under investigation. Networks constructed with traditional ego-centred and sampling techniques are too biased to compute bridging forces. Exhaustive graphs of small social groups will not work either, since such groups are, by definition, dominated by bounding relations. Since the essence of bridges is to connect individuals across distant social regions, they can only be computed in large and complete social graphs. Hopeless until a few years ago, such endeavour seems more and more reasonable as digital media spread through society. Thanks to digital traceability it is now possible to draw large and even huge social networks [4–6].

Let us discuss now the second point, the 'theory' needed to measure the bridging force of different edges or nodes.[2] Being able to identify bounding and bridging nodes has a clear interest for any type of network. In social networks, bounding and bridging measures (or 'closure' and 'brokerage', to use Burt's terms [7]) tell us which nodes build social territories and which allow items (ideas, pieces of information, opinions, money…) to travel through them. In scientometrics' networks, these notions tell us which authors define disciplines and paradigms and which breed inter-disciplinarity. In ecological networks, they identify relations, which create specific ecological communities and the ones connecting them to larger habitats.

In all these contexts, it is the very same question that we wish to ask: do nodes or edges reinforce the density of a cluster of nodes (bounding) or do they connect two separated clusters (bridging)? Formulated in this way, the bridging/bounding question seems easy to answer. After having identified the clusters of a network, one should simply observe if a node connects nodes of the same cluster (bounding) or of different clusters (bridging). However, the intra-cluster/inter-cluster approach is both too dependent on the method used to detect communities and flawed by its inherent circular logic: it uses

[1] We refer to the common use of the word 'bridge', and not to the technical meaning in graph theory as 'an edge whose deletion increases its number of connected components'.

[2] In this paper, we will focus on defining the bridgeness of nodes, but our definition can straightforwardly be extended to edges, just as the betweenness of edges is derived from that of nodes.

clustering to define bridging and bounding ties when it is precisely the balance of bridges and bounds that determines clusters. Remark that, far from being a mathematical subtlety, this question is a key problem in social theory. Defining internal (gemeinschaft) and external (gesellschaft) relations by presupposing the existence and the composition of social groups is absurd as groups are themselves defined by social relations.

In this paper, we introduce a measure of bridgeness of nodes that is independent on the community structure and thus escapes this vicious circle, contrary to other proposals [8,9]. Moreover, since the computation of bridgeness is straightforwardly related to that of the usual betweenness, Brandes' algorithm [10] can be used to compute it efficiently.[3] To demonstrate the power of our method and identify nodes acting as local or global bridges, we apply it on a synthetic network and two real ones: the world airport network and a scientometric network.

2. Measuring bridgeness

Identifying important nodes in a network structure is crucial for the understanding of the associated real-world system [12–14], for a review see [15]. The most common measure of centrality of a node for network connections on a global scale is betweenness centrality (BC), which 'measures the extent to which a vertex lies on paths between other vertices' [16,17]. We show in the following that, when trying to identify specifically *global* bridges, BC has some limitations as it assigns the same importance to paths between the immediate neighbours of a node as to paths between further nodes in the network. In other words, BC is built to capture the overall centrality of a node, and is not specific enough to distinguish between two types of centralities: local (centre of a community) and global (bridge between communities). Instead, our measure of bridging is more specific, as it gives a higher score to global bridges. The fact that BC may attribute a higher score to local centres than to global bridges is easy to see in a simple network (Fig. 1). The logics is that a 'star' node with degree k, i.e. a node without links between all its first neighbours (clustering coefficient 0) receives automatically a $\mathrm{BC} = k(k-1)/2$ arising from paths of length 2 connecting the node's first neighbours and crossing the central node. More generally, if there exist nodes with high degree but connected only locally (to nodes of the same community), their betweenness may be of the order of that measured for more globally connected nodes. Consistent with this observation, it is well known that for many networks, BC is highly correlated with degree [18–20]. A recent scientometrics study tried to use BC as 'an indicator of the inter-disciplinarity of journals' but noted that this idea only worked 'in local citation environments and after normalization because otherwise the influence of degree centrality (size) overshadows the BC measure' [21].

To avoid this problem and specifically spot out global centres, we decompose BC into a local and a global term, the latter being called 'bridgeness' centrality. Since we want to distinguish global bridges from local ones, the simplest approach is to discard shortest paths, which either start or end at a node's first neighbours from the summation to compute BC (equation (2.1)). This completely removes the paths that connect two non-connected neighbours for 'star nodes' (see Fig. 1) and greatly diminishes the effect of high degrees, while keeping those paths that connect more distant regions of the network.

More formally in a graph $\mathcal{G} = (V, E)$, where V assigns the set of nodes and E the set of links the definition of the BC for a node $j \in V$ stands as:

$$\mathrm{BC}(j) = \mathrm{Bri}(j) + \mathrm{local}(j), \tag{2.1}$$

[3] We have written a plug-in for Gephi [11] that computes this measure on large graphs. See Supplementary Information for a pseudo-algorithm for both node and edge bridgeness.

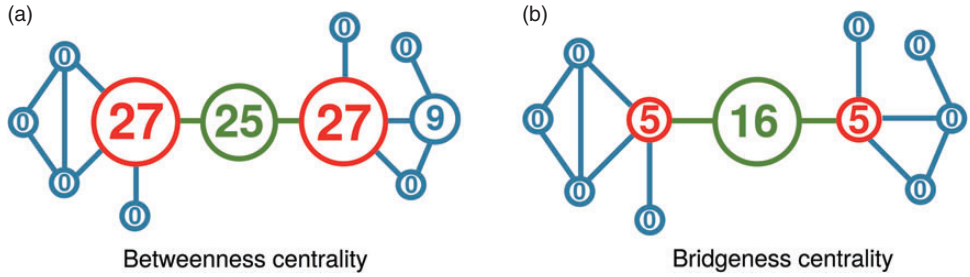

FIG. 1. The figures show the betweenness (a) and bridgeness (b) scores for a simple graph. Betweenness does not distinguish centres from bridges, as it attributes a slightly higher score (a, scores = 27) to high-degree nodes, which are local centres, than to the global bridge (a, score = 25). In contrast, bridgeness rightly spots out the node (b, score = 16) that plays the role of a global bridge.

where

$$\begin{aligned} \mathrm{BC}(j) &= \sum_{i \neq j \neq k} \frac{\sigma_{ik}(j)}{\sigma_{ik}}, \\ \mathrm{Bri}(j) &= \sum_{i \notin N_G(j) \wedge k \notin N_G(j)} \frac{\sigma_{ik}(j)}{\sigma_{ik}}, \\ \mathrm{local}(j) &= \sum_{i \in N_G(j) \vee k \in N_G(j)} \frac{\sigma_{ik}(j)}{\sigma_{ik}}. \end{aligned} \tag{2.2}$$

Here the summation runs over any distinct node pairs i and k; σ_{ik} represents the number of shortest paths between i and k; while $\sigma_{ik}(j)$ is the number of such shortest paths running through j. Decomposing BC into two parts (right-hand side) the first term defines actually the global term, *bridgeness centrality*, where we consider shortest paths between nodes not in the neighbourhood of j ($N_G(j)$), while the second *local term* considers the shortest paths starting or ending in the neighbourhood of j. This definition also demonstrates that the bridgeness centrality value of a node j is always smaller or equal to the corresponding BC value and they only differ by the local contribution of the first neighbours. Figure 1 illustrates the ability of bridgeness to specifically highlight nodes that connect different regions of a graph. Here the BC (Fig. 1(a)) and bridgeness centrality values (Fig. 1(b)) calculated for nodes of the same network demonstrate that bridgeness centrality gives the highest score to the node which is central globally (green), while BC does not distinguish among local or global centres, and actually assigns the highest score to nodes with high degrees (red).

In the following, to further explore the differences between these measures we define an independent reference measure of bridgeness using a known partitioning of the network. This measure provides us an independent ranking of the bridging power of nodes, that we correlate with the corresponding rankings using the BC and bridgeness values. In addition, we demonstrate via three example networks that bridgeness centrality is always more specific than BC to identify global bridges.

3. Computing global bridges from a community structure

To identify the global bridges independently from their score in BC or bridgeness, we use a simple indicator inspired by the well-known Rao–Stirling index [22–25], as this indicator is known to quantify

the ability of nodes to connect different communities. Moreover, it includes the notion of 'distance', which is important for distinguishing local and global connections. However, we note that this index needs as input a prior categorization of the nodes into distinct communities. Our global indicator G in equation (3.1) for node i is defined as:

$$G(i) = \sum_{J \in \text{communities}} l_{IJ} \delta_{i,J}, \tag{3.1}$$

where the sum runs over communities J (different from the community of node i, taken as I), $\delta_{i,J}$ being 1 if there is a link between node i and community J and 0 otherwise. Finally, l_{IJ} corresponds to the 'distance' between communities I and J, as measured by the inverse of the number of links between them: the more links connect two communities, the closer they are. Nodes that are only linked to nodes of their own community have $G = 0$, while nodes that connect two (or more) communities have a strictly positive indicator. Those nodes that bridge distant communities, for example, those that are the only link between two communities, have high G values.

As a next step, we use this reference measure (i.e. the global indicator) to rank nodes and compare it to the rankings obtained by the two tentative characteristics of bridging (BC and bridgeness) in three large networks.

4. Synthetic network: unbiased LFR

We start with a synthetic network (Fig. 2), obtained by a method similar to that of Lancichinetti *et al.* [26]. This method leads to the so-called 'LFR' networks with a clear community structure, which allows to easily identify bridges between communities. We have only modified the algorithm to obtain bridges without the degree bias which arises from the original method. Indeed, LFR first creates unconnected communities and then chooses randomly internal links that are reconnected outside the community. This leads to bridges, i.e. nodes connected to multiple communities, which have a degree distribution biased towards high degrees. In our method, we avoid this bias by randomly choosing nodes, and then one of their internal links, which we reconnect outside its community as in LFR. As reference, we use the global indicator defined above. As explained, this indicator depends on the community structure, which is not too problematic here since, by construction, communities are clearly defined in this synthetic network.

Figure 3(a) shows that bridgeness provides a ranking that is closer to that of the global indicator than BC. Indeed, we observe that the ratio for bridgeness is higher than for BC. This means that ordering nodes by their decreasing bridgeness leads to a better ranking of the 'global' scores—as measured by G—than the corresponding ordering by their decreasing BC values. As shown in the simpler example of a 1000-node network, BC fails because it ranks too high some nodes that have no external connection but have a high degree. A detailed analysis of the nodes of a cluster is given in Supplementary Information.

In addition, we directly measured $\langle locterm \rangle_i(k) = \langle (\mathrm{BC}(i,k) - \mathrm{Bri}(i,k))/\mathrm{BC}(i,k) \rangle_i$, the average relative contribution of the local term in BC for nodes of the same degree (see Fig. 3(b)). We observe a negative correlation, which means that the local term is dominating for low degree nodes, while high-degree nodes have higher bridgeness value as they have a higher chance to connect to different communities.

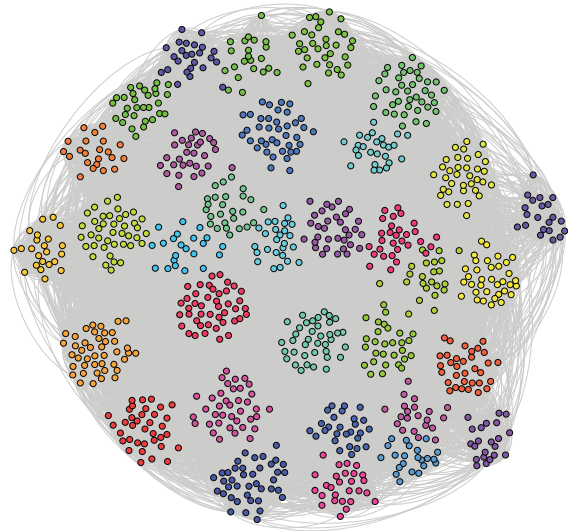

FIG. 2. Artificial network with a clear community structure using Lancichinetti *et al.* [26] method. For clarity, we show here a smaller network containing 1000 nodes, communities, links (20% inter-and 80% intra-community links). Each colour corresponds to a community as detected by modularity optimization [15,27].

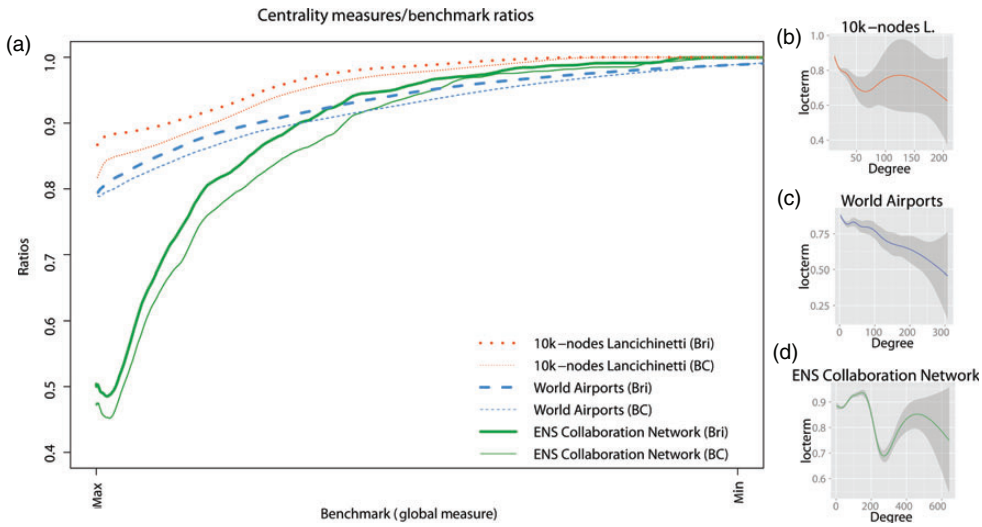

FIG. 3. (a) Ability of BC or bridgeness to reproduce the ranking of bridging nodes, taking as reference the global indicator (equation (2.2)). For each of the three networks, we first compute the cumulative sums for the global measure G, according to three sorting options: the G measure itself and the two centrality metrics, namely BC and bridgeness. By construction, sorting by G leads to the highest possible sum, since we rank the nodes starting by the highest G score and ending by the lowest. Then we test the ability of BC or bridgeness to reproduce the ranking of bridging nodes by computing the respective ratios of their cumulative sum, ranking by the respective metric (BC or Bri), to the cumulative obtained by the G ranking. A perfect match would therefore lead to a ratio equal to 1. Since we observe that the ratio for bridgeness is higher than for BC, this means that ordering nodes by their decreasing bridgeness leads to a better ranking of the 'global' scores as measured by G. To smooth the curves, we have averaged over 200 points. (b–d) Average relative local terms as function of node degree for the three investigated networks (for definition see text).

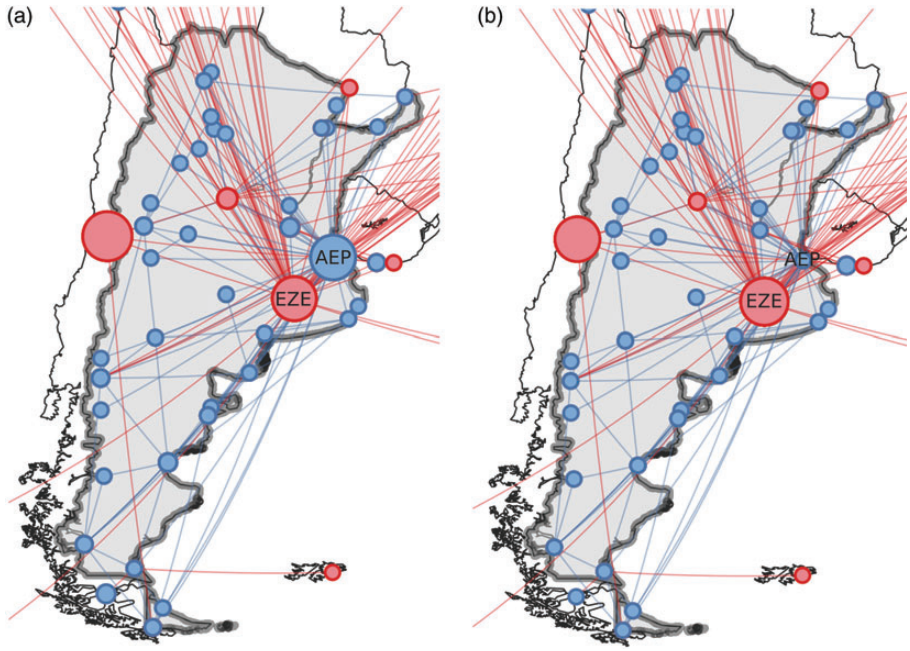

FIG. 4. Example of the two largest Argentinean airports, Ezeiza (EZE) and Aeroparque (AEP). Both have a similar degree (54 and 45, respectively), but while the first connects Argentina to the rest of the world (85% of international connections, average distance 2.848 miles, $G = 2327.2$), Aeroparque is only a local centre (18% of international connections, average distance 570 miles, $G = 9.0$). However, as in the simple graph (Fig. 1), BC gives the same score to both ($\mathrm{BC}_{EZE} = 79,000$ and $\mathrm{BC}_{AEP} = 82,000$), while bridgeness clearly distinguishes the local centre and the bridge to the rest of the world, by attributing to the global bridge a score 250 times higher ($\mathrm{Bri}_{EZE} = 46,000$ and $\mathrm{Bri}_{AEP} = 174$). (a) Betweenness centrality and (b) bridgeness centrality.

5. Real network 1: airport's network

Proving the adequacy of bridgeness to spot out global bridges on real networks is more difficult, because generally communities are not unambiguously defined, therefore neither are global bridges. Then, it is difficult to show conclusively that bridgeness is able to specifically spot these nodes. To answer this challenge, our strategy is the following:

(i) We use flight itinerary data providing origin destination pairs between commercial airports in the world (International Air Transport Association). The network collects 47,161 transportation connections between 7733 airports. Each airport is assigned to its country.

(ii) We consider each country to be a distinct 'community' and compute a global indicator based on this partitioning, as it allows for an objective (and arguably relevant) partition, independent from any community detection methods. Then we show that bridgeness offers a better ranking than BC to identify airports that act as global bridges, i.e. that connect countries internationally.

As an example, in Fig. 4 we show the two largest airports of Argentina, Ezeiza (EZE) and Aeroparque (AEP). Both have a similar degree (54 and 45, respectively), but while the first connects Argentina to the rest of the world, Aeroparque mostly handles domestic flights, thus functioning as a local centre.

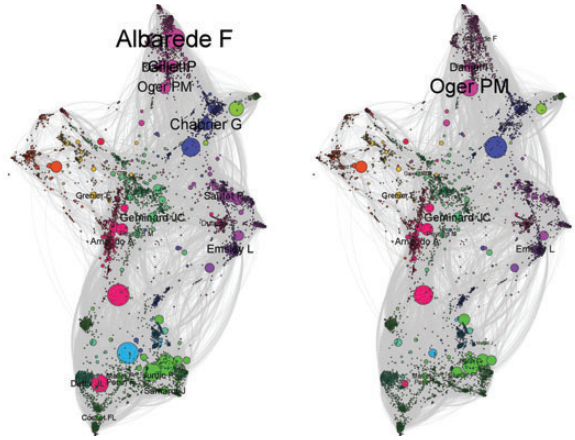

FIG. 5. Co-citation and co-author network of articles published by scientists at ENS de Lyon. Nodes represent the authors or references appearing in the articles, while links represent co-appearances of these features in the same article. The colour of the nodes corresponds to the modularity partition and their size is proportional to their BC (left) or to their bridgeness (right), which clearly leads to different rankings (references cited are used in the computations of the centrality measures but appear as dots to simplify the picture). We only keep nodes that appear on at least four articles and links that correspond to at least two co-appearances in the same paper. After applying these thresholds, the 8000 articles lead to 8883 nodes (author or references cited in the 8000 articles) and 347,644 links. The average degree is 78, the density 0.009 and the average clustering coefficient is 0.633. Special care was paid to avoid artefacts due to homonyms. Weights are attributed to the links depending on the frequency of co-appearances (cosine distance, see [28]).

This is confirmed by the respective G values: 2327.2 (EZE) and 9.0 (AEP). However, just like in our simple example in Fig. 1, BC gives the same score to both, while bridgeness clearly distinguishes between the local domestic centre and the global international bridge by attributing to the global bridge a score 250 times higher (see Fig. 4). This can partly be explained by the fact that AEP is a 'star' node (low clustering coefficient: 0.072), connected to 12 very small airports, for which it is the only link to the whole network. All the paths starting from those small airports are cancelled in the computation of the bridgeness (they belong to the 'local' term in equation (2.1)), while BC counts them equally as any other path.

More generally, Fig. 3 shows that, as for the Airport network, bridgeness provides again a ranking that is closer to that of the global indicator. Indeed, ordering nodes by their decreasing bridgeness leads to a ranking that is closer to the ranking obtained by the global score than the ranking by decreasing BC. In addition, we found again negative correlations between the average relative local term and node degrees (see Fig. 3(c)), assigning similar roles for low and high-degree nodes as in case of the synthetic network.

6. Real network 2: scientometric network of ENS Lyon

The second example of a real network is a scientometric graph of a scientific institution [28], the 'Ecole normale supérieure de Lyon' (ENS, see Fig. 5). This networks adds authors to the usual co-citation network, as we want to understand which authors connect different subfields and act as global, interdisciplinary bridges. To identify the different communities, we rely on modularity optimization [27], which leads to a relevant community partition because scientific networks are highly structured by disciplinary boundaries. This is confirmed by the high value of modularity generated by this partition

(0.89). In Fig. 5, the authors of different communities are shown with different colours, and their size corresponds to their betweenness (left) or bridgeness (right) centrality, which clearly leads to highlight different authors as the main global bridges, which connect different subfields. We compute the Stirling indicator (equation (2.1)) based on the modularity structure to identify the global bridges. As for the previous networks, Fig. 3 shows that bridgeness ranks the nodes in a closer way than BC to the ranking provided by the global measure based on community partition. On the other hand, the corresponding $\langle locterm \rangle(k)$ function (see Fig. 3(d)) suggests a slightly different picture in this case. Here nodes with large but moderate degrees (smaller than ~ 200) have high local terms suggesting that they act as local centres, while nodes with higher degrees have somewhat smaller local terms assigning their role to act as global bridges.

7. Discussion

In this paper, we introduced a measure to identify nodes acting as global bridges in complex network structures. Our proposed methodology is based on the decomposition of BC into a local and global term, where the local term considers shortest paths that start or end at one of the node's neighbours, while the global term, what we call bridgeness, is more specific to identify nodes which are globally central. We have shown, on both synthetic and real networks, that the proposed bridgeness measure improves the capacity to specifically find out global bridges as it is able to distinguish them from local centres. One crucial advantage of our measure of bridgeness over former propositions is that it is independent of the definition of communities.

However, the advantage in using bridgeness depends the precise topology of the network, and mainly on the degree distribution of bridges when compared with that of all the nodes in the network. When bridges are high-degree nodes, BC and bridgeness give an equally good approximation, since high-degree bias do not play an important role in this case. Instead, when some bridges have low degrees, while some high-degree nodes act like local centres of their own community, bridgeness is more effective to identify bridges as BC gives equally high rank to nodes with high degree, even if they are not connected to nodes outside of their community. We demonstrated that bridgeness is systematically more specific to spot out global bridges in all the networks we have studied here. Although the improvement was small on average, typically 5–10%, even a small amelioration of a widely used measure is in itself an interesting result.

We should also note that, except on simple graphs, comparing these two measures is difficult since there is no clear way to identify, independently, the 'real' global bridges. We have used community structure when communities seem clear-cut, but then we fall into the circularity problems stressed in the introduction. Using metadata on the nodes (i.e. countries for the airports) may solve this problem but raises others, as metadata do not necessarily correspond to structures obtained from the topology of the network, as shown recently on a variety of networks [29]. Another possible extension would be to identify overlapping communities to identify independently global bridges, as nodes involved in multiple communities, and correlate them with the actual measure, which provides a direction for future studies. However, in any case identifying global bridges remains a difficult problem as it is tightly linked to another difficult problem, that of community detection. Decomposing BC into a local and a global term helps to improve the solution, but many questions remain still open for further inquiry.

Supplementary data

Supplementary data are available at *Journal of Complex Networks* online.

Funding

Funding provided by the CNRS "Mission pour l'interdisciplinarité" (MASTODONS) is gratefully acknowledged.

References

1. Euler, L. (1736, 1741) Solutio problematis ad geometriam situs pertinentis. *Comment. Acad. Sci. Petropolitanae*, **8**, 128–140.
2. Granovetter, M. S. (1973) The strength of weak ties. *Amer. J. Sociol.*, **78**, 1360–1380.
3. Coser, R. (1975) The complexity of roles as seedbed of individual autonomy. *The Idea of Social Structure: Essays in Honor of Robert Merton* (L. Coser ed.). New York: Harcourt Brace Jovanovich.
4. Vespignani, A. (2009) Predicting the behavior of techno-social systems. *Science*, **325**, 425.
5. Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. & Van Alstyne, M. (2009) Life in the network: the coming age of computational social science. *Science*, **323**, 721–723.
6. Venturini, T. & Latour, B. (2010) The social fabric: digital traces and quali-quantitative methods. *Proceedings of Future En Seine 2009*. Paris: Cap Digital, pp. 87–101.
7. Burt, R. S. (2005) *Brokerage and Closure: An Introduction to Social Capital*. Oxford: Oxford University Press.
8. Nepusz, T., Petróczi, A., Négyessy, L. & Bazsó, F. (2008) Fuzzy communities and the concept of bridgeness in complex networks. *Phys. Rev. E*, **77**, 016–107.
9. Cheng, X.-Q., Ren, F.-X., Shen, H.-W. & Zhang, Z.-K. (2010) Bridgeness: a local index on edge significance in maintaining global connectivity. *J. Stat. Mech. Theory Exp.*, **2010**, P10–011.
10. Brandes, U. (2001) A faster algorithm for betweenness centrality. *J. Math. Sociol.*, **25**, 163–177.
11. Bastian, M., Heymann, S. & Jacomy, Q. (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceeding of the 3rd International AAAI Conference on Weblogs and Social Media*. Menlo Park: AAAAI Press, pp. 361–362.
12. Bonacich, P. (1987) Power and centrality: a family of measures. *Amer. J. Sociol.*, **92**, 1170–1182.
13. Borgatti, S. P. (2005) Centrality and network flow. *Soc. Netw.*, **27**, 55–71.
14. Estrada, E. & Rodriguez-Velazquez, J. A. (2005) Subgraph centrality in complex networks. *Phys. Rev. E*, **71**, 056103.
15. Newman, M. (2010) *Networks: An Introduction*. Oxford: Oxford University Press.
16. Freeman, L. C. (1977) A set of measures of centrality based on betweenness. *Sociometry*, 35–41.
17. Freeman, L. C. (1979) Centrality in social networks conceptual clarification. *Soc. Netw.*, **1**, 215–239.
18. Nakao, K. (1990) Distribution of measures of centrality: enumerated distributions of Freeman's graph centrality measures. *Connections*, **13**, 10–22.
19. Goh, K.-I., Oh, E., Kahng, B. & Kim, D. (2003) Betweenness centrality correlation in social networks. *Phys. Rev. E*, **67**, 017101.
20. Newman, M. E. J. (2005) A measure of betweenness centrality based on random walks. *Soc. Netw.*, **27**, 39–54.
21. Leydesdorff, L. (2007) Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *J. Amer. Soc. Inf. Sci. Technol.*, **58**, 1303–1319.
22. Rao, C. R. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theor. Popul. Biol.*, **21**, 24–43.
23. Stirling, A. (2007) A general framework for analysing diversity in science, technology and society. *J. R. Soc. Interface*, **4**, 707–719.
24. Rafols, I. (2014) Knowledge integration and diffusion: measures and mapping of diversity and coherence. *Measuring Scholarly Impact* (Y. Ding, R. Rousseau, & D. Wolfram, eds) pp. 169–190. Berlin: Springer.

25. Jensen, P. & Lutkouskaya, K. (2014) The many dimensions of laboratories interdisciplinarity. *Scientometrics*, **98**, 619–631.
26. Lancichinetti, A., Fortunato, S. & Radicchi, F. (2008) Benchmark graphs for testing community detection algorithms. *Phys. Rev. E*, **78**, 046110.
27. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. (2008) Fast unfolding of communities in large networks. *J. Stat. Mech.*, **10**, P10008.
28. Grauwin, S. & Jensen, P. (2011) Mapping scientific institutions. *Scientometrics*, **89**, 943–954.
29. Hric, D., Darst, R. K. & Fortunato, S. (2014) Community detection in networks: structural communities versus ground truth. *Phys. Rev. E*, **90**, 062805.