IMA Journal of Complex Networks (2015) Page 1 of 14
doi:10.1093/comnet/xxxxxx

Detecting global bridges in networks

PABLO JENSEN*
IXXI, Institut Rhonalpin des Systemes Complexes, ENS Lyon; Laboratoire de Physique, UMR 5672, ENS Lyon 69364 Lyon, France
*Corresponding author: pablo.jensen@ens-lyon.fr

MATTEO MORINI
IXXI, Institut Rhonalpin des Systemes Complexes, ENS Lyon; LIP, INRIA, UMR 5668, ENS de Lyon 69364 Lyon, France

MÁRTON KARSAI
IXXI, Institut Rhonalpin des Systemes Complexes, ENS Lyon; LIP, INRIA, UMR 5668, ENS de Lyon 69364 Lyon, France

TOMMASO VENTURINI
Médialab, Sciences Po, Paris

ALESSANDRO VESPIGNANI
MoBS, Northeastern University, Boston MA 02115 USA; ISI Foundation, Turin 10133, Italy

MATHIEU JACOMY
Médialab, Sciences Po, Paris

JEAN-PHILIPPE COINTET
Université Paris-Est, SenS-IFRIS

PIERRE MERCKLÉ
Centre Max Weber, UMR 5283, ENS Lyon 69364 Lyon, France

ERIC FLEURY
IXXI, Institut Rhonalpin des Systemes Complexes, ENS Lyon; LIP, INRIA, UMR 5668, ENS de Lyon 69364 Lyon, France

[Received on XX XX XXXX; revised on XX XX XXXX; accepted on XX XX XXXX]

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 is used to evaluate a node capacity to connect different graph regions. However, we argue here that this measure is deceptive as it gives equal scores to local centers (*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 organisation of the network structure. In this paper we propose a better indicator, called bridgeness centrality, which is capable to differentiate between local and global bridges in a network structure. Further we introduce an effective algorithmic implementation of this measure and demonstrate its capability to identify globally central nodes in air transportation and scientific collaboration networks.

Keywords: Centrality Measures, Betweenness Centrality, Bridgeness Centrality

JXX, JYY

1. Introduction

Although the history of graphs as scientific objects begins with Euler's [9] 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 [13]. 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 tonnesian cliche, strong ties generate Gemeinshaft, while weak ties generates Gesellshaft [7]. 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). Lets 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-centered 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 endeavor 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 [19, 29, 30].

Let's 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 [5]) 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 interdisciplinarity. 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 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

[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.

structure and thus escapes this vicious circle, contrary to other proposals [6, 23]. Moreover, since the computation of bridgeness is straightforwardly related to that of the usual betweenness, Brandes' algorithm [4] 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.

Measuring bridgeness

Identifying important nodes in a network structure is crucial for the understanding of the associated real-world system [2, 3, 8], for a review see [24]. 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 [10, 11]. We show in the following that *BC* is not specific enough, since it cannot distinguish between two types of centralities: local (center 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 centers than to global bridges is easy to see in a simple network (Figure 1). The logics is that a star node with degree k, *i.e.* a node without links between all its first neighbors (clustering coefficient 0) receives automatically a $BC = k(k-1)/2$ arising from paths of length 2 connecting the node's first neighbors 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 [12, 22, 25]. A recent scientometrics study tried to use betweenness centrality as an indicator of the interdisciplinarity of journals but noted that this idea only worked in local citation environments and after normalization because otherwise the influence of degree centrality dominated the betweenness centrality measure [20].

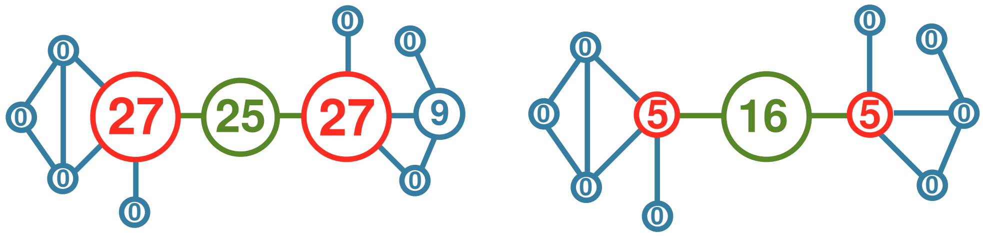
To avoid this problem and specifically spot out global centers, we introduce the bridgeness centrality. Since we want to distinguish global bridges from local ones the simplest approach is to discard shortest paths, which either start or end at a node's first neighbors from the summation to compute *BC* (Eq. 1.1). This completely removes the paths that connect two non connected neighbors for 'star nodes' (see Figure 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 $\mathscr{G} = (V,E)$, where V assigns the set of nodes and E the set of links the definition of the betweenness centrality for a node $j \in V$ stands as:

$$BC(j) = \sum_{i \neq j \neq k} \frac{\sigma_{ik}(j)}{\sigma_{ik}} = \sum_{\substack{i \notin N_G(j) \text{ and} \\ k \notin N_G(j)}} \frac{\sigma_{ik}(j)}{\sigma_{ik}} + \sum_{\substack{i \in N_G(j) \text{ or} \\ k \in N_G(j)}} \frac{\sigma_{ik}(j)}{\sigma_{ik}} \tag{1.1}$$

where 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,

[3]We have written a plug-in for Gephi [1] that computes this measure on large graphs. See Materials and Methods for a pseudo-algorithm for both node and edge bridgeness.



(a) Betweenness centrality (b) Bridgeness centrality

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

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. Fig. 1 illustrates the ability of bridgeness to specifically highlight nodes that connect different regions of a graph. Here the BC (Fig. 1a) and bridgeness centrality values (Fig. 1b) calculated for nodes of the same network demonstrate that bridgeness centrality assigns correctly the node, which is central globally (green), while BC suggests a confusing picture as it assigns higher centrality values 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 always outperforms BC in terms of the identification of real bridges.

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 [16, 26–28], 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 Eq.1.2 for node i is defined as:

$$G(i) = \sum_{j \in J} l_{IJ} \delta_{i,J} \tag{1.2}$$

where the sum is 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

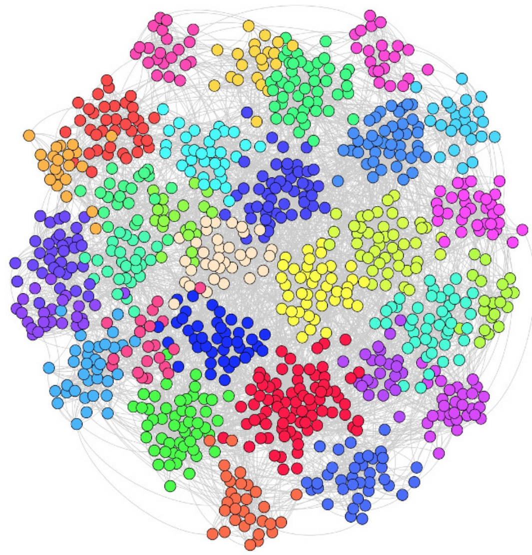


FIG. 2. Artificial network with a clear community structure using Lancichinetti et al [17] method. For clarity, we show here a smaller network containing 1000 nodes, 30 communities, 7539 links (20% inter-and 80% intra-community links). Each color corresponds to a community as detected by modularity optimization [24].

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.

Synthetic network: unbiased LFR

We start with a synthetic network obtained by a method similar to that of Lancichinetti et al [17]. 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 shows that bridgeness provides a better ranking than BC in this case, *i.e.* a ranking that is closer to that of the global indicator. 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

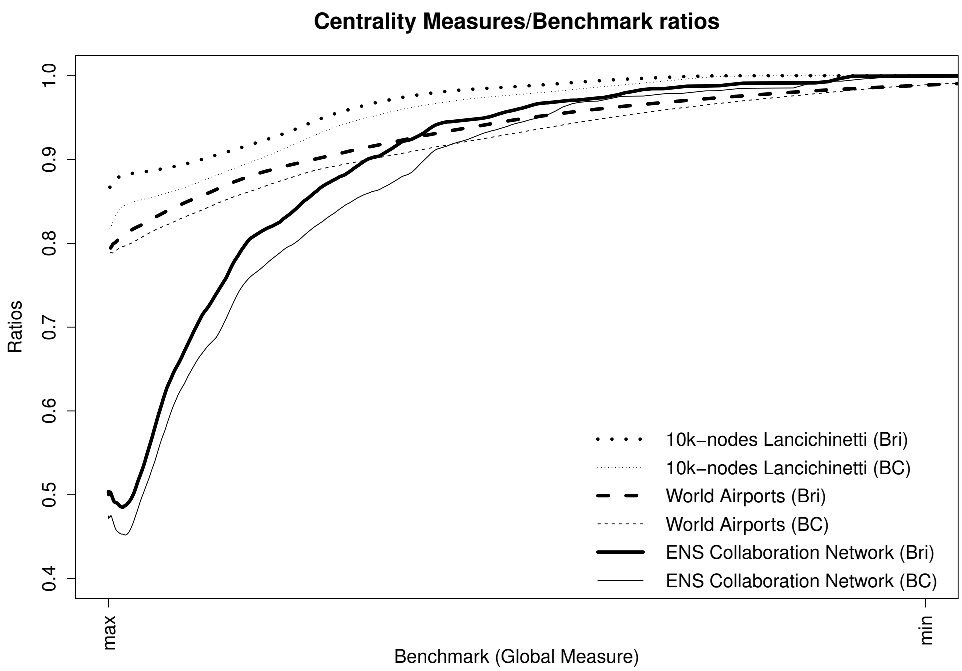


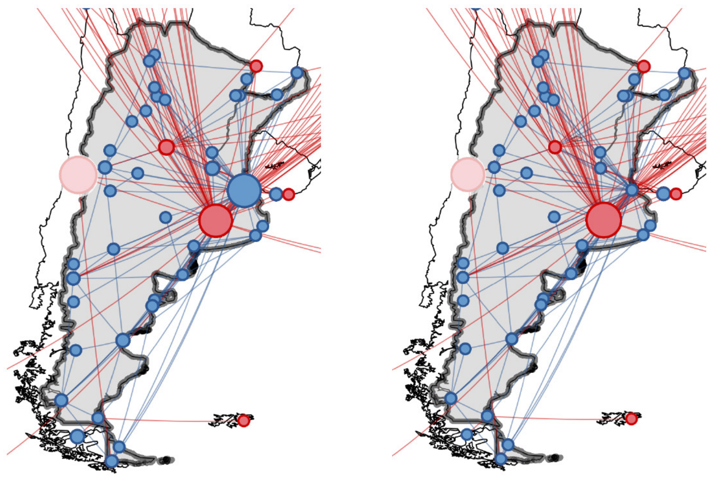
FIG. 3. Ability of BC or bridgeness to reproduce the ranking of bridging nodes, taking as reference the global indicator (Eq 2). For each of the three networks, we first compute the cumulative sums for the global measure G, according to three sorting options: the G measure itself and the two centrality 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.

'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 Supplementaty Information (Scluster5).

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 47161 transportation connections



(a) Betweenness centrality (b) Bridgeness centrality

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 center (18% of international connections, average distance 570 miles, G=9.0). However, as in the simple graph (Figure 1), BC gives the same score to both ($BC_{E}ZE$=79,000 and $BC_{A}EP$ = 82,000), while bridgeness clearly distinguishes the local center and the bridge to the rest of the world, by attributing to the global bridge a score 250 times higher ($Bri_{E}ZE$=46,000 and $Bri_{A}EP$ = 174). Red nodes represent international airports while blue nodes are domestic.

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 center. 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 center 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 Eq.1.1), while *BC* counts them equally as any other path.

More generally, Figure 3 shows, as for the synthetic network, that bridgeness provides a better ranking than *BC*, ie 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*. To give some intuition about this difference in the rankings, Table 1 shows the Top 20 airports according to the global score, together with their ranking according to

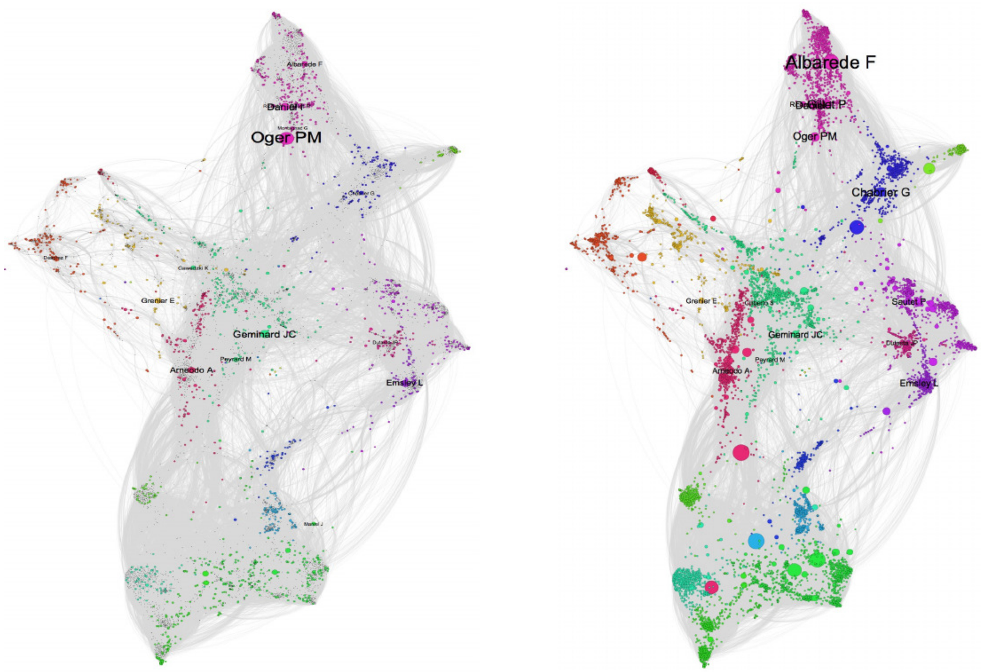
Table 1. Top 20 Airports, sorted by decreasing G

IATA Code	City	Country	BC	Bri	Rank BC	Rank Bri
FRA	Frankfurt	Germany	613100	401875	1	2
CDG	Paris	France	356542	195102	5	6
IST	Istanbul	Turkey	209588	65571	19	26
AMS	Amsterdam	Neth.	314519	154392	6	9
DXB	Dubai	UAE	356934	198960	4	5
FCO	Rome	Italy	143661	49493	32	35
LHR	London	UK	188712	138472	22	11
LGW	London	UK	95452	36113	53	54
MXP	Milano	Italy	64344	29875	84	63
BRU	Brussels	Belgium	59854	19672	91	83
JFK	New York	USA	213751	135433	16	12
MAD	Madrid	Spain	144820	73673	31	21
ZRH	Zurich	Switz.	60117	26343	90	69
MUC	Munich	Germany	101915	38904	48	50
VIE	Vienna	Austria	30157	5976	157	164
SVO	Moscow	Russia	89802	33712	57	58
YYZ	Toronto	Canada	274575	165073	9	8
DOH	Doha	Qatar	78238	56266	70	30
NBO	Nairobi	Kenya	93444	41016	54	48
BKK	Bangkok	Thailand	214030	88557	15	19

bridgeness and *BC*. In most cases, the bridgeness ranking is closer to that obtained from the global indicator.

Real network 2: scientometric network of ENS Lyon

The second example of a real network is a scientometric graph of a scientific institution [14], the "Ecole normale superieure de Lyon" (ENSL, see Figures 5a and b). This networks adds authors to the usual co-citation network, as we want to understand which authors connect different sub-fields and act as global, interdisciplinary bridges. To identify the different communities, we rely on modularity optimization, 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 Figures 5a and b, the authors of different communities are shown with different colors, and their size corresponds to their *BC* (a) or bridgeness (b) centrality, which clearly leads to highlight different authors as the main global bridges, which connect different subfields. We compute the Stirling indicator (Eq.1.1) 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.



(a) Betweenness centrality (b) Bridgeness centrality

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 color of the nodes corresponds to the modularity partition and their size is proportional to their bridgeness (a) or to their BC (b), 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 2 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 artifacts due to homonyms. Weights are attributed to the links depending on the frequency of co-appearances (cosine distance, see [14].

Discussion

We have shown, on synthetic and real networks, that using *BC* to identify global bridges leads to misclassifications, as *BC* is not always able to distinguish local centers from global bridges. Instead, bridgeness, by eliminating those shortest paths that start or end at one of the node's neighbors, improves the capacity to specifically find out global bridges. 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 on how much *BC* is misguided by its high degree bias. This in turn depends on the precise topology of the network, and mainly on the degree distribution of bridges as compared to that of all the nodes in the network. When bridges are also high-degree nodes, *BC* will not suffer from its high-degree bias to detect them. Instead, when some bridges have low degrees while some high-degree nodes are only centers of their own community, *BC* may misclassify these centers as global bridges, even if they are not connected to nodes outside their community. In the example of the Argentinean airports, *BC* spectacularly fails to distinguish a truly global bridge, namely the EZE international airport, from a local center, namely AEP. More generally, bridgeness is systematically better than *BC* in all the networks we have studied here, but only by a small amount on average, typically 5 to 10%. But 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) solves 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 [15]. Identifying global bridges remains a difficult problem because it is tightly linked to another difficult problem, that of community detection. Bridgeness allows to avoid some pitfalls of *BC* but many questions are still open for further inquiry.

Supplementary Information: algorithm

Bridgeness algorithm, inspired by Brandes' "faster algorithm" [4]

SP[s,t] $\leftarrow$precompute all shortest distances matrix/dictionary
CB[v] $\leftarrow$ 0, v $\in$ V ;
for s $\in$ V **do**
  S $\leftarrow$ empty stack;
  P[w] $\leftarrow$ empty list, w $\in$ V ;
  σ[t] $\leftarrow$ 0, t $\in$ V ; σ[s] $\leftarrow$ 1;
  d[t] $\leftarrow -1$, t $\in$ V ; d[s] $\leftarrow$ 0;
  Q $\leftarrow$ empty queue;
  enqueue s $\rightarrow$ Q;
  while Q not empty **do**
    dequeue v $\leftarrow$ Q;
    push v $\rightarrow$ S;
    foreach neighbor w of v **do**
      // w found for the first time?
      if d[w] $<$ 0 **then**

```
        enqueue w → Q;
        d[w] ← d[v] + 1;
      end
      // shortest path to w via v?
      if d[w] = d[v] + 1 then
        σ[w] ← σ[w] + σ[v];
        append v → P[w];
      end
    end
  end
  δ[v] ← 0, v ∈ V ;
  // S returns vertices in order of non-increasing distance from s
  while S not empty do
    pop w ← S;
    for v ∈ P[w] do δ[v] ← δ[v] + σ[v]/σ[w] · (1 + δ[w]);
    if SP[w,s]>1 then CB[w] ← CB[w] + δ[w];
  end
end
```

Supplementary Information $S\_Cluster5$

The specificity of bridgeness and the influence of the degree, which prevents BC from identifying correctly the most important bridges, can be exemplified by examining the scores of nodes in cluster 5 of the synthetic network. This cluster is linked to cluster 13 by 5 connections (through nodes 248, 861, 471, 576 and 758) and to cluster 1 by a single connection (through node 232). BC gives roughly the same score to nodes 232 and 248, while bridgeness attributes a score almost 4 times higher to node 232, correctly pointing out the importance of this single bridge between clusters 5 and 1. This is because BC is confused by the high degree of node 248 (41) as compared to node 232 low degree (20). Therefore, by counting all the shortest paths, BC attributes too high a bridging score to node 248. Second problem with BC, it gives a high score to nodes that are not connected to other communities, merely because they are local centers, *i.e.* they have a high degree. For example, node 515 obtains a higher BC score than node 758 (Table S1), even if node 515 has no connection to other communities (but degree 49), contrary to node 758 (connected to cluster 5, but degree 23). Bridgeness never ranks higher local centers than global bridges: here, it correctly assigns a 5 times higher score to node 758 than to node 515.

REFERENCES

1. Bastian, M., Heymann, S. & Jacomy, Q. (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media*.
2. Bonacich, P. (1987) Power and centrality: A family of measures. *American journal of sociology*, **92**(5), 1170–1182.
3. Borgatti, S. P. (2005) Centrality and network flow. *Social networks*, **27**(1), 55–71.
4. Brandes, U. (2001) A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, **25**(2), 163–177.
5. Burt, R. S. (2005) *Brokerage and closure: An introduction to social capital*. Oxford University Press, Oxford.

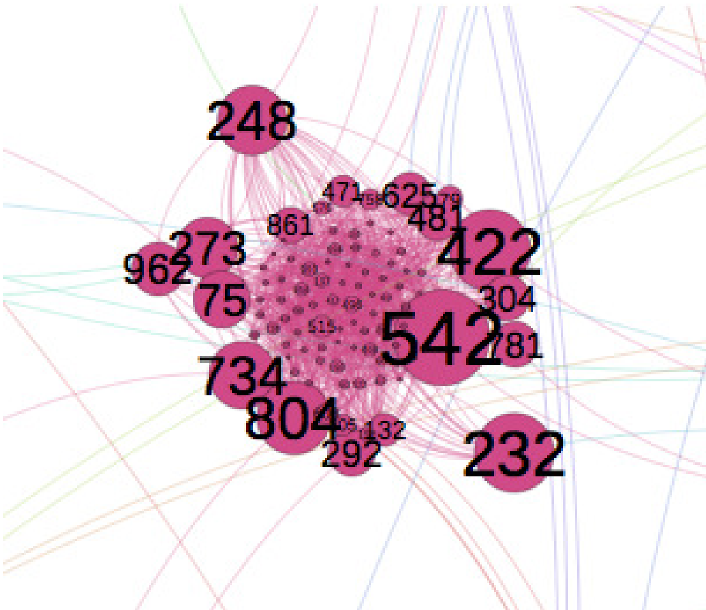


FIG. S1. Zoom on cluster 5 of the synthetic network. The numbers show node's labels, while the size of the nodes is proportional to their BC score.

6. Cheng, X.-Q., Ren, F.-X., Shen, H.-W. & Zhang, Z.-K. (2010) Bridgeness: a local index on edge significance in maintaining global connectivity. *Journal of Statistical Mechanics: Theory and Experiment*, **2010**(10), P10011.

7. Coser, R. (1975) The Complexity of Roles as Seedbed of Individual Autonomy.. In Coser, L., editor, *The Idea of Social Structure: Essays in Honor of Robert Merton.* Harcourt Brace Jovanovich, New York.

8. Estrada, E. & Rodriguez-Velazquez, J. A. (2005) Subgraph centrality in complex networks. *Physical Review E*, **71**(5), 056103.

9. Euler, L. (1741 (1736)) Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, **8**, 128–140.

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

11. Freeman, L. C. (1979) Centrality in social networks conceptual clarification. *Social networks*, **1**(3), 215–239.

12. Goh, K.-I., Oh, E., Kahng, B. & Kim, D. (2003) Betweenness centrality correlation in social networks. *Physical Review E*, **67**(1), 017101.

13. Granovetter, M. S. (1973) The strength of weak ties. *American journal of sociology*, **78**(6), 1360–1380.

14. Grauwin, S. & Jensen, P. (2011) Mapping scientific institutions. *Scientometrics*, **89**(3), 943–954.

15. Hric, D., Darst, R. K. & Fortunato, S. (2014) Community detection in networks: Structural communities versus ground truth. *Physical Review E*, **90**(6), 062805.

16. Jensen, P. & Lutkouskaya, K. (2014) The many dimensions of laboratories interdisciplinarity. *Scientometrics*, **98**(1), 619–631.

17. Lancichinetti, A., Fortunato, S. & Radicchi, F. (2008) Benchmark graphs for testing community detection algorithms. *Physical review E*, **78**(4), 046110.

18. Latour, B., Jensen, P., Venturini, T., Grauwin, S. & Boullier, D. (2012) The whole is always smaller than its parts–a digital test of Gabriel Tardes' monads. *The British journal of sociology*, **63**(4), 590–615.

19. 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

Table S1. Nodes in community 5 of the synthetic network, ranked by decreasing BC (see text)

Id	Stirling	Modularity Class	Betweenness	Bridgeness	Degree
542	0.0222	5	9173.71	2644.62	44
422	0.0278	5	7714.27	3855.62	35
232	0.0950	5	7551.22	5846.86	20
804	0.0285	5	6995.63	2824.64	34
248	0.0082	5	6588.65	1624.30	48
734	0.0907	5	6410.31	4373.72	21
273	0.0322	5	5698.28	2631.59	30
75	0.0868	5	5349.47	3558.31	22
962	0.0399	5	4989.66	2951.45	24
292	0.0399	5	4377.77	1939.06	24
481	0.0256	5	4305.68	1796.92	25
781	0.0475	5	4257.93	2200.21	20
304	0.0434	5	4221.64	2467.65	22
625	0.0202	5	3964.21	1314.62	32
861	0.0108	5	3295.01	714.44	36
132	0.0200	5	2985.45	1157.49	24
471	0.0154	5	2865.07	1296.38	25
79	0.0302	5	2256.02	1004.28	21
205	0.0208	5	1921.65	788.51	23
515	0.0000	5	1884.07	86.45	49
758	0.0166	5	1791.80	435.66	23
608	0.0200	5	1777.54	522.75	24

network: the coming age of computational social science. *Science*, **323**(5915), 721–723.

20. Leydesdorff, L. (2007) Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, **58**(9), 1303–1319.

21. Louf, R., Jensen, P. & Barthelemy, M. (2013) Emergence of hierarchy in cost-driven growth of spatial networks. *Proceedings of the National Academy of Sciences*, **110**(22), 8824–8829.

22. Nakao, K. (1990) Distribution of measures of centrality: enumerated distributions of Freeman's graph centrality measures. *Connections*, **13**(3), 10–22.

23. Nepusz, T., Petróczi, A., Négyessy, L. & Bazsó, F. (2008) Fuzzy communities and the concept of bridgeness in complex networks. *Physical Review E*, **77**(1), 016107.

24. Newman, M. (2010) *Networks: an introduction*. Oxford University Press.

25. Newman, M. E. J. (2005) A measure of betweenness centrality based on random walks. *Social networks*, **27**(1), 39–54.

26. Rafols, I. (2014) Knowledge Integration and Diffusion: Measures and Mapping of Diversity and Coherence. In Ding, Y., Rousseau, R. & Wolfram, D., editors, *Measuring Scholarly Impact*, pages 169–190. Springer.

27. Rao, C. R. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*, **21**(1), 24–43.

28. Stirling, A. (2007) A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, **4**(15), 707–719.

29. Venturini, T. & Latour, B. (2010) The social fabric: Digital traces and quali-quantitative methods. *Proceedings of Future En Seine*, pages 30–15.

30. Vespignani, A. (2009) Predicting the behavior of techno-social systems. *Science*, **325**(5939), 425.