Journal of Complex Networks Advance Access published August 6, 2014

Journal of Complex Networks (2014) Page 1 of 19
doi:10.1093/comnet/cnu031

Border sensitive centrality in global patent citation networks

Greg Morrison† and Eleftherios Giovanis
IMT Institute for Advanced Studies, Lucca 55100, Italy
†Corresponding author. Email: greg.morrison@imtlucca.it

Fabio Pammolli
IMT Institute for Advanced Studies, Lucca 55100, Italy and Department of Economics, Harvard University, Cambridge, MA 02138, USA

and

Massimo Riccaboni
IMT Institute for Advanced Studies, Lucca 55100, Italy and Department of MSI, KU Leuven, Leuven, Belgium

Edited by: Hocine Cherifi

[Received on 24 February 2014; accepted on 24 June 2014]

When resources are shared between interacting networks, the importance of each node depends strongly on how collaborative or competitive each sub-network is. In this paper, we develop a new method of measuring centrality in the complex network of patent citations that can take political borders into account, where the national benefit of domestic citations relative to foreign citations can be controlled by a free parameter. We find that while some patent classes are of high importance both in the global and the domestic economy, there often exist patent classes in individual countries that are more central nationally than in global economy. We characterize the most important classes globally and domestically for six different nations, and describe their robustness to various perturbations to the model and to noise.

Keywords: innovation networks; centrality; pagerank; global systems.

1. Introduction

The study of complex networks has been shown to be pertinent to a wide range of fields, including biology and medicine [1–4], economics [5,6], sociology [7,8] and information science [9]. The topology of a network is of great importance in a variety of contexts [2–5,10,11], with significant attention paid to a variety of topologically defined node centralities [6,12–15] measuring the importance of an individual node relative to the rest of the network. Most centrality measures address the issue of node importance for networks with edges of a single type, but many recent studies focus on networks-of-networks [16–20] or the related field of multiplexed networks [21–26] and have begun to shed further light on the behaviour of complex interacting systems by understanding the influence that one subsystem may have over another due to their interdependencies [16,25,26]. The influence of a node within its own sub-network as well as within the entire interacting system [16,21,27] is essential to quantify through a measure of centrality that incorporates the global topology of the network.

In many contexts, networks-of-networks may be used to represent the interdependence between distinct groups of nodes that may either directly compete [16] or have an intrinsic preference for self-interaction. Sharing resources with another sub-network B may be considered detrimental (or at least not directly beneficial) from the perspective of A, whereas resources flowing in the opposite direction would be considered beneficial from the perspective of A. One example of such a system, which we focus on here, is the global network of national patent classes and the citations between them. Here the sub-networks represent national economies; nodes represent research areas (the patent classes) within those countries and the information flow (the resource shared) models the technological spillovers [28,29] between the various patent classes. Growth in one domestic patent class that is heavily cited by foreign patents may induce greater spillovers outside of the domestic economy, which may be considered detrimental to the domestic economy. While classes that are central to the global citation network will generally be those that are heavily cited [9], one may expect that domestically central classes are those that are simultaneously heavily cited primarily by domestic classes rather than foreign classes.

In this paper, an expansion on our previous work [27], we model the information flow originating in a specific national network as a random walk in the global citation network with an imposed loss term [15,30] (a heightened discount rate [31,32] in economic terms) depending on whether the information is propagating within the originating country or circulating outside of it. The loss term acts as a bias against border crossing, where information that benefits a foreign economy is assigned an elevated probability of becoming ‘lost’ (or incapable of benefiting the domestic economy in the future) [30,33]. We define an expression for the centrality of domestic patent classes similar to PageRank [9,15,21] that varies continuously from a global measure of centrality (where there is no bias against information circulating in a foreign economy) to a domestic measure (which is maximally biased against border crossings) using a single parameter. We show that while heavily cited classes are often of importance both globally and domestically, there are many cases where smaller patent classes with a higher fraction of domestic citations are significantly more domestically central than they are globally. We show the model is robust to a wide range of parameter variations as well as random noise, and that the technological class centralities are less likely to be perturbed by noise in larger economies than those in smaller economies.

2. The patent citation network

Studying the structure of the patent citation network can provide insight into the dynamical processes of innovation that give rise to the filing of patents and the relationships between industries. Each office has a method for categorizing the patents, and the European Patent Office (EPO) categorization into International Patent Classes (IPCs) is hierarchically organized [34]: each patent is assigned 7-digit identifier(s) with the first digit indicating the category of the patent (e.g. section H corresponds to electricity), the second and third digits the patent’s class (e.g. H01 corresponds to ‘basic electric elements’), the fourth digit a sub-class (e.g. H01C corresponds to ‘Resistors’) and so on. By considering only triadic patents (related patents that cover the same claim in the EPO, the US Patent and Trademark Office (USPTO) and Japanese Patent Office (JPO)), we can apply the EPO hierarchy to aggregate similar patents that are filed in the EPO as well as the USPTO or JPO. Studying the citation network of triadic patents has the additional advantages of focusing on patents that are more internationally visible (and thus are more likely to be read by a world-wide audience) as well as those that the applicant deemed worthy of applying for in three different offices. The significant financial investment required to carry a patent through three different patent offices (with three application costs and different legal systems to navigate) suggests that triadic patents encompass many innovations that may be considered ‘important’ on a global scale. By focusing on triadic patents, we are selecting those patents that are likely to be read internationally (thus having a greater potential for cross-border spillovers) and to be have the potential to spur further innovations. We construct the citation networks using raw data provided by the Organization for Economic Cooperation and Development, covering all citations between patents in the EPO and the World Intellectual Property Organization from 1978 to July 2013. We focus our attention primarily on the years 1987–2005, as the number of triadic patents decreases beyond 2005 due to a time lag in the granting of patents covering the same technology in all three offices (data not shown).

In order to aggregate the individual patent families into classes, we focus group patents together based on the first three digits of their IPC codes, proving $n_d = 121$ patent classes that have at least one triadic patent filed in them since 1987. Patents with multiple IPC codes are treated as being members of all three-digit classes equally. The country of each patent is based on the address of the inventor(s), and patents with inventors in different countries are treated as being from both. Each country has different institutional expertise, and we therefore expect not all countries will be equally adept the technologies required for all patent classes. We treat patent classes in each country as distinct from one another—class H01 in Germany (DE) is treated differently from H01 in Japan (JP). Throughout this paper, we only consider the six largest economies measured by the total number of triadic patents filed since 1987 as distinct sub-networks: the United States of America (USA), JP, DE, France (FR), Great Britain (GB) and the Netherlands (NL). All other economies are grouped together as an external ‘world’ (WO) economy. The aggregated WO ‘region’ comprises 18.4% of all patents in the network averaged over all sectors and all years, producing on average only 12.5% of the citations received by the six non-aggregated economies, whereas 59.4% of the citations received by the WO region come from the other economies. The aggregation of these smaller economies into a single ‘region’ thus represents a moderate perturbation to the true disaggregated network (discussed further below). These geographical–technological classes are treated as nodes in a set of directed graphs for year T, with the weighted edge from node i towards node j equal to the number of citations from class–country i filed in year T towards a patent in class–country j published in any previous year (discussed in greater detail in Section 3). The number of nodes in the network is therefore $N = 7 \times n_d = 847$ nodes representing each class–country pair.

There are three levels of aggregation that are applied to the patent citation data in this approach: a geographical aggregation on the country level, a technological aggregation on the class level and a temporal aggregation on the year level. All levels of aggregation have been applied in previous studies [5,28,35,36] and are relevant to the triadic patent citation network. The IPC hierarchy is designed specifically to provide a meaningful relationship between classes, and political borders define significant regulatory, cultural and industrial differences between locations. While different aggregations are possible (on the sector level [37] or on the level of 4–7 digit aggregation and regionally on the level of geography [35,38]), we expect the three-digit coarse graining of the network to produce relevant information due to the diversity of classes (i.e. the representation is not too coarse) while remaining computationally approachable (i.e. an analysis can be performed quickly on a single computer). Temporal aggregation over a timescale shorter than a year may be noisy, while grouping the citations into more coarse grained time windows (5 or 10 years, for example) may make a meaningful study of the temporal behaviour of the citation network difficult.

3. Random walks in innovation space

A single patent may receive many citations from future patents, and in principle any of these citations may be viewed as a ‘spillover’, acting as a proxy for knowledge transfer [39]. We expect that the value of an innovation (represented by an individual patent) is blunted by the passage of time as new technologies are created [40], and it may therefore be reasonable to expect that a patent that is the 'primary beneficiary' of an innovation spillover (i.e. a patent that could not exist without the cited patent) is temporally close to the original patent. This spillover can be thought of as the first step in a random walk from the cited class to the citing class [9,15], and so long as the structure of the global innovation network (quantified by the number of citations received by patents in class j by patents in class i filed in year T, $c_{i\to j}(T)$) does not change appreciably by the time the spillover occurs, we can treat the citation network as approximately static when determining the 'primary beneficiary' of future spillovers. If spillovers between patent classes occur more rapidly than large-scale structural changes to the economy, it is reasonable to approach the behaviour of such spillovers as a random walk on the static citation network $c_{i\to j}(T)$.

The citation network $c_{i\to j}(T)$ can be used to model innovation by recognizing that the probability of an innovation in class i benefitting class j should be a function of the fraction of citations from j to i, $f_{i\leftarrow j}$ (with the spillover moving in the opposite direction of the citations [41]). Spillovers are not expected to continue indefinitely (due partially to the blunting of the value of an innovation with time as mentioned above, and also to the possibility that an eventual spillover produces an uncited patent). We capture this expectation through a uniform probability of the walk ceasing at each step [15,33], ϵ_0, which acts as a discount parameter [31,32] in an economic context. In our model of innovation spillovers as a random walk, the bare transition probability between classes is thus taken to be

$$p^{\text{bare}}_{i\to j} = (1-\epsilon_0) f_{i\leftarrow j} = (1-\epsilon_0)\frac{c_{i\leftarrow j}(T)}{\sum_k c_{i\leftarrow k}(T)}. \tag{3.1}$$

The probability of the random walk lasting at least k steps $(1-\epsilon_0)^k$ (so ϵ_0 sets the typical number of spillovers that are expected to occur before initial innovation becomes irrelevant). If the walker is randomly reinserted into the system when a loss occurs, ϵ_0 is equivalent to a teleportation probability commonly found in the PageRank centrality measure [15,30,42]. It is possible in such a model to define a steady-state probability distribution, and we recover a PageRank measure of centrality: $P_i = \sum_k [p^{\text{bare}}_{k\to i} + \epsilon_0/N] P_k$, with N the number of nodes in the network and $\epsilon_0 \equiv 0.15$ the typical value in the standard PageRank algorithm (although we note there is no *a priori* reason to choose this value in the context of citation networks). PageRank has already been highlighted as a useful centrality measure for citation networks in other contexts [9] and shown to be adaptable to multiplexed networks [21].

4. Border sensitive centrality and asymmetric random walks

Central classes using PageRank will be those that heavily cite other globally central classes (i.e. classes i that have a large $f_{k\leftarrow i}$ for other central classes k) and tend to also be heavily cited by important classes due to the near symmetry of the citation network. However, classes central to the domestic economy (rather than the global economy) will be heavily cited and simultaneously have a smaller fraction of citations originating from foreign economies (so that information does not flow as readily across the border). This can be modelled in the spirit of PageRank by introducing an asymmetric and border-dependent discount rate, equivalent to a non-uniform teleportation probability [30,33] if the loss is coupled with a reinsertion. Our model permits four types of information flow: domestic-to-domestic (with an associated loss term $\epsilon_{d\to d}$), foreign-to-domestic ($\epsilon_{f\to d}$), domestic-to-foreign ($\epsilon_{d\to f}$) and foreign-to-foreign ($\epsilon_{f\to f}$). The loss probability ϵ_0 reflects the expectation that a patent is not guaranteed to cause a beneficial spillover, as not all patents provide a beneficial spillover to another sector (due to being

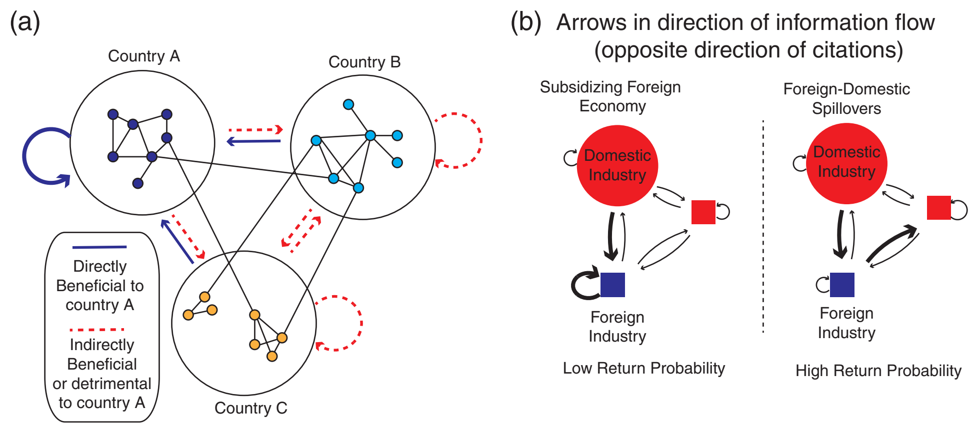

FIG. 1. In (a), the patent classes for each country are denoted by nodes and citations between the classes by thin edges. Information flow from domestic classes towards foreign classes (indicated by the dashed arrows) may be less beneficial to the domestic economy than information remaining in the economy. This expectation does not imply that information to a foreign economy is guaranteed to be lost, as shown in (b). If a domestic industry is funded (the large circle) and is heavily cited by a foreign class, the information will likely move across the border (the thick arrow). The likelihood of the information returning depends on the network topology.

uncited or the usefulness of the invention being reduced with time). From the perspective of the domestic economy, there may be an expectation that spillovers into a foreign economy are less beneficial than domestic spillovers (schematically diagrammed in Fig. 1). For example, the potential for a new company to be created in a foreign economy due to a cross-border spillover could reduce the expectation of benefit to the national economy relative to a domestic-to-domestic citation. Alternatively, language or cultural barriers that are not present domestically may reduce the expected domestic benefit of a foreign spillovers. In this paper, we focus on a simplified model of information sharing where flow into the domestic economy is considered 'good' (so $\epsilon_{d\to d} = \epsilon_{f\to d} \equiv \epsilon_0$ for the domestic-to-domestic and foreign-to-domestic discount rates), and information sharing benefiting foreign classes is considered 'bad' (so $\epsilon_0 \leqslant \epsilon_{f\to d} = \epsilon_{f\to f} \leqslant 1$ for the domestic-to-foreign and foreign-to-foreign discount rates). We denote the elevation in the discount rate for border crossing by the parameter $0 \leqslant \epsilon \leqslant 1$ and define $\epsilon_{d\to f} = \epsilon_{f\to f} = 1-(1-\epsilon)(1-\epsilon_0) \equiv \bar{\epsilon}$ (satisfying $\epsilon_0 \leqslant \bar{\epsilon} \leqslant 1$, so foreign spillovers are less beneficial than domestic spillovers).

In the previous paragraph, we have defined a transition probability between nodes that has a finite probability of the random walker becoming lost, which depends on the relationship between the start- and end-points of the walker's step in the global citation network (whether it is domestic-to-domestic, domestic-to-foreign, foreign-to-domestic or foreign-to-foreign). The walker must be reinserted into the network in order for a steady state to be attainable [15,30,42]. Since we are modelling the centrality of a class from the perspective of a particular sub-network (a national economy) with a step of the random walk representing a patent, it is natural to reinsert the walker only within that sub-network after loss rather than throughout the entire global network. For example, we might reasonably expect that a national government might provide an external stimulus in a patent citation network only to domestic classes and never directly to foreign classes. Such a choice for a teleportation probability is equivalent to a topic-sensitive or personalized PageRank [12,43]. The insertion, walk and loss are (schematically diagrammed in Fig. 1) schematically diagrammed in Fig. 2. If the insertion after loss is uniformly distributed amongst domestic classes, it is straightforward to show that the transition probabilities between

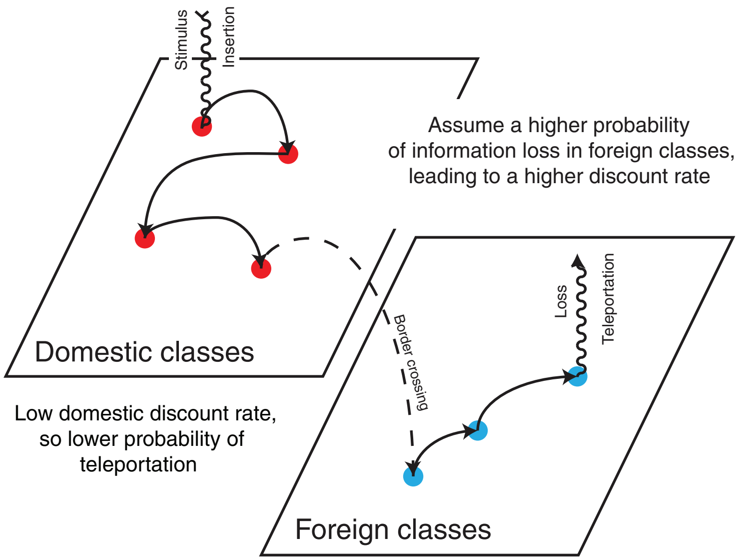

FIG. 2. In our model, the walker is always inserted into the domestic economy, since the national government will only fund domestic classes directly. While propagating within the national economy a low discount rate is assumed, but once the information flows into the foreign economy the discount rate is elevated. The future value is discounted at a higher rate through an increase the teleportation probability in foreign economies, with the random walk suddenly ending and the walker being reinserted as a new investment in the domestic economy.

classes becomes

$$p_{i\to j}^{d\to d}(\epsilon) = (1-\epsilon_0) f_{i\leftarrow j} + \frac{\epsilon_0}{n_d}\left(f_{i\leftarrow D} + \frac{\bar{\epsilon}}{\epsilon_0} f_{i\leftarrow F}\right), \tag{4.1}$$

$$p_{k\to j}^{f\to d}(\epsilon) = (1-\epsilon_0) f_{k\leftarrow j} + \frac{\epsilon_0}{n_d}\left(f_{k\leftarrow D} + \frac{\bar{\epsilon}}{\epsilon_0} f_{k\leftarrow F}\right), \tag{4.2}$$

$$p_{i\to l}^{d\to f}(\epsilon) = (1-\bar{\epsilon}) f_{i\leftarrow l}, \tag{4.3}$$

$$p_{k\to l}^{f\to f}(\epsilon) = (1-\bar{\epsilon}) f_{k\leftarrow l}, \tag{4.4}$$

where $f_{i\leftarrow D} = \sum_{j\in\mathbf{D}} f_{i\leftarrow j}$ is the fraction of citations received by f from any class in the domestic economy (denoted by $\mathbf{D}$), $f_{i\leftarrow F} = \sum_{l\in\mathbf{F}} f_{i\leftarrow l}$ is the fraction of citations from foreign classes (denoted by $\mathbf{F}$) and n_d is the number of classes in the domestic economy. To avoid traps, we assign a 100% teleportation probability for classes not simultaneously citing and having been cited by at least one class in the connected component by a minimum of two citations. In an earlier version of this work [27], this restriction was relaxed to a constraint on the number citations (not received citations), which in some cases leads to small quantitative differences but no quantitative changes to the results. The aggregation of smaller economies into a single WO 'region' can affect the path in the random walk model only if the walker (a) enters the WO region during the walk and (b) remains within the WO region for at least one step (so that there is ambiguity as to whether a border crossing has been ignored by the artificial aggregation). As the average fraction of citations received by the larger economies is $\approx 12.5\%$ and the fraction of WO-to-WO citations is $\approx 40.6\%$, we estimate that a fraction $0.125 \times 0.406 \approx 0.05$ of the random walk paths are likely to be perturbed by the aggregation.

While our network of interacting national innovation systems lacks the significant qualitative differences in the meaning of the layers seen in other studies (such as the power grid/internet interactions in ref. [25] or the florae/faunae interdependencies of ref. [44]), the concepts of interdependent networks or network-of-networks have proved useful even for the division of a single network into layers of qualitatively similar nodes or with qualitatively similar edges between layers in both theoretical [17,18] and real world [19,20] studies. We note that in the context of information flow through networks-of-networks with an inherent asymmetry when crossing between layers, we expect our approach to be potentially useful. In this particular case, we wish to emphasize that the division of the global innovation network into distinct layers is natural from the perspective of a domestic economy due to competition between nations, despite the fact that there is no inherent qualitative difference between a foreign or domestic citation.

All foreign and domestic classes can be assigned a domestic centrality ranking from the perspective of each country based on the steady-state probability of finding the walker at that node. Defining the class' rank $R_i^{(d)}(\epsilon)$ as the probability that the domestic class i is occupied at steady state during the random walk and $R_k^{(f)}(\epsilon)$ as the steady-state probability that foreign class k is occupied at steady state, it is straightforward to show that

$$R_i^{(d)}(\epsilon) = \sum_{j\in\mathbf{D}} p_{j\to i}^{d\to d}(\epsilon) R_j^{(d)}(\epsilon) + \sum_{l\in\mathbf{F}} p_{l\to i}^{f\to d}(\epsilon) R_l^{(f)}(\epsilon), \tag{4.5}$$

$$R_k^{(f)}(\epsilon) = \sum_{j\in\mathbf{D}} p_{j\to k}^{d\to f}(\epsilon) R_j^{(d)}(\epsilon) + \sum_{l\in\mathbf{F}} p_{l\to i}^{f\to f}(\epsilon) R_k^{(f)}(\epsilon). \tag{4.6}$$

The centrality of domestic classes can be determined without reference to the centrality of foreign classes through the eigenvalue equation

$$\mathbf{R}^{(d)}(\epsilon) = (\mathbf{P}^{d\to d}(\epsilon) + \mathbf{P}^{f\to d}(\epsilon)[\mathbf{1} - \mathbf{P}^{f\to f}(\epsilon)]^{-1}\mathbf{P}^{d\to f}(\epsilon))\mathbf{R}^{(d)}(\epsilon), \tag{4.7}$$

where $[\mathbf{P}^{x\to y}(\epsilon)]_{ij} = p_{j\to i}^{x\to y}(\epsilon)$ is one of the four matrices of transition probabilities and $\mathbf{R}^{(d)}(\epsilon)$ the vector of centralities of domestic classes. The domestic centrality $\mathbf{R}^{(d)}(\epsilon)$ is thus an eigenvector of a matrix involving the domestic-to-domestic transition probabilities coupled with a convolution of the transition probabilities involving foreign economies. This eigenvalue problem has a variety of well-known methods of solution [45], and we implement a simple power method to numerically determine $\mathbf{R}^{(d)}$. The eigenvector is normalized such that $\sum_i R_i^{(d)}(\epsilon) = 1$ for all countries and all ϵ.

The upper and lower bounds on the centrality of each country's innovation network will not in general be the same but rather depend strongly on the loss probability and the ratio of domestic-to-foreign citations in the former case and the skewness of the edge weight distribution in the latter case. In order to compare the variability of the importance of the same class between countries or for varying values of ϵ, it will often be useful to examine the scaled centrality

$$\mu_i = \frac{R_i^{(d)} - \langle R^{(d)}\rangle}{\sigma_d}, \tag{4.8}$$

where $\langle R^d\rangle = n_d^{-1}\sum_i R_i \equiv n_d^{-1}$ the mean centrality of the nodes in the domestic economy (the latter equality due to our choice of normalization) and $\sigma_d^2 = (n_d - 1)^{-1}\sum_i (R_i^{(d)} - n_d^{-1})^2$ is the variance of the centralities about the mean within the domestic economy. Large values of μ_i correspond to very central classes of the domestic economy, and similar values of μ_i in different countries indicate they are of similar significance in both countries, regardless of the upper and lower bounds on R_i.

5. Results for the patent citation network

We apply our measure of centrality to the patent citation network described in Section 2 and choose a small value for the domestic loss probability $\epsilon_0 = 0.01$. This choice is largely arbitrary in this paper and indicates an expectation that most innovations are likely to produce spillovers somewhere in the global economy at some point in the future. While it may seem possible in principle to estimate a ‘correct’ value of ϵ_0 from the data as the fraction of patents that receive no citations, the fact that newer patents are more likely to be uncited than older patents (as the latter have had more time to acquire citations) makes such an estimation difficult. The robustness of our model to this parameter is explored in detail in Section 6. Choosing $\epsilon_0 = 0.15$ (the typical value used in the PageRank algorithm) would indicate an expectation of shorter paths (with a 15% chance of a patent not useful to the global economy). The behaviour of important classes (those classes rated in the top 5 at either $\epsilon = 0$ or $\epsilon = 1$) are shown in Fig. 3 for 1995 and 2005. For comparison, in Fig. 4 we show the scaled centrality defined in Equation (4.8), which highlights the relative differences between these most important classes of the domestic economies. Only 21 classes appear in any country’s top-5 list, as shown in the legends of Figs 3 and 4, listed in descending order of the number of countries for which the class is central. ‘Organic Chemistry’ and ‘Medical or Veterinary Science’ are of overwhelming importance to most economies and are typically the two most central classes (with the prominent exception of JP), but there is greater variety in the remaining classes in the remaining list of central classes for each nation.

We observe a decrease in the centrality of high-ranked classes at $\epsilon = 0$ in most cases due to the overall increase in the probability of loss at higher ϵ in Fig. 3. Because the walker is reinserted randomly within the domestic economy, a higher overall probability of loss necessarily increases the centrality for unimportant classes (e.g. the centrality of a disconnected domestic node increases monotonically with ϵ) and therefore decreases the numerical value of R_i for important classes because of the normalization of

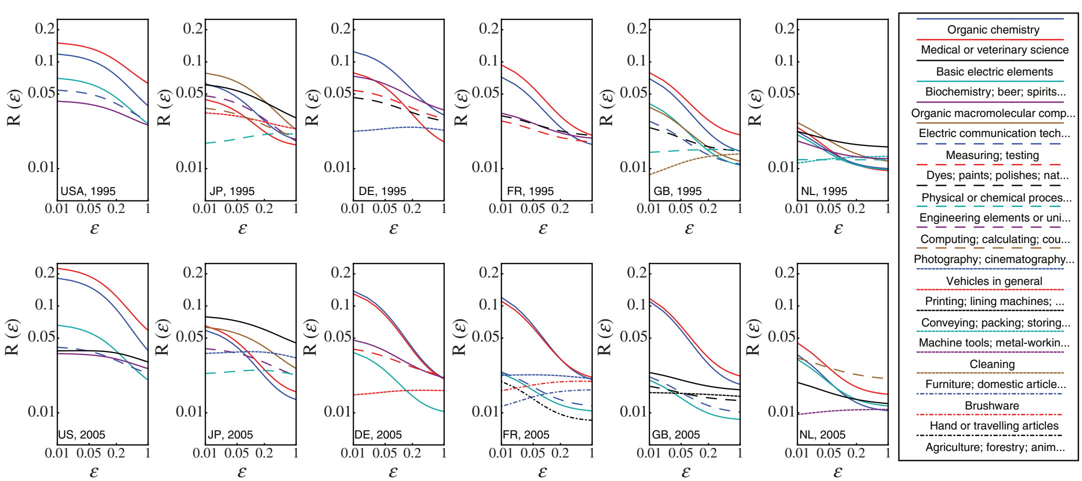

FIG. 3. Shown are the centralities for the most important global or domestic classes for each country (those in the top-5 lists at either $\epsilon = 0$ or $\epsilon = 1$) as a function of ϵ in 1995 (a) and 2005 (b) plotted on log–log axes. The decline observed in the most central classes is due to an increase in the overall loss probability at higher ϵ.

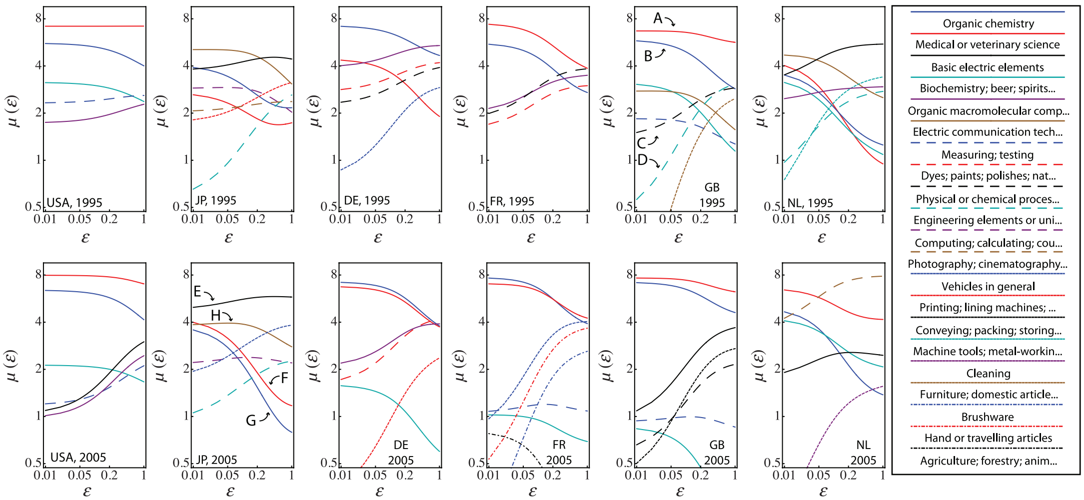

FIG. 4. Shown is the behaviour of scaled centrality as a function of ϵ for the same classes as shown in Fig. 3, plotted on log–log axes. The additional labels A–D in the GB-1995 panel and E–G in the JP-2005 panel correspond to the same labels in Fig. 5 and Table 3, used as exemplars of the sometimes significant differences of globally and domestically central classes.

the eigenvector. The *relative* change in centrality when comparing classes is entirely due to the topology of the citation network and is a signal of the difference between global and domestic centrality, most clearly indicated by the scaled centrality in Fig. 4. It is often the case that very central classes globally (at $\epsilon = 0$) are of significantly lower importance domestically (at $\epsilon = 1$). The primary exceptions to this trend are the 'Organic Chemistry' and 'Medical or Veterinary Science' in many countries (consistent with their place as most commonly central for all countries, discussed above). While there are multiple cases where these two classes do not depart from the top 5 for any ϵ, it is particularly interesting to note that there is negligible change in μ_i for 'Medical or Veterinary Science' in the USA in 1995, indicating that its importance to the US economy is essentially independent of the model parameters.

Additional details for the most central classes in Figs 3 and 4 are listed in Table 1 for 1995 and Table 2 for 2005. While neither the extreme of a fully open economy (with no expected detriment to information sharing across borders and $\epsilon = 0$) nor a fully protectionist economy (with no expected benefit from information sharing across borders and $\epsilon = 1$) is likely to be optimal, the differences between these extremes in the tables illustrate the potential variance in the centrality for intermediate values of ϵ. It is clear that the more heavily cited classes tend to be central globally (the left-hand side of the tables), which is unsurprising since at $\epsilon = 0$ the number of citations is the only determinant of centrality. At $\epsilon = 1$, the influence of the borders becomes significant, as indicated by the increases in the columns for $f_{i\leftarrow D}$ and $f_{i\rightarrow D}$ (the fraction of forward and backward domestic citations). That is, classes that give or receive a high fraction of domestic citations tend to be more central domestically. While this bias towards domestic citations is expected from our model, our centrality does not simply return the domestic classes with the most citations. Globally central classes are sometimes large enough to remain central domestically (e.g. compare JP's 'Electronic Communication Techniques' with 'Engineering Elements', the third and fourth most domestically central classes in 2005, Table 2), while local topology sometimes has a dominant effect on the centrality (e.g. compare JP's 'Basic Electric Elements' with 'Electronic Communication Techniques', the first and third most domestically central classes in 1995, Table 1).

The centralities and scaled centralities for each class in each country are shown in Fig. 5(a), (b) and

TABLE 1 *Details of the top-5 ranked classes in 1995 are listed for $\epsilon = 0$ (left) and $\epsilon = 1$ (right), with $\epsilon_0 = 0.01$ throughout. $\mu_i = (R_i - \langle R \rangle)/\sigma_R$ is the scaled centrality of each class, W is the total weight (sum of the received citations) received by that class from patents filed in 2005 (excluding those from the WO category of countries not individually considered) and $f_{i\leftarrow D}$ and $f_{i\rightarrow D}$ are the fraction of citations incoming from and outgoing towards the domestic economy, respectively. For $\epsilon = 0$, the more commonly patented classes tend to be highly ranked, while for $\epsilon = 1$ there is a bias towards patent classes that cite and are cited by other domestic classes*

USA, year = 1995, $\epsilon_0 = 0.01$. $\langle W \rangle = 663.479$, $W_{\max} = 11894$									
μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
7.17	11.9k	75	74	Medical or veterinary scien...	7.18	11.9k	75	74	Medical or veterinary scien...
5.6	10.5k	73	69	Organic chemistry	4.01	10.5k	73	69	Organic chemistry
3.15	6.3k	77	73	Biochemistry; beer; spirits...	2.59	4.9k	74	67	Measuring; testing
2.32	4.9k	74	67	Measuring; testing	2.37	6.3k	77	73	Biochemistry; beer; spirits...
1.73	4.3k	61	55	Organic macromolecular comp...	2.28	4.3k	61	55	Organic macromolecular comp...
JP, year = 1995, $\epsilon_0 = 0.01$. $\langle W \rangle = 338.975$, $W_{\max} = 3537$									
μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
5.07	3.5k	64	59	Electric communication tech...	4.43	2.9k	63	55	Basic electric elements
4.07	2.9k	50	51	Organic chemistry	3.14	1.5k	75	72	Printing; lining machines; ...
3.74	2.9k	63	55	Basic electric elements	3.07	3.5k	64	59	Electric communication tech...
2.87	2.3k	57	54	Computing; calculating; cou...	2.61	0.4k	58	67	Engineering elements or uni...
2.72	2.1k	44	41	Medical or veterinary scien...	2.37	1.7k	73	63	Photography; cinematography...
DE, year = 1995, $\epsilon_0 = 0.01$. $\langle W \rangle = 226.967$, $W_{\max} = 4143$									
μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
7.25	4.1k	46	59	Organic chemistry	5.4	2.9k	46	56	Organic macromolecular comp...
4.45	2.7k	39	49	Medical or veterinary scien...	4.67	4.1k	46	59	Organic chemistry
3.91	2.9k	46	56	Organic macromolecular comp...	4.2	1.8k	52	65	Dyes; paints; polishes; nat...
2.74	1.8k	52	65	Dyes; paints; polishes; nat...	3.91	1.4k	46	59	Physical or chemical proces...
2.26	1.4k	46	59	Physical or chemical proces...	2.91	0.5k	52	60	Vehicles in general
FR, year = 1995, $\epsilon_0 = 0.01$. $\langle W \rangle = 101.702$, $W_{\max} = 1852$									
μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
7.52	1.9k	38	45	Medical or veterinary scien...	3.87	1.9k	38	45	Medical or veterinary scien...
5.63	1.6k	40	43	Organic chemistry	3.84	0.6k	48	51	Physical or chemical proces...
2.01	0.8k	36	42	Organic macromolecular comp...	3.49	0.8k	36	42	Organic macromolecular comp...
1.82	0.6k	48	51	Physical or chemical proces...	2.99	0.4k	33	55	Dyes; paints; polishes; nat...
1.58	0.4k	33	55	Dyes; paints; polishes; nat...	2.71	1.6k	40	43	Organic chemistry
GB, year = 1995, $\epsilon_0 = 0.01$. $\langle W \rangle = 105.545$, $W_{\max} = 1833$									
μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
6.69	1.8k	43	38	Medical or veterinary scien...	5.66	1.8k	43	38	Medical or veterinary scien...
5.89	1.8k	40	32	Organic chemistry	3.06	0.2k	49	57	Engineering elements or uni...
3.13	1.k	34	33	Biochemistry; beer; spirits...	2.9	0.4k	33	35	Physical or chemical proces...
2.76	0.9k	20	29	Electric communication tech...	2.86	1.8k	40	32	Organic chemistry
1.84	0.7k	34	27	Measuring; testing	2.48	25	92	43	Furniture; domestic article...
NL, year = 1995, $\epsilon_0 = 0.01$. $\langle W \rangle = 34.2479$, $W_{\max} = 371$									
μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
4.74	0.4k	20	20	Electric communication tech...	5.52	0.2k	38	28	Basic electric elements
4.4	0.3k	23	22	Medical or veterinary scien...	3.4	18	83	28	Machine tools; metal-workin...
3.78	0.3k	24	15	Organic chemistry	2.95	0.3k	25	28	Organic macromolecular comp...
3.38	0.2k	24	16	Biochemistry; beer; spirits...	2.75	38	32	35	Engineering elements or uni...
3.08	0.2k	38	28	Basic electric elements	2.51	0.4k	20	20	Electric communication tech...

TABLE 2 *The same as in Table 1, but for classes in* 2005

USA, year = 2005, $\epsilon_0 = 0.01$. $\langle W \rangle = 548.413$, $W_{\max} = 13635$

μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
7.98	13.6k	71	69	Medical or veterinary scien...	7.03	13.6k	71	69	Medical or veterinary scien...
6.41	12.0k	67	66	Organic chemistry	4.15	12.k	67	66	Organic chemistry
2.13	5.7k	74	65	Biochemistry; beer; spirits...	3.00	3.0k	58	58	Basic electric elements
1.18	3.5k	76	61	Measuring; testing	2.44	2.6k	64	62	Organic macromolecular comp...
1.04	3.0k	58	58	Basic electric elements	2.11	3.5k	76	61	Measuring; testing

JP, year = 2005, $\epsilon_0 = 0.01$. $\langle W \rangle = 375.653$, $W_{\max} = 4139$

μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
4.81	4.1k	71	67	Basic electric elements	5.79	4.1k	71	67	Basic electric elements
4.27	3.0k	44	47	Medical or veterinary scien...	3.83	1.5k	74	77	Vehicles in general
3.84	3.0k	46	44	Organic chemistry	2.79	3.0k	65	74	Electric communication tech...
3.79	3.0k	65	74	Electric communication tech...	2.25	0.8k	74	77	Engineering elements or uni...
2.15	2.0k	58	71	Computing; calculating; cou...	2.2	2.0k	58	71	Computing; calculating; cou...

DE, year = 2005, $\epsilon_0 = 0.01$. $\langle W \rangle = 180.14$, $W_{\max} = 4242$

μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
7.26	4.2k	41	46	Organic chemistry	4.18	1.1k	49	54	Dyes; paints; polishes; nat...
6.79	3.6k	35	46	Medical or veterinary scien...	3.9	1.4k	50	58	Organic macromolecular comp...
2.08	1.4k	50	58	Organic macromolecular comp...	3.79	4.2k	41	46	Organic chemistry
1.6	1.1k	27	38	Biochemistry; beer; spirits...	3.74	3.6k	35	46	Medical or veterinary scien...
1.59	1.1k	49	54	Dyes; paints; polishes; nat...	2.36	0.3k	59	77	Printing; lining machines; ...

FR, year = 2005, $\epsilon_0 = 0.01$. $\langle W \rangle = 89.6116$, $W_{\max} = 2486$

μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
7.74	2.5k	30	43	Organic chemistry	4.25	2.4k	34	43	Medical or veterinary scien...
7.06	2.4k	34	43	Medical or veterinary scien...	4.03	0.2k	25	42	Vehicles in general
1.05	0.4k	26	44	Measuring; testing	3.92	2.5k	30	43	Organic chemistry
1.03	0.6k	20	34	Biochemistry; beer; spirits...	3.66	50	82	80	Hand or travelling articles
0.81	0.5k	16	45	Agriculture; forestry; anim...	2.61	23	100	82	Brushware

GB, year = 2005, $\epsilon_0 = 0.01$. $\langle W \rangle = 89.7603$, $W_{\max} = 2582$

μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
7.66	2.6k	38	32	Medical or veterinary scien...	6.25	2.6k	38	32	Medical or veterinary scien...
7.22	2.4k	40	32	Organic chemistry	4.61	2.4k	40	32	Organic chemistry
0.94	0.4k	39	38	Basic electric elements	3.69	0.4k	39	38	Basic electric elements
0.92	0.6k	33	21	Measuring; testing	2.71	65	32	48	Conveying; packing; storing...
0.86	0.8k	44	19	Biochemistry; beer; spirits...	2.15	0.4k	35	37	Physical or chemical proces...

NL, year = 2005, $\epsilon_0 = 0.01$. $\langle W \rangle = 43.6198$, $W_{\max} = 731$

μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class	μ	W	$f_{i\leftarrow D}$ (%)	$f_{i\rightarrow D}$ (%)	Class
6.6	0.7k	15	28	Medical or veterinary scien...	7.87	0.3k	42	55	Photography; cinematography...
4.98	0.7k	15	26	Organic chemistry	4.16	0.7k	15	28	Medical or veterinary scien...
4.26	0.6k	16	33	Biochemistry; beer; spirits...	2.46	0.3k	28	24	Basic electric elements
3.61	0.3k	42	55	Photography; cinematography...	2.08	0.6k	16	33	Biochemistry; beer; spirits...
1.73	0.3k	28	24	Basic electric elements	1.56	15	53	76	Cleaning

(c), (d), respectively, showing the global centrality (at $\epsilon = 0$) on the x-axes and domestic centrality (at $\epsilon = 1$) on the y-axes. The global and domestic centralities are strongly related, with $R_i(\epsilon = 1)$ typically falling within a factor of ~ 4 of the global centrality of $R_i(\epsilon = 1)$ over two decades of data. The scaled centrality more directly emphasizes the large impact on the rankings of very central classes due to the

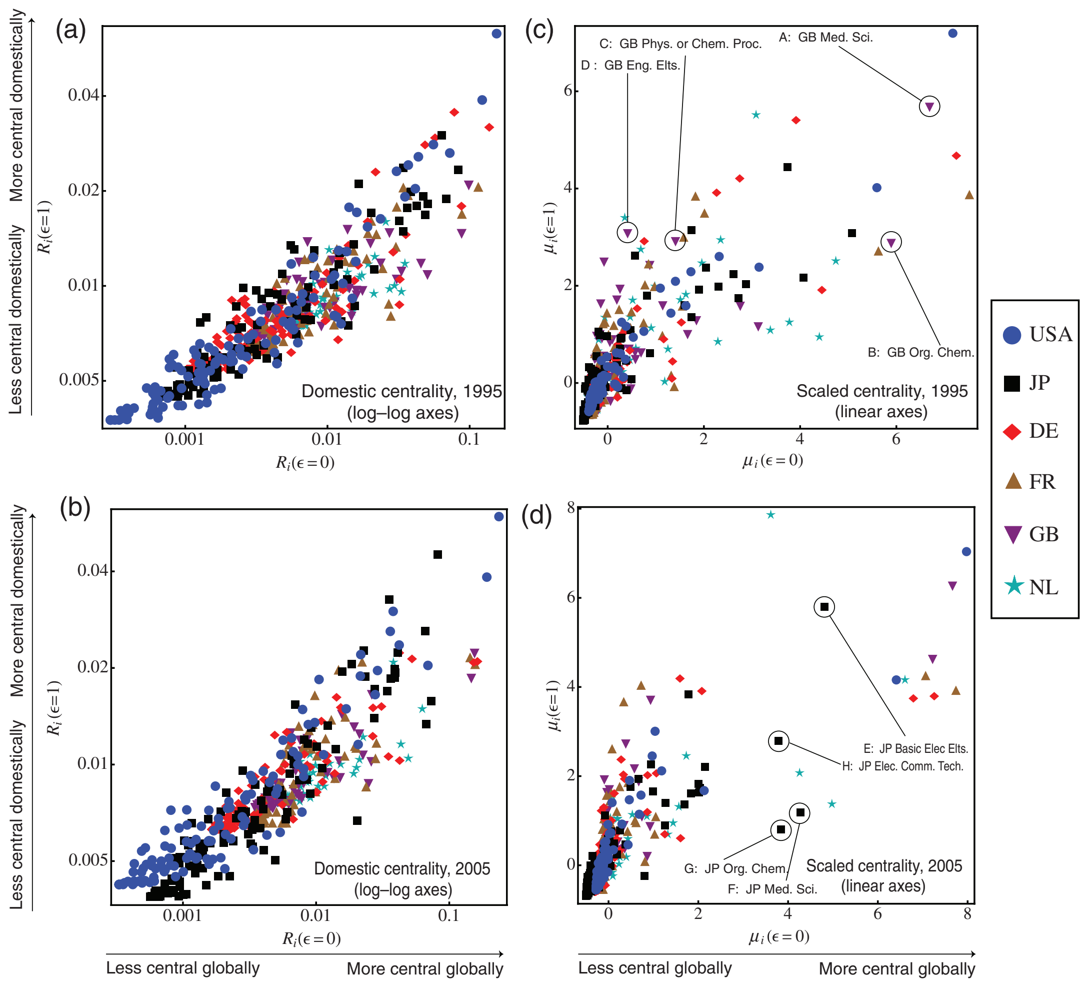

FIG. 5. Shown is a comparison of the centralities for all classes and countries in 1995 (a) and 2005 (b) and the scaled centralities in (c) and (d) in the two extremes $\epsilon = 0$ (measuring global centrality, x-axis) and $\epsilon = 1$ (measuring domestic centrality, y-axis) with the domestic discount rate $\epsilon_0 = 0.01$. While there is a clear correlation between globally and domestically important classes in (a) and (b), the wide variance in the scaled centralities are evident in (c) and (d) even for very central classes. The labels A–H in the scaled centrality figures correspond to the same labels in Fig. 4 and Table 3 (with the labels A–D corresponding to the top-4 GB classes at $\epsilon = 0$ in 1995 and the labels E–H corresponding to the top-4 JP domestic classes in 2005).

variation of a factor of ~ 4 (particularly when shown with linear axes). We highlight a few classes for two countries in particular: the strong variations of the four most domestically central classes for GB in 1995 (marked with labels A–D in Fig. 5(c)) and of the four most globally central classes for JP in 2005 (marked with labels E–H in Fig. 5(d)). Table 3 shows the extreme variations in ranking that can occur as the loss probability ϵ is varied using these classes as exemplars, with the 16-highest global GB class at $\epsilon = 0$ ranked second domestically (due to the high fraction of domestic citations, as listed in Table 1). Likewise, the second-highest class of global importance for JP in 2005 drops to 15th place as ϵ increases due to its relatively low fraction of domestic citations (see Table 2). The variations illustrated in Fig. 5

TABLE 3 *Details of the highlighted top-*5 *classes in Figs* 4 *and* 5 *and their global* ($\epsilon = 0$) *and domestic* ($\epsilon = 1$) *ranks. A–D correspond to the four most domestically central GB classes in* 1995, *and* (*E–H*) *correspond to the four most globally central JP classes. Classes that are globally unimportant* (*like GB Engineering Elements*) *may be very central domestically, and likewise globally central classes* (*like JP's Organic Chem.*) *may have a low domestic centrality*

Label	Class	Global rank	Domestic rank
A	GB Med. Sci.	1	1
B	GB Organic Chem.	2	4
C	GB Phys. or Chem. Proc.	7	3
D	GB Engineering Elts.	16	2
E	JP Basic Elec. Elts.	1	1
F	JP Med. Sci.	2	15
G	JP Organic Chem.	3	18
H	JP Elec. Comm. Techniques	4	3

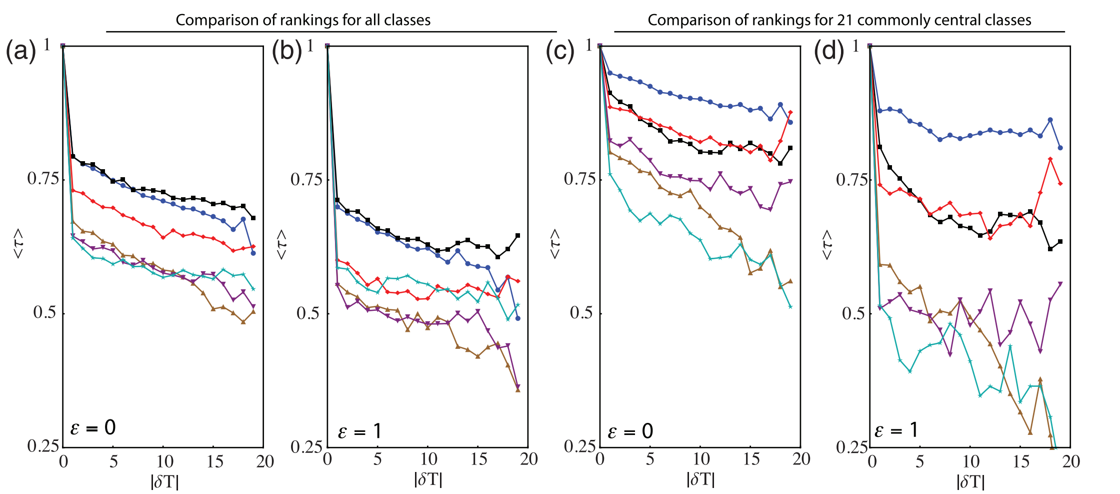

FIG. 6. The Kendall-τ coefficient for the centralities of all classes at $\epsilon = 0$ (a) and $\epsilon = 1$ (b), and for the 21 major classes listed in Figs 3–4 and Tables 1 and 2 at $\epsilon = 0$ (c) and $\epsilon = 1$ (d). The sharp initial decline seen in (a and b) is greatly reduced in (c), indicating that globally central classes are slow to change. The initial decline of domestically central classes may be either very slow or very fast depending on the country, shown in (d). The symbols are the same as in Fig. 5.

are large in some cases, but an examination of the topology of the citation network makes it clear that our model captures a meaningful difference between globally and domestically important classes.

We find the most central classes in the USA and JP are predominantly the same in 1995 and 2005 (over 66% of the members of the top-5 classes are found in both years), but there is a drastic restructuring observed for both FR and the NL (with below 40% of the top-5 classes the same in both years). This suggests that the class centrality in smaller economies may be less stable than those in larger economies as time progresses, which is confirmed by examining the Kendall-τ coefficient [46] (measuring the fraction of pairs of values in each list that have the same ordering) between centralities in different years for each country. In Fig. 6(a) and (b), the average of the Kendall-τ coefficient for each pair of years T_1 and T_2 is shown as a function of the time interval between them, $|\delta T| = |T_1 - T_2|$. Regardless of the value of ϵ, Fig. 6(a) and (b) shows a steep drop by about 20–40% at $|\delta T| = 1$ followed by a slow decline on the scale of decades for rankings for $|\delta T| > 1$. The initial drop at $|\delta T| = 1$ is due primarily to changes in classes with low centrality as can be seen by restricting the Kendal-τ average to only the 21 important classes that are found in the top 5 for any country (those listed in Figs 3 and 4 and Tables 1 and 2). The restricted average still shows a similar slow decline for global centralities in Fig. 6(c) with $|\delta T| > 1$ but with a greatly reduced initial decline at $|\delta T| = 1$, indicating a smaller change in the relative ordering of these more central classes. For the domestic centralities (at $\epsilon = 1$) in Fig. 6(d), the sharp decline at $|\delta T| = 1$ is only reduced for the larger economies of USA, JP and DE. Smaller economies retain their large initial decline in the 21 most central classes and fluctuate more heavily as $|\delta T|$ increases.

6. Robustness: parameter variation and noise

It is valuable to understand the robustness of the model to variations in the parameters or aggregation and to statistical noise. In Fig. 7, we compare the centralities using two different values of the domestic loss parameter ϵ_0 using the correlation coefficient and Kendall-τ (determining the fraction of pairs of rankings that differ between the two orderings [46]) as a function of ϵ. We compare the values of $\epsilon_0 = 0.01$, the loss probability considered throughout the paper, with the commonly used PageRank teleportation probability of $\epsilon_0 = 0.15$. While these loss probabilities differ by an order of magnitude,

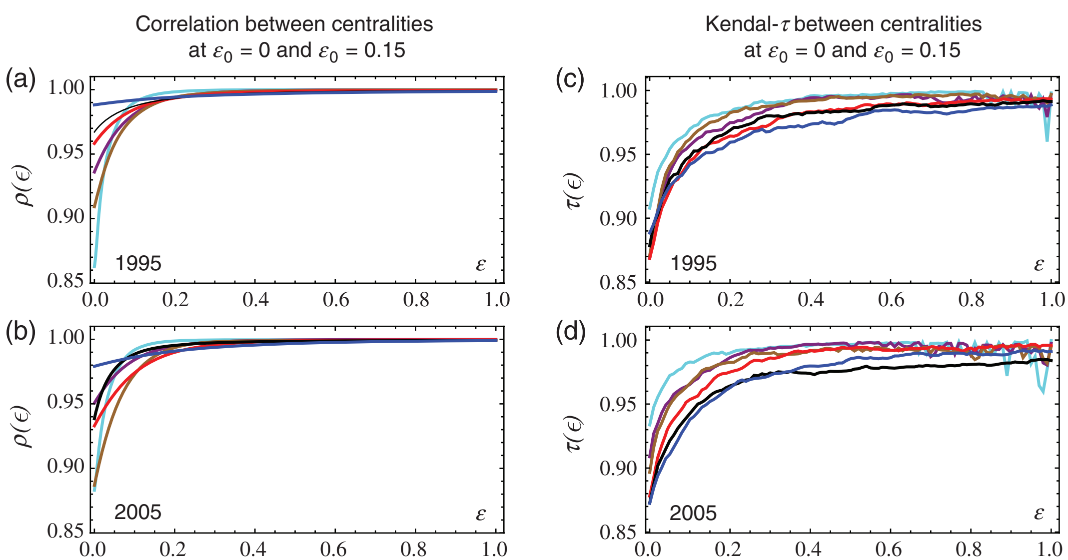

FIG. 7. As a robustness check on the method, we quantify the similarity between the centrality vector $\mathbf{R}^{(d)}(\epsilon)$ for two different teleportation probabilities: $\epsilon_0 = 0.01$ (the very low value used throughout this paper) and $\epsilon_0 = 0.15$ (the traditional PageRank value). The correlation coefficient and the Kendall-τ coefficients are shown in (a and b) and (c and d), respectively, for varying ϵ using the same colorings as Figs 5 and 6. (a) and (c) show 1995, (b) and (d) show 2005. There is a good agreement between the centralities for the global measure (near $\epsilon = 0$) which improves with increasing ϵ, approaching near-perfect agreement at $\epsilon \sim 1$.

the agreement between centrality measures is very high: the correlation coefficient ρ above $\sim$0.9 and Kendall τ above 0.85 for $\epsilon = 0$ and improves steadily as ϵ increases. This increase is due to the rarity of teleportation events at $\epsilon = 0$ compared with $\epsilon = 1$ (where in the latter case border crossings boost the overall probability of teleportation) and indicates that the domestic centrality should be considered at least as robust as the well-accepted topic-sensitive or personalized [12,43] PageRank centralities to variations in the model parameters.

We also examine the effect of changing the temporal aggregation of the citation network by replacing the citations between classes filed in year T, $c_{i\leftarrow j}(T)$, with that of a 3-year total $c^*_{i\leftarrow j}(T) = c_{i\leftarrow j}(T-1) + c_{i\leftarrow j}(T) + c_{i\leftarrow j}(T+1)$. A comparison between the centralities for all classes and countries for the 1- and 3-year aggregations is shown in Fig. 8 for $T = 1995$, and that there is a strong correlation between the centralities for both citation networks. Consistent with the results of Fig. 6, temporal variations have a greater effect on domestic centrality than global centrality (as evidenced by the greater variance in Fig. 8(b)), and the centrality of unimportant classes tends to be more strongly effected than those of highly central classes.

As a final check on the robustness of the model, we simulate the effect of noise in the number of citations between classes by randomly rewiring the observed network and recomputing the centralities of the classes. The rewiring is diagrammed in Fig. 9(a): a pair of edges connecting four distinct nodes are randomly chosen and the locations of their end-points are swapped. This rewiring procedure guarantees that the degree of each node is preserved since no edges are added or removed (just exchanged) but does not preserve the fraction of domestic or foreign citations. We therefore expect the rewire to have a small effect on the global centrality (at $\epsilon = 0$) and a more significant effect on domestic centrality (at $\epsilon = 1$) where the nationality of the citations are relevant. This expectation is confirmed in Fig. 9(b) and (c), which shows the average root mean square deviation (RMSD) between the rewired centralities and the unmodified centralities in 1995: $\langle \delta R(\epsilon)\rangle = n_d^{-1}\sum_i [n_r^{-1}[\sum_r (R_{i,r}(\epsilon) - R_i(\epsilon))^2]^{1/2}$ with $R_{i,r}(\epsilon)$ the

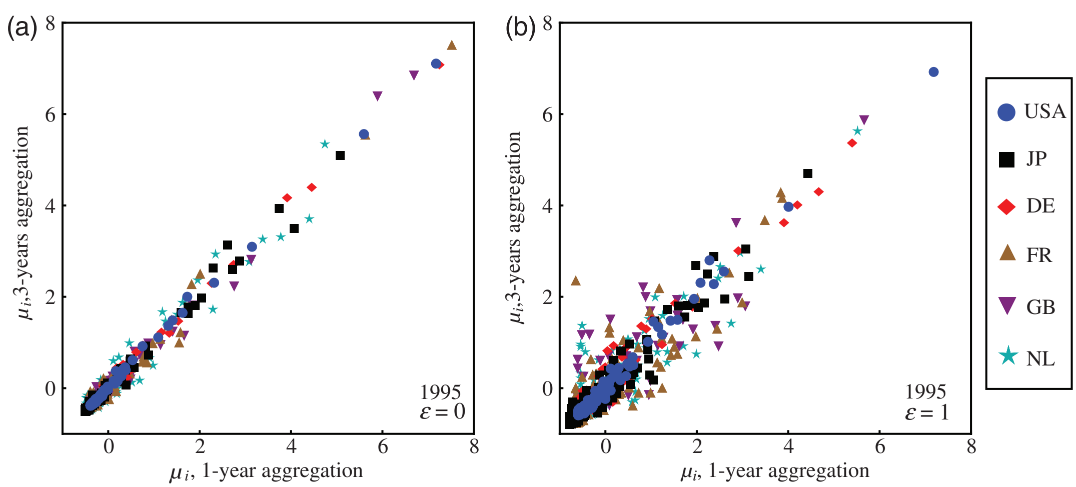

Fig. 8. Shown is a comparison of centralities as different levels of temporal aggregation. The x-axes correspond to node centrality, computed from networks based on citations to classes in 1995 only, while the y-axis corresponds to centralities for a temporal aggregation of 1994, 1995 and 1996 (a 3-year window). The scaled global (a) and domestic (b) centralities are highly correlated for central classes (those with large μ_i), but the domestic centrality shows greater variability for non-central classes (those with $\mu_i \lesssim 0$).

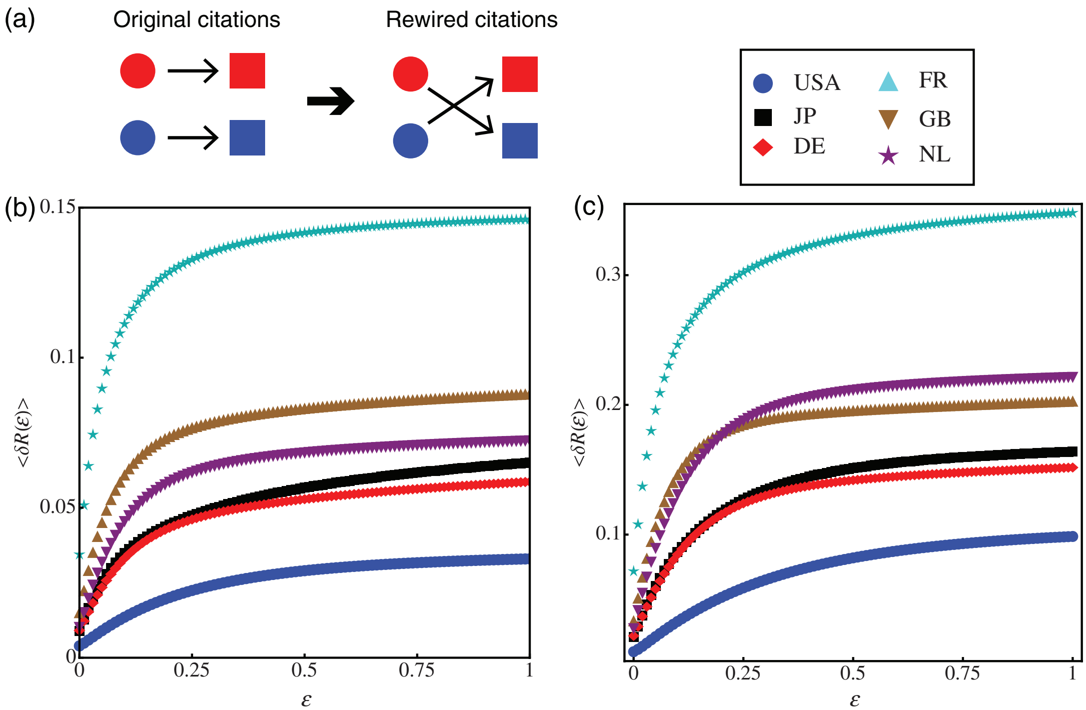

Fig. 9. Shown is the robustness of our measure of centrality under a degree-preserving random rewiring, schematically diagrammed in (a). (b) shows the RMSD difference $\langle \delta R(\epsilon) \rangle$ as a function of ϵ after rewiring 5% of the edges in 1995, while (c) shows the same for a 25% rewire fraction. The monotonic increase in the RMSD indicates that domestic centrality, which depends most strongly on the details of cross-border citations, is far more susceptible to the perturbation than is global centrality.

centrality of class i in simulation r and $n_r = 50$ the number of rewire simulations performed. Figure 9 confirms a number of expected behaviours. First, the average RMSD increases monotonically with ϵ, meaning that global centrality is more robust to random rewirings than domestic centrality. We also see that increasing the fraction of rewires increases the overall RMSD for all countries (compare the axes of Figs 9(b) and (c)). Finally, the centralities of classes in smaller economies like NL, GB and FR tend to be more susceptible to fluctuations in the citation network than the larger economies of USA, JP and DE, consistent with the behaviour seen in Fig. 6.

7. Conclusions

Centrality has long been recognized as a useful concept in determining the importance of individual nodes in complex networks, and patent citation networks are known to be amenable to such analysis as well [6,47]. In the global network of interdependent (and competing) national innovation systems, the importance of a patent class in the context of a domestic economy may not coincide with its importance in a global sense. In this paper, we define a new measure of centrality that takes into account a national preference for keeping innovation spillovers within the domestic economy rather than allowing it to flow into a foreign economy. We construct a model of innovation spillovers as a random walk and introduce the preference for domestic spillovers as an asymmetric loss term that depends on the political borders. We use this model to define a modified PageRank measure of centrality that provides insights into the patent classes that are important domestically, in terms of a single parameter ϵ (the relative increase in the probability of loss as information crosses a border).

The centrality of each sector of the global economy in our method is measured from the perspective of the domestic economy of interest, and we have applied our method to understand the central structures of 6 major economies in the global innovation network. Our model produces a wealth of information about the structure of national economies in a global network, identifying patent classes of importance both globally and from a more self-interested domestic perspective. We find that it is commonly the case that the most globally central classes are not the most domestically central due to the competition between the importance of the overall citation counts with the expectation that foreign citations are in some sense less desirable (reflected in the choice of ϵ). We highlight a number of specific examples of highly cited classes that have a low domestic centrality, illustrating how the competition between the raw number of citations and the fraction of domestic citations can lead to large variations in centrality.

We also show that our approach is robust to perturbations in the parameter ϵ_0 (the probability that a patent fails to provide a spillover), temporal aggregation and random noise through simulated rewiring. These results indicate that this measure of centrality can be reliably applied to potentially noisy data that would be expected in many systems. We therefore expect this approach can also be fruitfully applied to a variety of other systems of interacting networks [16–20], where sub-networks are both interdependent and competing to determine the most central nodes globally or in each sub-network.

Acknowledgements

The authors would like to thank Andrea Morescalchi for numerous helpful discussions.

Funding

We acknowledge funding from the CRISIS Lab project and from the Multiplex FP7 project (Foundational Research on MULTIlevel comPLEX networks and systems).

References

1. Goh, K., Cusick, M. E., Valle, D., Childs, B., Vidal, M. & Barabasi, A. (2007) The human disease network. *Proc. Natl Acad. Sci.*, **104**, 8685.
2. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. (2002) Network motifs: simple building blocks of complex networks. *Science*, **298**, 5594.
3. Yan, K., Fang, G., Bhardwaj, N., Alexander, R. P. & Gerstein, M. (2010) Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. *Proc. Natl Acad. Sci.*, **107**, 9186.
4. Meunier, D., Lambiotte, R. & Bullmore, E. T. (2010) Modular and hierarchically modular organization of brain networks. *Front. Neurosci.*, **4**, 1.
5. Hidalgo, C. A., Klinger, B., Barabasi, A.-L. & Hausmann, R. (2007) The product space conditions the development of nations. *Science*, **317**, 482.
6. Blochl, F., Theis, F., Vega-Redondo, F. & Fisher, E. (2011) Vertex centralities in input-output networks reveal the structure of modern economies. *Phys. Rev. E*, **83**, 046127.
7. Traud, A. L., Kelsic, E. D., Mucha, P. J. & Porter, M. A. (2011) Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev.*, **53**, 526.

8. Eagle, N., Macy, M. & Claxton, R. (2010) Network diversity and economic development. *Science*, **328**, 1029.

9. Chen, P., Xie, H., Maslov, S. & Redner, S. (2007) Finding scientific gems with Google's PageRank algorithm. *J. Infometrics*, **1**, 8.

10. Morrison, G. & Mahadevan, L. (2011) Asymmetric network connectivity using weighted harmonic averages. *Europhys. Lett.*, **93**, 40002.

11. Cerina, F., Chessa, A., Pammolli, F. & Riccaboni, M. (2014) Network communities within and across borders. *Sci. Rep.*, **4**, 4546.

12. Haveliwala, T. H. (2003) Topic sensitive PageRank: a context sensitive ranking algorithm for web search. *IEEE Trans. Knowl. Data Eng.*, **15**, 784.

13. Bonachich, P. (1987) Power and centrality: a family of measures. *Amer. J. Soc.*, **92**, 1170.

14. Newman, M. E. J. (2005) A measure of betweenness centrality based on random walks. *Soc. Net.*, **27**, 39.

15. Franceschet, M. (2011) PageRank: standing on the shoulders of giants. *Com. ACM*, **54**, 92.

16. Aguirre, J., Papo, D. & Buldu, J. (2013) Successful strategies for competing networks. *Nat. Phys.*, **9**, 230.

17. Dickinson, M., Havlin, S. & Stanley, H. E. (2012) Epidemics on interconnected networks. *Phys. Rev. E*, **85**, 066109.

18. Gao, J., Buldyrev, S. V., Stanley, H. E. & Havlin, S. (2012) Networks formed from interdependent networks. *Nat. Phys.*, **8**, 40.

19. Cardillo, A., Gmez-Gardees, J., Zanin, M., Romance, M., Papo, D., del Pozo, F. & Boccaletti, S. (2013) Emergence of network features from multiplexity. *Sci. Rep.*, **3**, 1.

20. Brummitt, C. D., R. M. D'Souza & Leicht, E. A. (2012) Suppressing cascades of load in interdependent networks. *Proc. Natl Acad. Sci.*, **109**, E681.

21. Halu, A., Mondragon, R. J., Panzarasa, P. & Bianconi, G. (2013) Multiplex pageRank. *PLoS One*, **8**, e78293.

22. Nicosia, V., Bianconi, G., Latora, V. & Barthelemy, M. (2013) Growing multiplex networks. *Phys. Rev. Lett.*, **111**, 058701.

23. Gomez, S., Daz-Guilera, A., Gmez-Gardees, J., Prez-Vicente, C. J., Moreno, Y. & Arenas, A. (2013) Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.*, **110**, 028701.

24. Gomez-Gardenes, J., Reinares, I., Arenas, A. & Floria, L. M. (2012) Evolution of cooperation in multiplex networks. *Sci. Rep.*, **2**, 620.

25. Buldyrev, S. V., Parshani, R., Paul, G., Stanely, H. E. & Havlin, S. (2010) Catastrophic cascade of failures in interdependent networks. *Nature*, **464**, 1025.

26. Gao, J., Buldyrev, S. V., Havlin, S. & Staneley, H. E. (2011) Robustness of a network of networks. *Phys. Rev. Lett.*, **107**, 195701.

27. Morrison, G., Giovannis, E., Riccaboni, M. & Pammolli, F. & Riccaboni, M. (2013) Border sensitive centralities in patent citation networks using asymmetric random walks. *International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, p. 546.

28. Coe, D. T. & Helpman, E. (1995) International R&D spillovers. *Eur. Econ. Rev.*, **39**, 859.

29. Erdi, P., Makovi, K., Somogyvari, Z., Strandburg, K., Tobochnik, J., Volf, P. & Zalanyi, L. (2013) Prediction of emerging technologies based on analysis of the US patent citation network. *Scienom*, **95**, 225.

30. Rossi, R. & Gleich, D. (2012) Dynamic PageRank using evolving teleportation. *Algorithms and Models for the Web Graph*. Berlin: Springer, p. 126.

31. Pearce, D., Groom, B., Hepburn, C. & Koundouri, P. (2003) Valuing the future: recent advances in social discounting. *World Econ.*, **4**, 121.

32. Arrow, K., Cropper, M., Gollier, C., Groom, B., Heal, G., Newell, R., Nordhaus, W., Pindyck, R., Pizer, W., Portney, P., Sterner, T., Tol, R. S. J. & Weitzman, M. (2013) Determining benefits and costs for future generations. *Science*, **341**, 349.

33. Constantine, P. G. & Gleich, D. F. (2009) Random alpha PageRank. *Int. Math.*, **6**, 189.

34. A full list of EPO categories can be found at http://web2.wipo.int/ipcpub.

35. Alcacer, J. & Chung, W. (2007) Location strategies and knowledge spillovers. *Manage. Sci.*, **53**, 760.

36. Papaconstantinou, G., Sakurai, N. & Wyckoff, A. (1998) Domestic and international product-embodied R&D diffusion. *Res. Policy*, **27**, 301.

37. Schmoch, U., Loaville, F., Patel, P. & Frietsch, R. (2003) *Linking Technology Areas to Industrial Sectors*. European Commission.

38. Chessa, A., Morescalchi, A., Pammolli, F., Penner, O., Petersen, A. M. & Riccaboni, M. (2013) Is Europe evolving towards an integrated research area? *Science*, **339**, 650.

39. Jaffe, A. B., Trajtenberg, M. & Romer, P. M. (2002) *Patents, Citations & Innovations: A Window on the Knowledge Economy*. Cambridge, MA: MIT Press.

40. Jaffe, A. B. & Trajtenberg, M. (1996) Flows of knowledge from universities and federal laboratories: modeling the flow of patent citations over time and across institutional and geographic boundaries. *Proc. Natl Acad. Sci.*, **93**, 12671.

41. Zhirov, A. O., Zhirov, O. V. & Shepelyansky, D. L. (2010) Two dimensional ranking of wikipedia articles. *Eur. Phys. J. B*, **77**, 523.

42. Gleich, D. F., Constantine, P. G., Flaxman, A.D. & Gunawardana, A. (2010) Tracking the random surfer: empirically measured teleportation parameters in PageRank. *Proceedings of the 19th international conference on World wide web*, p. 381.

43. Fogaras, D. & Racz, B. (2004) Towards scaling fully personalized pagerank. *Algorithms and Models for the Web-Graph*. Heidelberg, Berlin: Springer, p.105.

44. Pocock, M. J. O., Evans, D. M. & Memmott, J. (2012) The robustness and restoration of a network of ecological networks. *Science*, **335**, 97.

45. Berkhin, P. (2005) A survey on PageRank methods. *Internet Math.*, **2**, 73.

46. Kendall, M. (1938) A new measure of rank correlation. *Biometrika*, **30**, 81.

47. Gress, B. (2010) Properties of the USPTO patent citation network: 1963–2002. *World Patent Info.*, **32**, 3.