Maps of random walks on complex networks reveal community structure

Martin Rosvall† and Carl T. Bergstrom‡

†Department of Biology, University of Washington, Seattle, WA 98195-1800; and ‡Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

Edited by Brian Skyrms, University of California, Irvine, CA, and approved December 10, 2007 (received for review July 21, 2007)

To comprehend the multipartite organization of large-scale biological and social systems, we introduce an information theoretic approach that reveals community structure in weighted and directed networks. We use the probability flow of random walks on a network as a proxy for information flows in the real system and decompose the network into modules by compressing a description of the probability flow. The result is a map that both simplifies and highlights the regularities in the structure and their relationships. We illustrate the method by making a map of scientific communication as captured in the citation patterns of >6,000 journals. We discover a multicentric organization with fields that vary dramatically in size and degree of integration into the network. Along the backbone of the network—including physics, chemistry, molecular biology, and medicine—information flows bidirectionally, but the map reveals a directional pattern of citation from the applied fields to the basic sciences.

Biological and social systems are differentiated, multipartite, integrated, and dynamic. Data about these systems, now available on unprecedented scales, often are schematized as networks. Such abstractions are powerful (1, 2), but even as abstractions they remain highly complex. It therefore is helpful to decompose the myriad nodes and links into modules that represent the network (3–5). A cogent representation will retain the important information about the network and reflect the fact that interactions between the elements in complex systems are weighted, directional, interdependent, and conductive. Good representations both simplify and highlight the underlying structures and the relationships that they depict; they are maps (6, 7).

To create a good map, the cartographer must attain a fine balance between omitting important structures by oversimplification and obscuring significant relationships in a barrage of superfluous detail. The best maps convey a great deal of information but require minimal bandwidth: the best maps are superfluous detail. The best maps convey a great deal of information but require minimal bandwidth: the best maps are.

Network Maps and Coding Theory

In this article, we use maps to describe the dynamics across the links and nodes in directed, weighted networks that represent the local interactions among the subunits of a system. These local interactions induce a system-wide flow of information that characterizes the behavior of the full system (8–12). Consequently, if we want to understand how network structure relates to system behavior, we need to understand the flow of information on the network. We therefore identify the modules that compose the network by finding an efficiently coarse-grained description of how information flows on the network. A group of nodes among which information flows quickly and easily can be aggregated and described as a single well connected module; the links between modules capture the avenues of information flow between those modules.

Succinctly describing information flow is a coding or compression problem. The key idea in coding theory is that a data stream can be compressed by a code that exploits regularities in the process that generates the stream (13). We use a random walk as a proxy for the information flow, because a random walk uses all of the information in the network representation and nothing more. Thus, it provides a default mechanism for generating a dynamics from a network diagram alone (8).

Taking this approach, we develop an efficient code to describe a random walk on a network. We thereby show that finding community structure in networks is equivalent to solving a coding problem (14–16). We exemplify this method by making a map of science, based on how information flows among scientific journals by means of citations.

Describing a Path on a Network. To illustrate what coding has to do with map-making, consider the following communication game. Suppose that you and I both know the structure of a weighted, directed network. We aim to choose a code that will allow us to efficiently describe paths on the network that arise from a random walk process in a language that reflects the underlying structure of the network. How should we design our code?

If maximal compression were our only objective, we could encode the path at or near the entropy rate of the corresponding Markov process. Shannon showed that one can achieve this rate by assigning to each node a unique dictionary over the outgoing transitions (17). But compression is not our only objective; here, we want our language to reflect the network structure, we want the words we use to refer to things in the world. Shannon’s approach does not do this for us because every codeword would have a different meaning depending on where it is used. Compare maps: useful maps assign unique names to important structures. Thus, we seek a way of describing or encoding the random walk in which important structures indeed retain unique names. Let us look at a concrete example. Fig. 1A shows a weighted network with *n* = 25 nodes. The link thickness indicates the relative probability that a random walk will traverse any particular link. Overlaid on the network is a specific 71-step realization of a random walk that we will use to illustrate our communication game. In Fig. 1, we describe this walk with increasing levels of compression (B–D), exploiting more and more of the regularities in the network.

Huffman Coding. A straightforward method of giving names to nodes is to use a Huffman code (18). Huffman codes save space...
Fig. 1. Detecting communities by compressing the description of information flows on networks. (A) We want to describe the trajectory of a random walk on the network such that important structures have unique names. The orange line shows one sample trajectory. (B) A basic approach is to give a unique name to every node in the network. The Huffman code illustrated here is an efficient way to do so. The 314 bits shown under the network describe the sample trajectory in A, starting with 1111100 for the first node on the walk in the upper left corner, 1100 for the second node, etc., and ending with 00011 for the last node on the walk in the lower right corner. (C) A two-level description of the random walk, in which major clusters receive unique names, but the names of nodes within clusters are reused, yields on average a 32% shorter description for this network. The codes naming the modules and the codes used to indicate an exit from each module are shown to the left and the right of the arrows under the network, respectively. Using this code, we can describe the walk in A by the 243 bits shown under the network in C. The first three bits 111 indicate that the walk begins in the red module, the code 0000 specifies the first node on the walk, etc. (D) Reporting only the module names, and not the locations within the modules, provides an efficient coarse graining of the network.
by assigning short codewords to common events or objects and long codewords to rare ones, much as common words are short in spoken languages (19). Fig. 1B shows a prefix-free Huffman coding for our sample network. Each codeword specifies a particular node, and the codeword lengths are derived from the ergodic node visit frequencies of an infinitely long random walk. With the Huffman code pictured in Fig. 1B, we are able to describe the specific 71-step walk in 314 bits. If we instead had chosen a uniform code, in which all codewords are of equal length, each codeword would be \( \lceil \log 25 \rceil = 5 \) bits long and 71-5 = 355 bits would have been required to describe the walk.

Although in this example we assign actual codewords to the nodes for illustrative purposes, in general, we will not be interested in the codewords themselves but rather in the theoretical limit of how concisely we can specify the path. Here, we invoke Shannon’s source coding theorem (17), which implies that when you use \( n \) codewords to describe the \( n \) states of a random variable \( X \) that occur with frequencies \( p_i \), the average length of a codeword can be no less than the entropy of the random variable \( X \) itself: \( H(X) = -\sum p_i \log(p_i) \). This theorem provides us with the necessary apparatus to see that, in our Huffman illustration, the average number of bits needed to describe a single step in the random walk is bounded below by the entropy \( H(P) \), where \( P \) is the distribution of visit frequencies to the nodes on the network. We define this lower bound on code length to be \( L \). For example, \( L = 4.50 \) bits per step in Fig. 1B.

### Highlighting Important Objects.

Matching the length of codewords to the frequencies of their use gives us efficient codewords for the nodes, but no map. Merely assigning appropriate-length names to the nodes does little to simplify or highlight aspects of the important structures from the insignificant details. We therefore divide the network into two levels of description. We retain unique names for large-scale objects, the clusters or modules to be identified within our network, but we reuse the names associated with fine-grain details, the individual nodes within each module. This is a familiar approach for assigning names to objects on maps: most U.S. cities have unique names, but street names are reused from one city to the next, such that each city has a Main Street and a Broadway and a Washington Avenue and so forth. The reuse of street names rarely causes confusion, because most routes remain within the bounds of a single city.

A two-level description allows us to describe the path in fewer bits than we could do with a one-level description. We capitalize on the network’s structure and, in particular, on the fact that a random walker is statistically likely to spend long periods of time within certain clusters of nodes. Fig. 1C illustrates this approach. We give each cluster a unique name but use a different Huffman code to name the nodes within each cluster. A special codeword, the exit code, is chosen as part of the within-cluster Huffman coding and indicates that the walk is leaving the current cluster. The exit code always is followed by the “name” or module code of the new module into which the walk is moving [see supporting information (SI) for more details]. Thus, we assign unique names to coarse-grain structures (the cities in the city metaphor) but reuse the names associated with fine-grain details (the streets in the city metaphor). The savings are considerable; in the two-level description of Fig. 1C the limit \( L \) is 3.05 bits per step compared with 4.50 for the one-level description.

Herein lies the duality between finding community structure in networks and the coding problem: to find an efficient code, we look for a module partition \( M \) of \( n \) nodes into \( m \) modules so as to minimize the expected description length of a random walk. By using the module partition \( M \), the average description length of a single step is given by

\[
L(M) = q_{\gamma} H(\mathcal{E}) + \sum_{i=1}^{m} p_i H(\mathcal{O}_i). \tag{1}
\]

This equation comprises two terms: first is the entropy of the movement between modules, and second is the entropy of movements within modules (where exiting the module also is considered a movement). Each is weighted by the frequency with which it occurs in the particular partitioning. Here, \( \gamma \) is the probability that the random walk switches modules on any given step. \( H(\mathcal{E}) \) is the entropy of the module names, i.e., the entropy of the underlined codewords in Fig. 1D. \( H(\mathcal{O}_i) \) is the entropy of the within-module movements, including the exit code for module \( i \). The weight \( p_i \) is the fraction of within-module movements that occur in module \( i \), plus the probability of exiting module \( i \) such that \( \sum_{i=1}^{m} p_i = 1 + \gamma \) (see SI for more details).

For all but the smallest networks, it is infeasible to check all possible partitions to find the one that minimizes the description.
length in the map equation (Eq. 1). Instead, we use computational search. We first compute the fraction of time each node is visited by a random walker using the power method, and, using these visit frequencies, we explore the space of possible partitions by using a deterministic greedy search algorithm (20, 21). We refine the results with a simulated annealing approach (6) using the heat-bath algorithm (see SI for more details).

Fig. 1D shows the map of the network, with the within-module descriptors faded out; here the significant objects have been highlighted and the details have been filtered away. In the interest of visual simplicity, the illustrative network in Fig. 1 has weighted but undirected links. Our method is developed more generally, so that we can extract information from networks with links that are directed in addition to being weighted. The map equation remains the same; only the path that we aim to describe must be slightly modified to achieve ergodicity. We introduce a small “teleportation probability” τ in the random walk: with probability τ, the process jumps to a random node anywhere in the network, which converts our random walker into the sort of “random surfer” that drives Google’s PageRank algorithm (22). Our clustering results are highly robust to the particular choice of the small fraction τ. For example, so long as τ < 0.45 the optimal partitioning of the network in Fig. 1 remains exactly the same. In general, the more significant the regularities, the higher τ can be before frequent teleportation swamps the network structure. We choose τ = 0.15 corresponding to the well known damping factor d = 0.85 in the PageRank algorithm (22).

Mapping Flow Compared with Maximizing Modularity

The traditional way of identifying community structure in directed and weighted networks has been simply to disregard the directions and the weights of the links. But such approaches discard valuable information about the network structure. By mapping the system-wide flow induced by local interactions between nodes, we retain the information about the directions and the weights of the links. We also acknowledge their interdependence in networks inherently characterized by flows. This distinction makes it interesting to compare our flow-based approach with recent topological approaches based on modularity optimization that also makes use of information about weight and direction (23–26). In its most general form, the
modularity for a given partitioning of the network into $m$ modules is the sum of the total weight of all links in each module minus the expected weight

$$Q = \sum_{i=1}^{m} \frac{w_{ii}}{w} - \frac{w_{in} w_{out}}{w^2}.$$

Here, $w_{ii}$ is the total weight of links starting and ending in module $i$, $w_{in}$ and $w_{out}$ are the total in- and out-weight of links in module $i$, and $w$ is the total weight of all links in the network. To estimate the community structure in a network, Eq. 2 is maximized over all possible assignments of nodes into any number $m$ of modules. Eqs. 1 and 2 reflect two different senses of what it means to have a network. The former, which we pursue here, finds the essence of a network in the patterns of flow that its structure induces. The latter effectively situates the essence of network in the topological properties of its links (as we did in ref. 16).

Does this conceptual distinction make any practical difference? Fig. 2 illustrates two simple networks for which the map equation and modularity give different partitionings. The weighted, directed links shown in the network in Fig. 2A induce a structured pattern of flow with long persistence times in, and limited flow between, the four clusters as highlighted on the left. The map equation picks up on these structural regularities, and thus the description length is much shorter for the partitioning in Fig. 2A Left (2.67 bits per step) than for Fig. 2A Right (4.13 bits per step). Modularity is blind to the interdependence in networks characterized by flows and thus cannot pick up on this type of structural regularity. It only counts weights of links, in-degree, and out-degree in the modules, and thus prefers to partition the network as shown in Fig. 2A Right with the heavily weighted links inside of the modules.

In Fig. 2B, by contrast, there is no pattern of extended flow at all. Every node is either a source or a sink, and no movement along the links on the network can exceed more than one step in length. As a result, random teleportation will dominate (irrespective of teleportation rate), and any partition into multiple modules will lead to a high flow between the modules. For a network such as in Fig. 2B, where the links do not induce a pattern of flow, the map equation always will partition the network into one single module. Modularity, because it looks at pattern in the links and in- and out-degree, separates the network into the clusters shown at right.

Which method should a researcher use? It depends on which of the two senses of network, described above, that the researcher is studying. For analyzing network data where links represent patterns of movement among nodes, flow-based approaches such as the map equation are likely to identify the most important aspects of structure. For analyzing network data where links represent not flows but rather pairwise relationships, it may be useful to detect structure even where no flow exists. For these systems, topological methods such as modularity (11) or cluster-based compression (16) may be preferable.

**Mapping Scientific Communication**

Science is a highly organized and parallel human endeavor to find patterns in nature; the process of communicating research findings is as essential to progress as is the act of conducting the research in the first place. Thus, science is not merely a set of ideas but also the flow of these ideas through a multiparticle and highly differentiated social system. Citation patterns among journals allow us to glimpse this flow and provide the trace of communication between scientists (27–31). To highlight important fields and their relationships, to uncover differences and changes, to simplify and make the system comprehensible—we need a good map of science.

Using the information theoretic approach presented above, we map the flow of citations among 6,128 journals in the sciences (Fig. 3) and social sciences (Fig. 4). The 6,434,916 citations in this...
cross-citation network represent a trace of the scientific activity during 2004 (32). Our data tally on a journal-by-journal basis the citations from articles published in 2004 to articles published in the previous 5 years. We exclude journals that publish <12 articles per year and those that do not cite other journals within the data set. We also exclude the only three major journals that span a broad range of scientific disciplines: Science, Nature, and Proceedings of the National Academy of Sciences; the broad scope of these journals otherwise creates an illusion of tighter connections among disciplines, when in fact few readers of the physics articles in Science also are close readers of the biomedical articles therein. Because we are interested in relationships between journals, we also exclude journal self-citations.

Through the operation of our algorithm, the fields and the boundaries between them emerge directly from the citation data rather than from our preconceived notions of scientific taxonomy (see Figs. 3 and 4). Our only subjective contribution has been to suggest reasonable names for each cluster of journals that the algorithm identifies: economics, mathematics, geosciences, and so forth.

The physical size of each module or “field” on the map reflects the fraction of time that a random surfer spends following citations within that module. Field sizes vary dramatically. Molecular biology includes 723 journals that span the areas of genetics, cell biology, biochemistry, immunology, and development biology; a random surfer spends 26% of her time in this field, indicated by the size of the module. Tribology (the study of friction) includes only seven journals, in which a random surfer spends 0.064% of her time.

The weighted and directed links between fields represent citation flow, with the color and width of the arrows indicating flow volume. The heavy arrows between medicine and molecular biology indicate a massive traffic of citations between these disciplines. The arrows point in the direction of citation: $A \rightarrow B$ means “$A$ cites $B$” as shown in the key. These directed links reveal the relationship between applied and basic sciences. We find that the former cite the latter extensively, but the reverse is not true, as we see, for example, with geotechnology citing material sciences but not being cited by them. The thickness of the module borders reflect the probability that a random surfer within the module will follow a citation to a journal outside of the module. These outflows show a large variation; for example the outflow is 30% in general medicine but only 12% in economics.

The map reveals a ring-like structure in which all major disciplines are connected to one another by chains of citations, but these connections are not always direct because fields on opposite sides of the ring are linked only through intermediate fields. For example, although psychology rarely cites general physics or vice versa, psychology and general physics are connected via the strong links to and between the intermediaries molecular biology and chemistry. Once we consider the weights of the links among fields, however, it becomes clear that the structure of science is more like the letter U than like a ring, with the social sciences at one terminal and engineering at the other, joined mainly by a backbone of medicine, molecular biology, chemistry, and physics. Because our map shows the pattern of citations to research articles published within 5 years, it represents what de Solla Price called the “research frontier” (27) rather than the long-term interdependencies among fields. For example, although mathematics are essential to all natural sciences, the field of mathematics is not central in our map because only certain subfields (e.g., areas of physics and statistics) rely heavily on the most recent developments in pure mathematics and contribute in return to the research agenda in that field.

When a cartographer designs a map, the scale or scope of the map influences the choice of which objects are represented. A regional map omits many of the details that appear on a city map. Similarly, in the approach that we have developed here, the appropriate size or resolution of the modules depends on the universe of nodes that are included in the network. If we compare the map of a network to a map of a subset of the same network, we would expect to see the map of the subset reveal finer divisions, with modules composed of fewer nodes. Fig. 4 illustrates this by partitioning a subset of the journals included in the map of science: the 1,431 journals in the social sciences. The basic structure of the fields and their relations remains unchanged, with psychiatry and psychology linked via sociology and management to economics and political science, but the map also reveals further details. Anthropology fractures along the physical/cultural divide. Sociology divides into behavioral and institutional clusters. Marketing secedes from management. Psychology and psychiatry reveal a set of applied subdisciplines.

The additional level of detail in the more narrowly focused map would have been clutter on the full map of science. When we design maps to help us comprehend the world, we must find that balance where we eliminate extraneous detail but highlight the relationships among important structures. Here, we have shown how to formalize this cartographer’s precept by using the mathematical apparatus of information theory.

ACKNOWLEDGMENTS. We are grateful to Jevin West for processing the data used to construct the maps in Figs. 3 and 4 and to Cynthia A. Brewer, www.ColorBrewer.org, for providing the color schemes we have used in Figs. 1–4. This work was supported by the National Institute of General Medical Sciences Models of Infectious Disease Agent Study Program Cooperative Agreement SU01GM07649.

Here we present a detailed description of the map equation that serves as our objective function, followed by a step-by-step overview of the computational procedure that we use to minimize it. Thereafter we illustrate the procedure for the greedy search in 22 subsequent slides, which show how the method operates to find successively shorter encodings that highlight and exploit regularities in the network structure. Finally we present the maps from Figs. 3 and 4 in the paper, along with a listing of the complete set of fields and a listing of all journals within the fields.

The map equation

Define a module partition \( M \) as a hard partition of a set of \( n \) nodes into \( m \) modules such that each node is assigned to one and only one module. The map equation \( L(M) \) gives the average number of bits per step that it takes to describe an infinite random walk on a network partitioned according to \( M \):

\[
L(M) = q_{i\rightarrow} H(Q) + \sum_{i=1}^{m} p_{i\rightarrow} H(P^i). 
\]

Below, we define and expand these terms, but first a note about the general approach. The map equation calculates the minimum description length of a random walk on the network for a two-level code that separates the important structures from the insignificant details based on the partition \( M \). As described in the main text, this two-level code uses unique codewords to name the modules specified by partition \( M \) but reuses the codewords used to name the individual nodes within each module. The first term of this equation (in red) gives the average number of bits necessary to describe movement between modules, and the second term (in blue) gives the average number of bits necessary to describe movement within modules.

To efficiently describe a random walk using a two-level code of this sort, the choice of partition \( M \) must reflect the patterns of flow within the network, with each module corresponding to a cluster of nodes in which a random walker spends a long period of time before departing for another module. To find the best such partition, we therefore seek to minimize the map equation over all possible partitions \( M \). We begin by expanding the terms in the map equation. For clarity we here use \( i, j \) to enumerate modules, \( \alpha, \beta \) to enumerate nodes, red terms to describe movements between the modules, and blue terms to describe movements within the modules.

The per step probability that the random walker switches modules is

\[
q_{i\rightarrow} = \sum_{i=1}^{m} q_{i\rightarrow}, \tag{2}
\]

where \( q_{i\rightarrow} \) is the per step probability that the random walker exits module \( i \). This probability depends on the partitioning of the network and will be derived in Eq. 7.

The entropy of movements between modules is

\[
H(Q) = \sum_{i=1}^{m} q_{i\rightarrow} \log \left( \frac{q_{i\rightarrow}}{\sum_{j=1}^{m} q_{j\rightarrow}} \right), \tag{3}
\]

which is the lower limit of the average length of a codeword used to name a module. Here we have used Shannon’s source coding theorem (1) and treated the modules as \( m \) states of a random variable \( X \) that occur with frequencies \( q_{i\rightarrow} / \sum_{j=1}^{m} q_{j\rightarrow} \). Combining Eqs. 2 and 3, the first term in the map equation is the per step average description length of movements between modules within the random walk.

To weight the entropy of movements within module \( i \), we compute

\[
p_{i\rightarrow} = q_{i\rightarrow} + \sum_{\alpha \in i} p_{\alpha}, \tag{4}
\]

where the notation \( \alpha \in i \) means “over all nodes \( \alpha \) in module \( i \)” and \( p_{\alpha} \) is the ergodic node visit frequency at node \( \alpha \) within the random walk. We use the power method, to be explained in detail on next page, to calculate this probability. Because the exit codewords are necessary to separate within-module movements from between-module movements, we include the probability of exiting module \( i \), \( q_{i\rightarrow} \), in the weight of within-module movements in module \( i \). In this way we can guarantee efficient coding: by encoding the exit codewords together with the within-module codewords, we appropriately adjust the length of the exit codewords to the frequency of their use.

Finally, the entropy of movements within module \( i \) is

\[
H(P^i) = \frac{q_{i\rightarrow}}{q_{i\rightarrow} + \sum_{\beta \in i} p_{\beta}} \log \left( \frac{q_{i\rightarrow}}{q_{i\rightarrow} + \sum_{\beta \in i} p_{\beta}} \right) + \sum_{\alpha \in i} \frac{p_{\alpha}}{q_{i\rightarrow} + \sum_{\beta \in i} p_{\beta}} \log \left( \frac{p_{\alpha}}{q_{i\rightarrow} + \sum_{\beta \in i} p_{\beta}} \right), \tag{5a}
\]

which is the lower limit of the average length of a codeword used to name a node (exit code included) in module \( i \). The single term in Eq. 5a is the contribution from the exit codeword and the sum in Eq. 5b is the contribution from the codewords naming the nodes. Combining Eqs. 4 and 5 and summing over all modules makes it easy to identify the second term in the map equation as the per step average description length of movements within modules of the random walk.

By collecting the terms and simplifying, we get the final ex-
pression for the map equation

\[ L(M) = \left( \sum_{i=1}^{m} q_{i \sim \infty} \right) \log \left( \sum_{i=1}^{m} q_{i \sim \infty} \right) - (1 + 1) \sum_{i=1}^{m} q_{i \sim \infty} \log (q_{i \sim \infty}) - \sum_{\alpha=1}^{n} p_{\alpha} \log (p_{\alpha}) \]  

(6a)

\[ + \sum_{i=1}^{m} \left( q_{i \sim \infty} + \sum_{\alpha \in i} p_{\alpha} \right) \log \left( \frac{q_{i \sim \infty} + \sum_{\alpha \in i} p_{\alpha}}{n_{i}} \right). \]  

(6c)

Note that the map equation is only a function of the ergodic node visit frequencies \( p_{\alpha} \) and the exit probabilities \( q_{i \sim \infty} \), which both can be easily calculated. Moreover, because the term \( \sum_{\alpha} p_{\alpha} \log (p_{\alpha}) \) is independent of partitioning and \( p_{\alpha} \) otherwise only shows up summed over all nodes in a module, it is sufficient to keep track of changes in \( q_{i \sim \infty} \) and \( \sum_{\alpha \in i} p_{\alpha} \) in the optimization algorithm. They can easily be derived for any partition of the network or quickly updated when they change in each step of the optimization procedure using the ergodic node visit frequencies.

- **Ergodic node visit frequencies.** We use the power method to calculate the steady state visit frequency for each node. To guarantee a unique steady state distribution for directed networks, we introduce a small teleportation probability \( \tau \) in the random walk that links every node to every other node with positive probability and thereby convert the random walker into a random surfer. The movement of the random surfer can now be described by an irreducible and aperiodic Markov chain that has a unique steady state by the Perron-Frobenious theorem. To generate the ergodic node visit frequencies, we start with a distribution of \( p_{\alpha} = 1/n \) for the random surfer to start at each node \( \alpha \). The surfer moves as follows: at each time step, with probability \( 1 - \tau \) the random surfer follows one of the outgoing links from the node \( \alpha \) that it currently occupies, with probability proportional to the weights of the outgoing links \( w_{\alpha \beta} \) from \( \alpha \) to \( \beta \). It is therefore convenient to set \( \sum_{\beta} w_{\alpha \beta} = 1 \). With the remaining probability \( \tau \), or with probability 1 if the node does not have any outlinks, the random surfer “teleports” with uniform probability to a random node anywhere in the system. As in Google’s PageRank algorithm (2), we use \( \tau = 0.15 \), but emphasize that the results are robust to this choice.

- **Exit probabilities.** Given the ergodic node visit frequencies \( p_{\alpha}, \alpha = 1, \ldots, n \) and an initial partitioning of the network, it is easy to calculate the ergodic module visit frequencies \( \sum_{\alpha \in i} p_{\alpha} \) for module \( i \). The exit probability for module \( i \), with teleportation taken into account is then

\[ q_{i \sim \infty} = \frac{\tau n - n_{i}}{n} \sum_{\alpha \in i} p_{\alpha} + (1 - \tau) \sum_{\alpha \in i} \sum_{\beta \notin i} p_{\alpha} w_{\alpha \beta}, \]  

(7)

where \( n_{i} \) is the number of nodes in module \( i \). This equation follows since every node teleports a fraction \( \tau (n - n_{i})/(n - 1) \) and guides a fraction \((1 - \tau) \sum_{\beta \notin i} w_{\alpha \beta}\) of its weight \( p_{\alpha} \) to nodes outside of the module.

**Implementation**

Finding the map that provides the minimal description length of the data, given the requirement that modules receive unique names, is now a standard computational optimization problem. Below we describe how we first use a deterministic greedy search algorithm (3, 4) and then refine the results with a simulated annealing approach (5, 6) with the heat-bath algorithm. Our implementation, written in C++, is available upon request.

1. **Greedy search.** We first calculate the ergodic node visit frequencies and then assign every node to a unique module and derive the exit probabilities as described above. We use the map equation to calculate the description length and repeatedly merge the two modules that give the largest decrease in description length until further merging gives a longer description (3, 4). With the improved version in ref. (4) of the greedy search algorithm in ref. (3), we have successfully partitioned networks with 2.6 million nodes and 29 million links. Starting at the next page, we illustrate the greedy search for the example network in Fig. 1 of the paper.

2. **Simulated annealing.** The result of the previous step can typically be refined by simulated annealing (5, 6). We use the heat-bath algorithm (7) and start with the module configuration achieved by the greedy search. Starting the heat-bath algorithm at several different temperatures, we select the run that gives the shortest description of the map, i.e., the minimal value of the map equation. This step can improve the description length by up to several percent over that found by the greedy search alone.

3. **Visualization.** We set the area of every module to be proportional to the fraction of time a random surfer spends in the module, and the area of the bordering ring to be proportional to the exit probability. Similarly, we vary the widths of the links in accord with the transition probabilities between modules (excluding teleportation).

**References**

Fig. 5 The greedy search algorithm begins with each node $\alpha = 1, \ldots, n$ in its own module $i = 1, \ldots, m$, as shown here. The slides on the following 22 pages show how the greedy search operates, finding successively shorter encodings that highlight and exploit regularities in the network structure. Node colors indicate module identity. To illustrate the duality between module detection and coding, the code structure corresponding to the current partition is shown by the stacked boxes on the right. The height of each box corresponds to the per step rate of codeword use.

The left stack represents the codes associated with movements between modules. The height of each box is equal to the exit probability $q_\statealker/\rho_\statealker$, of the corresponding module $i$. Boxes are ordered according to their heights. The codewords naming the modules are the Huffman codes calculated from the probabilities $q_\statealker/\rho_\statealker$, where $q_\statealker = \sum_{i=1}^{m} q_\statealker$ is the total height of the left stack. The length of the codeword naming module $i$ is approximately $-\log q_\statealker/\rho_\statealker$, the Shannon limit in the map equation. In the figure, the per step description length of the random walker’s movements between modules is the sum of the length of the codewords weighted by their use. This length is bounded below by the limit $-\rho_\statealker \sum_{i=1}^{m} q_\statealker \log q_\statealker/\rho_\statealker$ that we use in the paper.

The right stack is associated with movements within modules. The height of each box in module $i$ is equal to the ergodic node visit probabilities $p_{\alpha \in i}$ or the exit probability $q_\statealker$. The boxes corresponding to the same module are collected together and ordered according to their weight; in turn the modules are ordered according to their total weights $p_\statealker = q_\statealker + \sum_{\alpha \in i} p_{\alpha}$. The codewords naming the nodes and exit in each module $i$ are the Huffman codes calculated from the probabilities $p_{\alpha \in i}/p'_i$ (nodes) and $q_\statealker/p'_i$ (exit). The length of codewords naming nodes $\alpha \in i$ and exit from module $i$ are approximately $-\log (p_{\alpha \in i}/p'_i)$ (nodes) and $-\log (q_\statealker/p'_i)$ (exit). In the figure, the per step description length of the random walker’s movements within modules is the sum of the length of the codewords weighted by their use. This length is bounded below by the limit $-\rho_\statealker \sum_{i=1}^{m} \left[ q_\statealker \log (q_\statealker/p'_i) + \sum_{\alpha \in i} p_{\alpha} \log (p_{\alpha}/p'_i) \right]$.

The total description length $L(M)$ for the module partition $M$ of the network is the sum of the contributions from movements between and within modules and it takes a minimum value for the partitioning into four models on the last slide.
$L(M) = 4.53 + 2.00 = 6.53 \text{ bits}$
$L(M) = 4.31 + 2.05 = 6.37\text{ bits}$
$L(M) = 3.98 + 2.15 = 6.13$ bits
$L(M) = 3.77 + 2.20 = 5.97 \text{ bits}$
$L(M) = 3.47 + 2.22 = 5.69 \text{ bits}$
\[ L(M) = 3.19 + 2.20 = 5.39 \text{ bits} \]
The information entropy $L(M)$ is calculated as:

$$L(M) = 2.99 + 2.26 = 5.24 \text{ bits}$$
\[ L(M) = 2.72 + 2.30 = 5.03 \text{ bits} \]
\[ L(M) = 2.42 + 2.36 = 4.78 \text{ bits} \]
$L(M) = 2.13 + 2.35 = 4.48$ bits
\[ L(M) = 1.97 + 2.38 = 4.35 \text{ bits} \]
\[ L(M) = 1.76 + 2.43 = 4.19 \text{ bits} \]
\( \mathcal{L}(M) = 1.59 + 2.46 = 4.05 \text{ bits} \)
\[ L(M) = 1.30 + 2.57 = 3.88 \text{ bits} \]
$L(M) = 1.16 + 2.58 = 3.74 \text{ bits}$
\[ L(M) = 0.97 + 2.66 = 3.63 \text{ bits} \]
\[ L(M) = 0.75 + 2.73 = 3.48 \text{ bits} \]
\[ L(M) = 0.58 + 2.79 = 3.37 \text{ bits} \]
\[ L(M) = 0.51 + 2.79 = 3.30 \text{ bits} \]
$L(M) = 0.36 + 2.88 = 3.24$ bits
\[ L(M) = 0.13 + 2.97 = 3.09 \text{ bits} \]
Fig. 6 A map of science with fields according to a partitioning with the presented method. All 88 field names in this caption and the 51 field names in the map are clickable links to lists with the journals in the corresponding fields. The fields and journals are ranked by the fraction of time they are visited by a random surfer. 1 Molecular & Cell Biology, 2 Medicine, 3 Physics, 4 Neuroscience, 5 Ecology & Evolution, 6 Economics, 7 Geosciences, 8 Psychology, 9 Chemistry, 10 Psychiatry, 11 Environmental Chemistry & Microbiology, 12 Mathematics, 13 Computer Science, 14 Analytic Chemistry, 15 Business & Marketing, 16 Political Science, 17 Fluid Mechanics, 18 Medical Imaging, 19 Material Engineering, 20 Sociology, 21 Probability & Statistics, 22 Astronomy & Astrophysics, 23 Gastroenterology, 24 Law, 25 Chemical Engineering, 26 Education, 27 Telecommunication, 28 Orthopedics, 29 Control Theory, 30 Environmental Health, 31 Operations Research, 32 Ophthalmology, 33 Crop Science, 34 Geography, 35 Anthropology, 36 Veterinary, 37 Computer Imaging, 38 Agriculture, 39 Parasitology, 40 Dentistry, 41 Dermatology, 42 Urology, 43 Rheumatology, 44 Applied Acoustics, 45 Pharmacology, 46 Pathology, 47 History & Philosophy of Science, 48 Otolaryngology, 49 Electromagnetic Engineering, 50 Information Science, 51 Circuits, 52 Media & Communication, 53 Power Systems, 54 Tribology, 55 History, 56 Geotechnology, 57 Wood Products, 58 Radiation, 59 Linguistics, 60 Social Work, 61 Psychoanalysis, 62 Middle Eastern Studies, 63 Forensic Science, 64 Transfusion, 65 Mycology, 66 Nuclear Energy, 67 offshore Engineering, 68 Environmental Ethics, 69 Insect Taxonomy, 70 Higher Education, 71 Refineries, 72 Reliability Engineering, 73 Other, 74 Civil Engineering, 75 Lab Veterinary, 76 Music, 77 Tourism, 78 Textiles, 79 Creativity Research, 80 Travel Sociology, 81 Medical Informatics, 82 Leprosy, 83 Sociology (Russian), 84 Cryobiology, 85 Death Studies, 86 Rehabilitation Counseling, 87 Steel, 88 Futurist.
Fig. 7 A map of social science with fields according to a partitioning with the presented method. All 54 field names in this caption and the 36 field names in the map are clickable links to lists with the journals in the corresponding fields. The fields and journals are ranked by the fraction of time they are visited by a random surfer. 1 Economics, 2 Psychology, 3 Psychiatry, 4 Healthcare, 5 Political Science, 6 Sociology (Behavioral), 7 Management, 8 Law, 9 Education, 10 Geography, 11 Physical Anthropology, 12 Cultural Anthropology, 13 Marketing, 14 Information Science, 15 Philosophy of Science, 16 Sociology (Institutional), 17 Communication, 18 Educational Assessment, 19 Educational Psychology, 20 Human-Computer Interface, 21 Applied Linguistics, 22 Experimental Psychology, 23 History, 24 Social Work, 25 Speech And Hearing, 26 Disabilities, 27 Transportation, 28 Psychoanalysis, 29 Guidance Counseling, 30 Middle Eastern Studies, 31 East Asian Studies, 32 Ergonomics, 33 Medical Ethics, 34 Public Administration, 35 Ethics, 36 Economic History, 37 Sport Psychology, 38 Public Affairs, 39 Social Policy, 40 Family Relations, 41 Law And Behavior, 42 Criminology, 43 Sexuality, 44 Higher Education, 45 Leisure Studies, 46 Neurorehabilitation, 47 Pacific Studies, 48 Tourism, 49 Creativity, 50 Death And Dying, 51 Sociology (French), 52 Sociology (Eastern Europe), 53 Maritime Law, 54 Hypnosis
Fields of science

1 Molecular & Cell Biology (20%)
3 Physics (10%)

5 Ecology & Evolution (4.3%)

6 Economics (3.5%)
9 Chemistry (2.7%)


10 Psychiatry (2.4%)

11 Environmental Chemistry & Microbiology (2.3%)

12 Mathematics (2.0%)


13 Computer Science (1.4%)


14 Analytic Chemistry (1.4%)

15 Business & Marketing (1.2%)


16 Political Science (1.2%)

17 Fluid Mechanics (1.1%)


18 Medical Imaging (1.1%)


19 Material Engineering (1.1%)

### 20 Sociology (0.97%)

<table>
<thead>
<tr>
<th>Journal Title</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int, 52 Computation Stat, 53 Statistics, 54 Environ Ecol Sta</td>
<td></td>
</tr>
<tr>
<td>Stat Comput Sim, 50 J Agr Biol Envir St, 51 Qual Reliab Eng Int, 52 Computation Stat</td>
<td></td>
</tr>
<tr>
<td>53 Statistics, 54 Environ Ecol Stat, 55 Stat Neerl</td>
<td></td>
</tr>
<tr>
<td>56 Stat Pap, 57 Qual Quant</td>
<td></td>
</tr>
</tbody>
</table>

### 22 Astronomy & Astrophysics (0.86%)

<table>
<thead>
<tr>
<th>Journal Title</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Astrophys J, 2 Astron Astrophys, 3 Mon Not R Astron Soc</td>
<td></td>
</tr>
<tr>
<td>4 Gut, 5 J Hepatol, 6 Am J Physiol-Gastr L, 7 Gastrointest Endose</td>
<td></td>
</tr>
<tr>
<td>32 Hepatol Res, 33 Best Pract Res Cl Ga, 34 Z Gastroentero, 35 Dis Esophagus, 36 Metab Brain Dis, 37 Digest Dis, 38 Acta Gastro-Ent Belg, 39 Curr Opin Gastroen, 40 Rev Esp Enferm Dig</td>
<td></td>
</tr>
</tbody>
</table>

### 23 Gastroenterology (0.80%)

<table>
<thead>
<tr>
<th>Journal Title</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Gastroenterology, 2 Hepatology, 3 Am J Gastroenterol, 4 Gut, 5 J Hepatol, 6 Am J Physiol-Gastr L, 7 Gastrointest Endose</td>
<td></td>
</tr>
<tr>
<td>32 Hepatol Res, 33 Best Pract Res Cl Ga, 34 Z Gastroentero, 35 Dis Esophagus, 36 Metab Brain Dis, 37 Digest Dis, 38 Acta Gastro-Ent Belg, 39 Curr Opin Gastroen, 40 Rev Esp Enferm Dig</td>
<td></td>
</tr>
</tbody>
</table>

### 24 Law (0.79%)

<table>
<thead>
<tr>
<th>Journal Title</th>
<th>Page Numbers</th>
</tr>
</thead>
</table>
25 Chemical Engineering (0.77%)


26 Education (0.75%)


27 Telecommunication (0.68%)


28 Orthopedics (0.68%)

### 29 Control Theory (0.64%)

<table>
<thead>
<tr>
<th>Journal</th>
</tr>
</thead>
</table>

### 30 Environmental Health (0.63%)

<table>
<thead>
<tr>
<th>Journal</th>
</tr>
</thead>
</table>

### 32 Ophthalmology (0.48%)

<table>
<thead>
<tr>
<th>Journal</th>
</tr>
</thead>
</table>

### 33 Crop Science (0.47%)

<table>
<thead>
<tr>
<th>Journal</th>
</tr>
</thead>
</table>

---

**Beograd**
34 Geography (0.47%)


35 Anthropology (0.45%)


36 Veterinary (0.45%)


37 Computer Imaging (0.43%)


38 Agriculture (0.40%)
46 Pathology (0.27%)


47 History & Philosophy of Science (0.19%)


48 Otolaryngology (0.19%)


49 Electromagnetic Engineering (0.18%)


50 Information Science (0.17%)


51 Circuits (0.17%)


52 Media & Communication (0.16%)


53 Power Systems (0.16%)


54 Tribology (0.16%)


55 History (0.15%)

56 Geotechnology (0.14%)

57 Wood Products (0.12%)

58 Radiation (0.10%)

59 Linguistics (0.093%)
1 J Pragmatics, 2 Discourse Soc, 3 Lang Soc, 4 Mod Lang J, 5 Tesol Quart, 6 Appl Linguist, 7 Res Lang Soc Interac, 8 Aphasiology, 9 Lang Learn, 10 Lang Commun, 11 Can Mod Lang Rev, 12 Foreign Lang Ann, 13 Am Speech

60 Social Work (0.079%)

61 Psychoanalysis (0.077%)
1 J Am Psychoanal Ass, 2 Int J Psychoanal, 3 Psychoanal Quart, 4 Psychoanal Inq, 5 Psychoanal Psychol, 6 Psyche-Z Psychoanal, 7 Contemp Psychoanal, 8 Forum Psychoanal, 9 J Anal Psychol

62 Middle Eastern Studies (0.073%)
1 Int J Middle E Stud, 2 New Left Rev, 3 Middle Eastern Stud, 4 Middle East J, 5 Mon Rev, 6 Sci Soc, 7 Race Class

63 Forensic Science (0.067%)

64 Transfusion (0.066%)

65 Mycology (0.055%)
1 Mycol Res, 2 Mycologia, 3 Mycotoxan, 4 Lichenologist, 5 Sydowia, 6 Persoonia, 7 Cryptogamie Mycol, 8 Mikol Fritopatol

66 Nuclear Energy (0.055%)

67 offshore Engineering (0.036%)
1 Coast Eng, 2 Ocean Eng, 3 J Waterw Port C-Asce, 4 Appl Ocean Res, 5 J Offshore Mech Arct, 6 Mar Technol Sname N, 7 Nav Eng J

68 Environmental Ethics (0.034%)
1 Ethics, 2 Soc Philos Policy, 3 Environ Ethics, 4 Environ Value, 5 J Agr Environ Ethic

69 Insect Taxonomy (0.033%)
1 P Entomol Soc Wash, 2 Entomol News, 3 T Am Entomol Soc, 4 J New York Entomol S, 5 Pan-Pac Entomol, 6 Orient Insects

70 Higher Education (0.028%)
1 J Coll Student Dev, 2 Res High Educ, 3 J High Educ, 4 Rev High Educ
71 Refineries (0.024%)
1 Oil Gas J, 2 Hydrocarb Process, 3 Neft Khoz

72 Reliability Engineering (0.023%)
1 Reliab Eng Syst Safe, 2 IEEE T Reliab, 3 P Rel Maint S

73 Other (0.023%)

74 Civil Engineering (0.022%)
1 J Comput Civil Eng, 2 J Constr Eng M Asce, 3 Build Res Inf, 4 J Surv Eng-Asce, 5 J Prof Iss Eng Ed Pr

75 Lab Veterinary (0.017%)
1 Lab Anim-Uk, 2 Comparative Med, 3 Contemp Top Lab Anim, 4 Lab Animal, 5 Scand J Lab Anim Sci

76 Music (0.017%)
1 Music Percept, 2 J New Music Res, 3 Comput Music J

77 Tourism (0.017%)
1 Ann Tourism Res, 2 Tourism Manage

78 Textiles (0.015%)
1 Text Res J, 2 Sen-I Gakkaishi, 3 Indian J Fibre Text, 4 Tekstil

79 Creativity Research (0.014%)
1 Creativity Res J, 2 J Creative Behav

80 Travel Sociology (0.012%)
1 Sociol Trav, 2 Rev Fr Sociol, 3 Mouvement Soc

81 Medical Informatics (0.011%)
1 Comput Meth Prog Bio, 2 Comput Biol Med, 3 Meas Control-Uk

82 Leprosy (0.011%)
1 Leprosy Rev, 2 Int J Leprosy

83 Sociology (Russian) (0.011%)
1 Psikhol Zh, 2 Vop Psikhol+, 3 Sotsiol Issled+

84 Cryobiology (0.010%)
1 Cryobiology, 2 Cryoletters

85 Death Studies (0.0095%)
1 Death Stud, 2 Omega-J Death Dying

86 Rehabilitation Counseling (0.0074%)
1 Rehabil Couns Bull, 2 J Rehabil

87 Steel (0.0072%)
1 Steel Res Int, 2 Stahl Eisen

88 Futurist (0.003%)
1 Futurist

Fields of social science

1 Economics (19%)

2 Psychology (18%)


3 Psychiatry (16%)


4 Healthcare (7.2%)


5 Political Science (4.9%)

6 Sociology (Behavioral) (4.4%)


7 Management (4.1%)


8 Law (3.9%)


9 Education (2.7%)


10 Geography (2.5%)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Anthropology</td>
<td>1.5%</td>
</tr>
<tr>
<td>Cultural Anthropology</td>
<td>1.2%</td>
</tr>
<tr>
<td>Marketing</td>
<td>1.0%</td>
</tr>
<tr>
<td>Information Science</td>
<td>0.99%</td>
</tr>
<tr>
<td>Philosophy of Science</td>
<td>0.98%</td>
</tr>
<tr>
<td>Sociology (Institutional)</td>
<td>0.95%</td>
</tr>
<tr>
<td>Communication</td>
<td>0.80%</td>
</tr>
<tr>
<td>Educational Assessment</td>
<td>0.70%</td>
</tr>
<tr>
<td>Educational Psychology</td>
<td>0.66%</td>
</tr>
<tr>
<td>Human-Computer Interface</td>
<td>0.54%</td>
</tr>
</tbody>
</table>

**Physical Anthropology (1.5%)**

**Cultural Anthropology (1.2%)**

**Marketing (1.0%)**

**Information Science (0.99%)**

**Philosophy of Science (0.98%)**

**Sociology (Institutional) (0.95%)**

**Communication (0.80%)**

**Educational Assessment (0.70%)**

**Educational Psychology (0.66%)**

**Human-Computer Interface (0.54%)**
<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
<th>Journals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applied Linguistics</td>
<td>0.51%</td>
<td>1 J Pragmatics, 2 Discourse Soc, 3 Lang Soc, 4 Res Lang Soc Interac, 5 Tesol Quart, 6 Mod Lang J, 7 Appl Linguist, 8 Language, 9 Lang Learn, 10 Lang Commun, 11 Linguistics, 12 Can Mod Lang Rev, 13 Lingua, 14 Am Speech, 15 Foreign Lang Ann</td>
</tr>
<tr>
<td>Experimental Psychology</td>
<td>0.49%</td>
<td>1 J Exp Anal Behav, 2 J Exp Psychol Anim B, 3 Physiol Behav, 4 Behav Process, 5 J Comp Psychol, 6 Q J Exp Psychol-B, 7 Psychol Rec, 8 Learn Motiv, 9 Behav Analyst, 10 Neurobiol Learn Mem, 11 Integr Phys Beh Sci, 12 Jpn Psychol Res</td>
</tr>
<tr>
<td>History</td>
<td>0.46%</td>
<td>1 Am Hist Rev, 2 J Am Hist, 3 Environ Hist, 4 Slavic Rev, 5 J Mod Hist, 6 Past Present, 7 Continuity Change, 8 J Fam Hist, 9 Labor Hist, 10 Int Rev Soc Hist, 11 J Soc Hist, 12 J Urban Hist, 13 J Hist Sexuality, 14 E Eur Quart, 15 Crit Rev</td>
</tr>
<tr>
<td>Speech And Hearing</td>
<td>0.40%</td>
<td>1 J Speech Lang Hear R, 2 Appl Psycholinguist, 3 Int J Lang Comm Dis, 4 J Phonetics, 5 Lang Speech Hear Ser, 6 Clin Linguist Phonet, 7 J Commun Disord, 8 Phonetica, 9 J Fluency Disord, 10 Top Lang Disord, 11 Folia Phoniatri Logo</td>
</tr>
<tr>
<td>Psychoanalysis</td>
<td>0.38%</td>
<td>1 J Am Psychoanal Ass, 2 Int J Psychoanal, 3 Psychoanal Quart, 4 Psychoanal Inq, 5 Psychoanal Psychol, 6 Psyche-Z Psychoanal, 7 Contemp Psychoanal, 8 Forum Psychoanal, 9 J Anal Psychol</td>
</tr>
<tr>
<td>Guidance Counseling</td>
<td>0.36%</td>
<td>1 J Couns Psychol, 2 Couns Psychol, 3 J Couns Dev, 4 J Career Assessment, 5 Meas Eval Couns Dev, 6 Career Dev Q, 7 J Multicult Couns D, 8 J Career Dev, 9 Brit J Guid Couns, 10 Fed Prob, 11 Women Ther</td>
</tr>
<tr>
<td>Middle Eastern Studies</td>
<td>0.35%</td>
<td>1 Int J Middle E Stud, 2 New Left Rev, 3 Middle Eastern Stud, 4 Middle East J, 5 Mon Rev, 6 Sci Soc, 7 Race Class</td>
</tr>
<tr>
<td>East Asian Studies</td>
<td>0.35%</td>
<td>1 China Quart, 2 China J, 3 Asian Surv, 4 Mod China, 5 Issues Stud, 6 Pac Rev, 7 J Contemp Asia, 8 Pac Aff</td>
</tr>
<tr>
<td>Ergonomics</td>
<td>0.34%</td>
<td>1 Ergonomics, 2 Accident Anal Prev, 3 Hum Factors, 4 Appl Ergon, 5 Int J Ind Ergonom, 6 J Safety Res, 7 Int J Aviat Psychol, 8 J Occup Rehabil</td>
</tr>
<tr>
<td>Medical Ethics</td>
<td>0.33%</td>
<td>1 Hastings Cent Rep, 2 J Med Ethics, 3 J Law Med Ethics, 4 Bioethics, 5 Kennedy Inst Ethic J, 6 J Clin Ethic, 7 J Med Philos, 8 Health Care Anal, 9 Camb Q Healthe Ethic, 10 Sci Eng Ethics, 11 Theor Med Bioeth</td>
</tr>
<tr>
<td>Ethics</td>
<td>0.26%</td>
<td>1 Ethics, 2 Inquiry, 3 Soc Philos Policy, 4 Environ Ethics, 5 Hum Stud, 6 Environ Value, 7 J Agr Environ Ethic</td>
</tr>
</tbody>
</table>
36 Economic History (0.26%)
1 J Econ Hist, 2 Econ Hist Rev, 3 Explor Econ Hist, 4 Bus Hist, 5 Soc Sci Hist, 6 J Interdiscipl Hist, 7 Agr Hist

37 Sport Psychology (0.25%)
1 Res Q Exercise Sport, 2 J Sport Exercise Psy, 3 Quest, 4 J Appl Sport Psychol, 5 J Teach Phys Educ, 6 Sport Psychol, 7 Int J Sport Psychol, 8 Sociol Sport J, 9 J Sport Hist

38 Public Affairs (0.22%)
1 Public Admin, 2 Parliament Aff, 3 Polit Quart, 4 Local Gov Stud, 5 Public Money Manage, 6 Public Admin Develop, 7 Aust J Publ Admin

39 Social Policy (0.19%)

40 Family Relations (0.18%)
1 Fam Relat, 2 Fam Process, 3 J Marital Fam Ther, 4 Am J Fam Ther, 5 J Fam Ther, 6 Contemp Fam Ther

41 Law And Behavior (0.17%)
1 Law Human Behav, 2 Psychol Public Pol L, 3 Psychol Crime Law

42 Criminology (0.16%)

43 Sexuality (0.16%)
1 Arch Sex Behav, 2 J Sex Res, 3 J Sex Marital Ther, 4 J Homosexual

44 Higher Education (0.14%)
1 J Coll Student Dev, 2 Res High Educ, 3 J High Educ, 4 Rev High Educ

45 Leisure Studies (0.14%)
1 Environ Behav, 2 J Environ Psychol, 3 J Leisure Res, 4 Leisure Sci

46 Neurorehabilitation (0.14%)
1 J Head Trauma Rehab, 2 Rehabil Psychol, 3 Neurorehabilitation, 4 Rehabil Couns Bull, 5 J Rehabil, 6 Int J Rehabil Res

47 Pacific Studies (0.10%)
1 Contemp Pacific, 2 Oceania, 3 J Polynesian Soc, 4 J Mat Cult

48 Tourism (0.093%)
1 Ann Tourism Res, 2 Tourism Manage

49 Creativity (0.073%)
1 Creativity Res J, 2 J Creative Behav

50 Death And Dying (0.061%)
1 Death Stud, 2 Omega-J Death Dying

51 Sociology (French) (0.059%)
1 Psikhol Zh, 2 Vop Psikhol+, 3 Sociol Cas, 4 Sotsiol Issled+

52 Sociology (Eastern Europe) (0.054%)
1 Sociol Trav, 2 Rev Fr Sociol, 3 Mouvement Soc

53 Maritime Law (0.048%)
1 Mar Policy, 2 Ocean Dev Int Law

54 Hypnosis (0.028%)
1 Int J Clin Exp Hyp, 2 Am J Clin Hypn