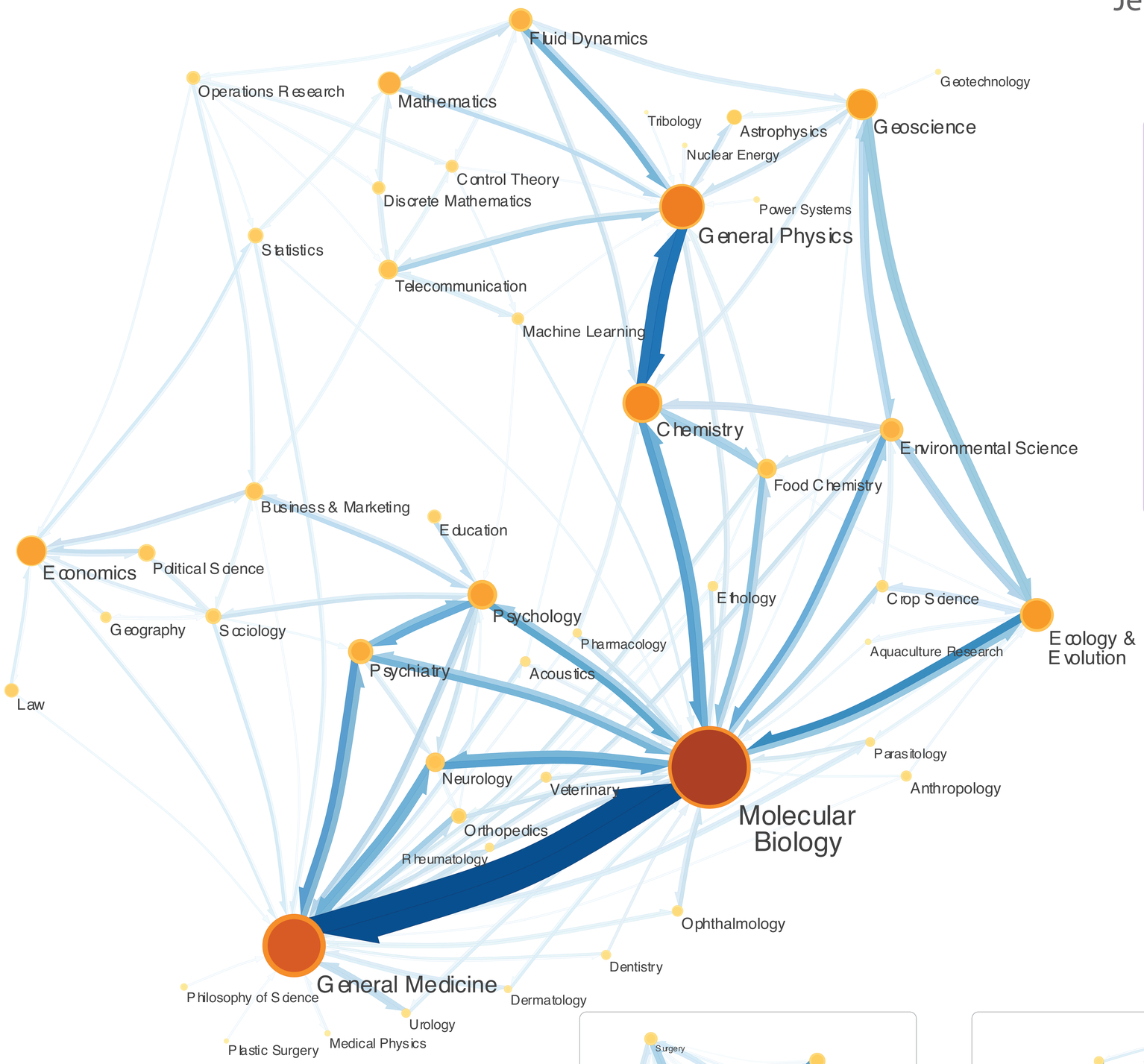
Ranking and Mapping Scholarly Literature

Jevin West, Martin Rosvall, Ben Althouse, Carl Bergstrom

University of Washington, Department of Biology, Seattle, WA, 98105



Eigenfactor.org: A New Measure of Journal Quality

Using a method similar to that used by Google to rank webpages, Eigenfactor.org is a new way to measure the importance of scholarly literature. By this approach, journals are considered to be influential if they are cited often by other influential journals. Iterative ranking schemes of this type are known as eigenvector centrality methods.

The Eigenfactor score of a journal is an estimate of the percentage of time that library users spend with that journal. The Eigenfactor algorithm corresponds to a simple model of research in which readers follow chains of citations as they move from journal to journal. Imagine that a researcher goes to the library and selects a journal article at random. After reading the article, the researcher selects at random one of the citations from the article. She then proceeds to the journal that was cited, reads a random article there, and selects a citation to direct her to her next journal volume. The researcher does this ad infinitum.

The amount of time that the researcher spends with each journal gives us a measure of that journal's importance within network of academic citations. Moreover, if real researchers find a sizable fraction of the articles that they read by following citation chains, the amount of time that our random researcher spends with each journal gives us an estimate of the amount of time that real researchers spend with each journal. While we cannot carry out this experiment in practice, we can use mathematics to simulate this process (see Eigenfactor.org).

Top 10 Science Journals

1. Nature
2. Science
3. Journal of Biological Chemistry
4. PNAS
5. Physical Review Letters
6. Physical Review B
7. Journal of the Americal Chem. Soc.
8. Cell
9. Applied Physics Letters
10. New England Journal of Medicine

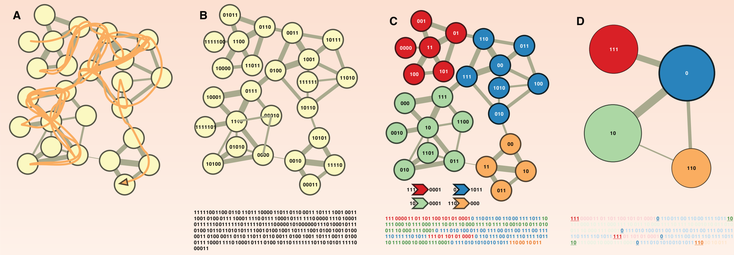
Top 10 Social Science Journals

1. Am. Journal of Psychiatry
2. American Economic Review
3. Archives of General Psychiatry
4. Journal of Finance
5. Journal of Personality and Social Psych.
6. Quarterly Journal of Econonmics
7. Am. Journal of Public Health
8. Econometrica
9. Journal of Political Economy
10. Journal of Clin. Psychiatry

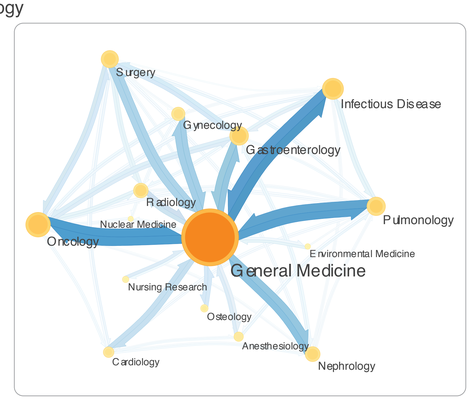
Top 10 Microbiology Journals

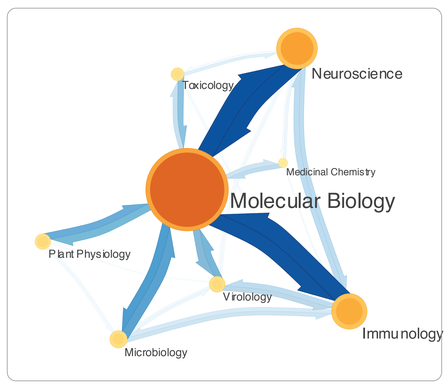
1. Journal of Bacteriology
2. Applied and Environmental Microbio.
3. Molecular Microbiology
4. Journal of Clinical Microbiology
5. Clinical Infectious Diseases
6. Antimicrobial Agents and Chemotherapy
7. FEMS Microbiology Letters
8. Microbiology - SGM
9. Journal of Antimicrobial Chemotherapy
10. Current Opinion in Microbiology

The Mapping Idea



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. (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. (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. Module codes and exit codes are shown left and right of the arrows under the network. (D) Reporting only the module names, and not the locations within the modules, provides an optimal coarse-graining of the network.





The Map Equation

To find an optimal code, we look for a modular map M that partitions the n nodes into m modules so as to minimize the expected description length of a random walk. Using map M, the average description length of a single step is given by

$$L(M) = q_{\curvearrowright} H(Q) + \sum_{i=1}^{m} p^{i}_{\circlearrowright} H(P^{i}).$$

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 is also considered a movement). Each is weighted by the frequency with which it occurs in the particular partitioning. Here $q_{\curvearrowright}$ is the probability that the random walk switches modules on any given step. H(Q) is the entropy of the module names, i.e., the entropy of the underlined codewords in Fig. 1D. $H(P^{i})$ is the entropy of the within module movements — including the exit code for module i. The weight $p^{i}_{\circlearrowright}$ is the fraction of within module movements that occur in module i, plus the probability of exiting module i such that $\sum_{i=1}^{m} p^{i}_{\circlearrowright} = 1 + q_{\curvearrowright}$.

eigenFACTOR.org

RANKING AND MAPPING SCIENTIFIC KNOWLEDGE