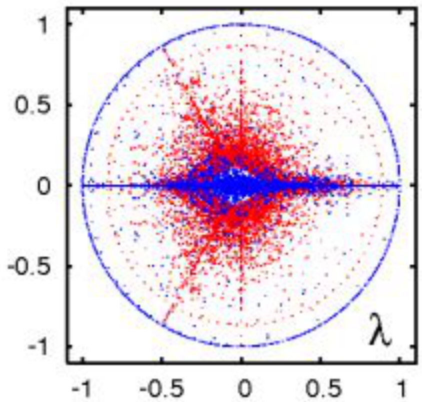
IOPscience

Universal emergence of PageRank

In the past few years, social and communication networks have become of significant importance. The World Wide Web alone has about 50 billion indexed webpages, so their classification and information retrieval becomes a formidable task that users have to face everyday. Various search engines have been developed by private companies, such as Google, Yahoo! and others, which are extensively used by internet users. To handle such enormous databases, fundamental mathematical tools and algorithms should be developed.

The PageRank algorithm forms the basis of Google's search engine. It builds Google's matrix of web links and determines the PageRank eigenvector with unit eigenvalue. This vector gives an importance value to all nodes thus enabling ranking of them. In general, Google's matrix has all eigenvalues inside the unit circle in the complex plane. The matrix depends on the damping parameter which creates a gap between the unit eigenvalue and all others; the gap disappears when the parameter goes to unity.

It is striking that the emergence of the PageRank vector at small gap values is characterized by a number of universal features related to the eigenvalue spectrum. Powerful numerical methods have been developed to compute the largest few thousand eigenvalues and associated eigenvectors for Google matrices with up to several million network nodes, with a special accent on British university web networks. The obtained results show that, in general, the whole network can be divided into a core part and a group of invariant subspaces. Consequently, as can be seen in the example of the Cambridge University web network shown (spectrum shown in the figure), the complex spectrum of the Google matrix is composed of a part associated with the core space (in red in the figure) and another associated with the invariant subspaces which have many unit eigenvalues at zero gap (in blue in the figure). In the limit of a small gap, these subspaces determine the universal algebraic decay of the PageRank vector with the rank index. The size distribution of invariant subspaces also follows a universal power law. The above properties create large relaxation times on the network as, for example, in spin glasses. The obtained results allow us to understand the dependence of PageRank on the damping parameter in a deeper manner.



(http://images.iop.org/objects/jio/insights/1/11/1/fig1.jpg)
Figure 1. The spectrum of the Google matrix of the web network of Cambridge University in 2006 at unit damping parameter. (http://images.iop.org/objects/jio/insights/1/11/1/fig1.jpg)

The paper was originally published in *J. Phys. A: Math. Theor.* **46** 465101 (http://iopscience.iop.org/1751-8121/44/46/465101) .


(http://images.iop.org/objects/jio/insights/1/11/1/authors.jpg)
Authors from left to right: B Georgeot, K M Frahm, D L Shepelyansky (photo courtesy of D L Shepelyansky) (http://images.iop.org/objects/jio/insights/1/11/1/authors.jpg)

About the author

Klaus M Frahm is a professor at the University Paul Sabatier. Bertrand Georgeot and Dima L Shepelyansky are Directeurs de Recherche at CNRS. They all work at the Laboratoire de Physique Theorique, Toulouse. (http://w w w .quantw are.ups-tlse.fr)