

Google matrix

From Wikipedia, the free encyclopedia

A **Google matrix** is a particular stochastic matrix that is used by Google's PageRank algorithm. The matrix represents a graph with edges representing links between pages. The rank of each page can be generated iteratively from the Google matrix using the power method. However, in order for the power method to converge, the matrix must be stochastic, irreducible and aperiodic.

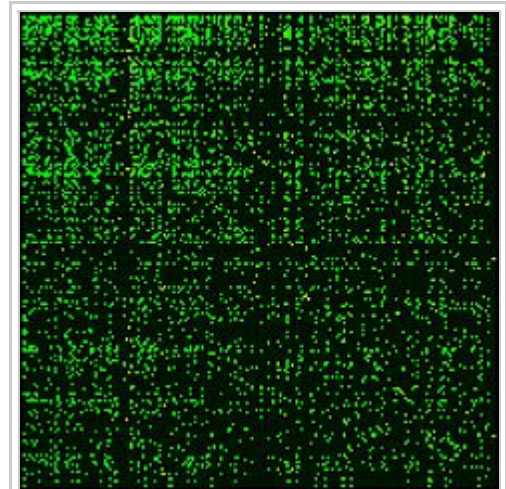


Fig.1. Google matrix of Wikipedia articles network, written in the bases of PageRank index; fragment of top 200 X 200 matrix elements is shown, total size $N=3282257$ (from [19])

Contents

- 1 Adjacency matrix A and Markov matrix S
- 2 Construction of Google matrix G
- 3 Examples of Google matrix
- 4 Spectrum and eigenstates of G matrix
- 5 Historical notes
- 6 See also
- 7 References

Adjacency matrix A and Markov matrix S

In order to generate the Google matrix G , we must first generate an adjacency matrix A which represents the relations between pages or nodes.

Assuming there are N pages, we can fill out A by doing the following:

1. A matrix element $A_{i,j}$ is filled with 1 if node j has a link to node i , and 0 otherwise; this is the adjacency matrix of links.
2. A related matrix S corresponding to the transitions in a Markov chain of given network is constructed from A by dividing the elements of column "j" by a number of k_j where k_j is the total number of outgoing links from node j to all other nodes. The columns having zero matrix elements, corresponding to dangling nodes, are replaced by a constant value $1/N$. Such a procedure adds a link from every sink, dangling state a to every other node.
3. Now by the construction the sum of all elements in any column of matrix S is equal to unity. In this way the matrix S is mathematically well defined and it belongs to the class of Markov chains and the class of Perron-Frobenius operators. That makes S suitable for the PageRank algorithm.

Construction of Google matrix G

Then the final Google matrix G can be expressed via S as:

$$G_{ij} = \alpha S_{ij} + (1 - \alpha) \frac{1}{N} \quad (1)$$

By the construction the sum of all non-negative elements inside each matrix column is equal to unity. The numerical coefficient α is known as a damping factor.

Usually S is a sparse matrix and for modern directed networks it has only about ten nonzero elements in a line or column, thus only about $10N$ multiplications are needed to multiply a vector by matrix G [1,2].

Examples of Google matrix

An example of the matrix S construction via Eq.(1) within a simple network is given in the article CheiRank.

For the actual matrix, Google uses a damping factor α around 0.85 [1,2,3]. The term $(1 - \alpha)$ gives a surfer probability to jump randomly on any page. The matrix G belongs to the class of Perron-Frobenius operators of Markov chains [1]. The examples of Google matrix structure are shown in Fig.1 for Wikipedia articles hyperlink network in 2009 at small scale and in Fig.2 for University of Cambridge network in 2006 at large scale.

Spectrum and eigenstates of G matrix

For $0 < \alpha < 1$ there is only one maximal eigenvalue $\lambda = 1$ with the corresponding right eigenvector which has non-negative elements P_i which can be viewed as stationary probability distribution [1]. These probabilities ordered by their decreasing values give the PageRank vector P_i with the Ragerank K_i used by Google search to rank webpages. Usually one has for the World Wide Web that $P \propto 1/K^\beta$ with $\beta \approx 0.9$. The number of nodes with a given PageRank value scales as $N_P \propto 1/P^\nu$ with the exponent $\nu = 1 + 1/\beta \approx 2.1$ [4,5]. The left eigenvector at $\lambda = 1$ has constant matrix elements. With $0 < \alpha$ all eigenvalues move as $\lambda_i \rightarrow \alpha \lambda_i$ except the maximal eigenvalue $\lambda = 1$, which remains unchanged [1]. The PageRank vector varies with α but other eigenvectors with $\lambda_i < 1$ remain unchanged due to their orthogonality to the constant left vector at $\lambda = 1$. The gap between $\lambda = 1$ and

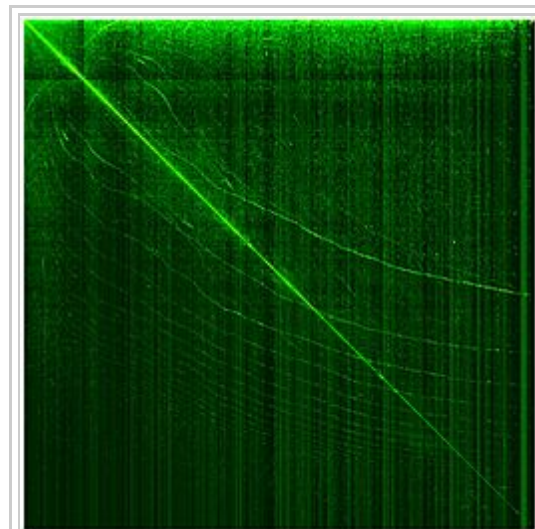


Fig.2. Google matrix of Cambridge University network (2006), coarse-grained matrix elements are written in the bases of PageRank index, total size $N=212710$ is shown (from [19])

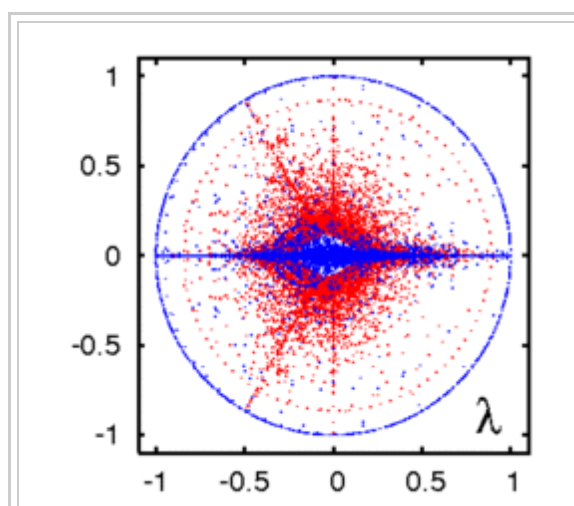


Fig3. The spectrum of eigenvalues of the Google matrix of University of Cambridge from Fig.2 at $\alpha = 1$, blue points show eigenvalues of isolated subspaces, red points show eigenvalues

other eigenvalue is $1 - \alpha \approx 0.15$ gives a rapid convergence of a random initial vector to the PageRank approximately after 50 multiplications on G matrix.

At $\alpha = 1$ the matrix G has generally many degenerate eigenvalues $\lambda = 1$ (see e.g. [6,7]). Examples of the eigenvalue spectrum of the Google matrix of various directed networks is shown in Fig.3 from [14] and Fig.4 from [7].

The Google matrix can be also constructed for the Ulam networks generated by the Ulam method [8] for dynamical maps. The spectral properties of such matrices are discussed in [9,10,11,12,13,14,15,16]. In a number of cases the spectrum is described by the fractal Weyl law [10,12].

The Google matrix can be constructed also for other directed networks, e.g. for the procedure call network of the Linux Kernel software introduced in [15]. In this case the spectrum of λ is described by the fractal Weyl law with the fractal dimension $d \approx 1.3$ (see Fig.5 from [16]). Numerical analysis shows that the eigenstates of matrix G are localized (see Fig.6 from [16]). Arnoldi iteration method allows to compute many eigenvalues and eigenvectors for matrices of rather large size [13,14,16].

Other examples of G matrix include the Google matrix of brain [17] and business process management [18], see also [19].

Historical notes

The Google matrix with damping factor was described by Sergey Brin and Larry Page in 1998 [20], see also articles PageRank and [21],[22].

See also

- PageRank, CheiRank
- Arnoldi iteration
- Markov chain, Transfer operator, Perron–Frobenius theorem
- Web search engines

References

of core component (from [14])

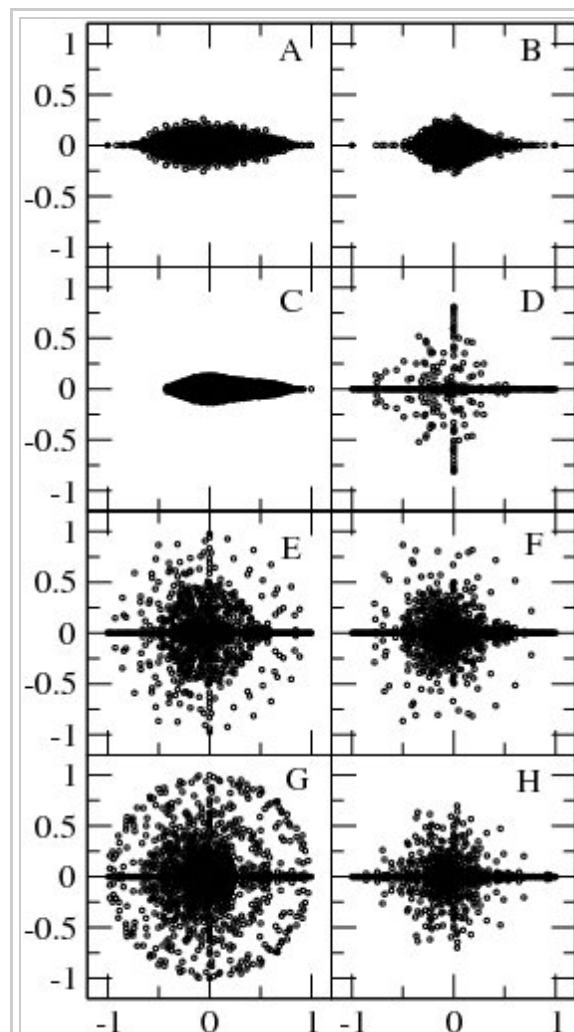


Fig4. Distribution of eigenvalues λ_i of Google matrices in the complex plane at $\alpha = 1$ for dictionary networks: Roget (A, N=1022), ODLIS (B, N=2909) and FOLDOC (C, N=13356); UK university WWW networks: University of Wales (Cardiff) (D, N=2778), Birmingham City University (E, N=10631), Keele University (Staffordshire) (F, N=11437), Nottingham Trent University (G, N=12660), Liverpool John Moores University (H, N=13578)(data for universities are for 2002) (from [7])

- [1] Langville, Amity N; Carl Meyer (2006). *Google's PageRank and Beyond*. Princeton University Press. ISBN 0-691-12202-4.
- [2] Austin, David (2008). "How Google Finds Your Needle in the Web's Haystack" (<http://www.ams.org/samplings/feature-column/fcarc-pagerank>) . AMS Feature Columns. <http://www.ams.org/samplings/feature-column/fcarc-pagerank>
- [3] Law, Edith (2008). "PageRank" (http://scienceoftheweb.org/15-396/lectures/PageRank_Lecture12.pdf) (PDF). http://scienceoftheweb.org/15-396/lectures/PageRank_Lecture12.pdf
- [4] Donato D.; Laura L., Leonardi S., Millozzi S. (2004). "Large scale properties of the Webgraph". *Eur. Phys. J. B* v.38, p.239
- [5] Pandurangan G.; Raghavan P., Upfal E. (2005). "Using PageRank to Characterize Web Structure". *Internet Math.* v.3, p. 1
- [6] Serra-Capizzano, Stefano (2005). "Jordan Canonical Form of the Google Matrix: a Potential Contribution to the PageRank Computation". *SIAM J. Matrix. Anal. Appl.* v.27, p.305
- [7] Georgeot B.; Giraud O., Shepelyansky D.L. (2010). "Spectral properties of the Google matrix of the World Wide Web and other directed networks". *Phys. Rev. E* v.81, p.056109
- [8] Ulam, Stanislaw (1960). "A Collection of mathematical problems, Interscience tracs in pure and applied mathematics". Interscience, New York p.73
- [9] Froyland G.; Padberg K. (2009). "Almost-invariant sets and invariant manifolds — Connecting probabilistic and geometric descriptions of coherent structures in flows". *Physica D* v.238, p.1507
- [10] Shepelyansky D.L.; Zhironov O.V. (2010). "Google matrix, dynamical attractors and Ulam networks". *Phys. Rev. E* v.81, p.036213
- [11] Ermann L.; Shepelyansky D.L. (2010). "Google matrix and Ulam networks of intermittency maps". *Phys. Rev. E* v.81, p.036221
- [12] Ermann L.; Shepelyansky D.L. (2010). "Ulam method and fractal Weyl law for Perron-Frobenius operators". *Eur. Phys. J. B* v.75, p.299
- [13] Frahm K.M.; Shepelyansky D.L. (2010). "Ulam method for the Chirikov standard map". *Eur. Phys. J. B* v.76, p.57
- [14] Frahm K.M.; Georgeot B, Shepelyansky D.L. (2011). *Universal emergence of PageRank*. *J. Phys. A: Mat. Theor.* v.44, p.465101. arXiv:1105.1062 (<http://arxiv.org/abs/1105.1062>)
- [15] Chepelienskii, Alexei D. (2010). *Towards physical laws for software architecture*. arXiv:1003.5455 (<http://arxiv.org/abs/1003.5455>)
- [16] Ermann L.; Chepelienskii A.D., Shepelyansky D.L. (2011). "Fractal Weyl law for Linux Kernel Architecture". *Eur. Phys. J. B* v.79, p.115. arXiv:1005.1395 (<http://arxiv.org/abs/1005.1395>)
- [17] Shepelyansky D.L.; Zhironov O.V. (2010). "Towards Google matrix of brain". *Phys. Lett. A* v.374, p.3206

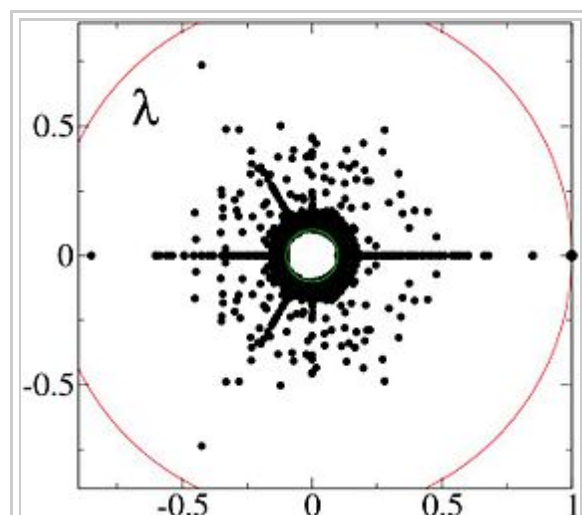


Fig5. Distribution of eigenvalues λ in the complex plane for the Google matrix G of the Linux Kernel version 2.6.32 with matrix size $N = 285509$ at $\alpha = 0.85$, unit circle is shown by solid curve (from [16])

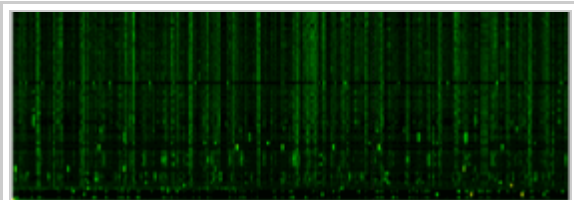


Fig.6 Coarse-grained probability distribution for the eigenstates of the Google matrix for Linux Kernel version 2.6.32. The horizontal lines show the first 64 eigenvectors ordered vertically by $|\lambda_i|$ (from [16])

- [18] Abel M.; Shepelyansky D.L. (2011). "Google matrix of business process management". *Eur. Phys. J. B* v.84, p.493. arXiv:1009.2631 (<http://arxiv.org/abs/1009.2631>)
- [19] Ermann L.; Chepelianskii A.D, Shepelyansky D.L. (2011). *Towards two-dimensional search engines*. arXiv:1106.6215 (<http://arxiv.org/abs/1106.6215>)
- [20] Brin S.; Page L. (1998). "The anatomy of a large-scale hypertextual Web search engine". *Computer Networks and ISDN Systems* v.30, p.107
- [21] Franceschet, Massimo (2010). *PageRank: Standing on the shoulders of giants*. arXiv:1002.2858 (<http://arxiv.org/abs/1002.2858>)
- [22] Vigna, Sebastiano (2010). "Spectral Ranking" (<http://vigna.dsi.unimi.it/ftp/papers/SpectralRanking.pdf>) . personal webpage. <http://vigna.dsi.unimi.it/ftp/papers/SpectralRanking.pdf>

Retrieved from "http://en.wikipedia.org/w/index.php?title=Google_matrix&oldid=535519317"

Categories: [Google](#) | [Link analysis](#) | [Markov models](#)

- This page was last modified on 29 January 2013 at 14:27.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of Use for details.
Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.