Available online at www.sciencedirect.com
ScienceDirect
Journal of Informetrics 1 (2007) 8–15
Journal of INFORMETRICS
An International Journal
www.elsevier.com/locate/joi

Finding scientific gems with Google's PageRank algorithm

P. Chen[a,*], H. Xie[b,c], S. Maslov[c], S. Redner[a]

[a] *Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, United States*
[b] *New Media Lab, The Graduate Center, CUNY, New York, NY 10016, United States*
[c] *Department of Condensed Matter Physics and Materials Science, Brookhaven National Laboratory, Upton, NY 11973, United States*

Received 27 April 2006; received in revised form 12 June 2006; accepted 12 June 2006

Abstract

We apply the Google PageRank algorithm to assess the relative importance of all publications in the Physical Review family of journals from 1893 to 2003. While the Google number and the number of citations for each publication are positively correlated, outliers from this linear relation identify some exceptional papers or "gems" that are universally familiar to physicists.
© 2006 Published by Elsevier Ltd.

PACS: 02.50.Ey; 05.40.-a; 05.50.+q; 89.65.-s

Keywords: Google PageRank algorithm; Scientific gems; Physical Review; Citations

1. Introduction

With the availability of electronically available citation data, it is now possible to undertake comprehensive studies of citations that were unimaginable just a few years ago. In most previous studies of citation statistics, the metric used to quantify the importance of a paper is its number of citations. In terms of the underlying citation network, in which nodes represent publications and directed links represent a citation from a *citing* article to a *cited* article, the number of citations to an article translates to the in-degree of the corresponding node. The distribution of in-degree for various citation data sets has a broad tail (Price, 1965, 1976) that is reasonably approximated by a power law (Laherrère & Sornette, 1998; Redner, 1998, 2005).

While the number of citations is a natural measure of the impact of a publication, we probably all have encountered examples where citations do not seem to provide a full picture of the influence of a publication. We are thus motivated to study alternative metrics that might yield a truer measure of importance than citations alone. Such a metric already exists and is provided by the Google PageRank (Brin & Page, 1998) or similar algorithms proposed for the analysis of social (Bonacich, 1972, 1987) and information (Kleinberg, 1999) networks. By simulating random traffic on a network these algorithms calculate the importance of papers in a self-consistent fashion. They naturally take into account the following two factors: (1) the effect of receiving a citation from a more important paper should be greater that that coming from a less popular one; (2) citation links coming from a paper with a long reference list should count less than those coming from one with a short list. In other words, the importance of a paper should be divided over a number of references that inspired this line of research.

* Corresponding author.
E-mail addresses: patrick@bu.edu (P. Chen), hxie@bnl.gov (H. Xie), maslov@bnl.gov (S. Maslov), redner@bu.edu (S. Redner).

1751-1577/$ – see front matter © 2006 Published by Elsevier Ltd.
doi:10.1016/j.joi.2006.06.001

A variant of the PageRank algorithm was recently applied to better calibrate the impact factor of scientific journals (Bollen, Rodriguez, & Van de Sompel, 2006). In this work, we apply Google PageRank to the Physical Review citation network with the goal of measuring the importance of individual scientific publications published in the APS journals. This network consists of 353,268 nodes that represent all articles published in the Physical Review family of journals from the start of publication in 1893 until June 2003, and 3,110,839 links that represent all citations *to* Physical Review articles *from* other Physical Review articles. As found previously (Redner, 2005), these internal citations represent 1/5 to 1/3 of all citations for highly-cited papers. This range provides a sense of the degree of completeness of the Physical Review citation network.

With the Google PageRank approach, we find a number of papers with a modest number of citations that stand out as exceptional according to the Google PageRank analysis. These exceptional publications, or gems, are familiar to almost all physicists because of the very influential contents of these articles. Thus, the Google PageRank algorithm seems to provide a new and useful measure of scientific quality.

2. The Google PageRank algorithm

To set the stage for our use of Google PageRank to find scientific gems, let us review the elements of the PageRank algorithm. Given a network of N nodes $i = 1, 2, \ldots, N$, with directed links that represent references from an initial (citing) node to a target (cited) node, the Google number G_i for the ith node is defined by the recursion formula (Brin & Page, 1998):

$$G_i = (1 - d) \sum_{j \text{ nn } i} \frac{G_j}{k_j} + \frac{d}{N}. \tag{1}$$

Here the sum is over the neighboring nodes j in which a link points to node i. The first term describes propagation of the probability distribution of a random walk in which a walk at node j propagates to node i with probability $1/k_j$, where k_j is the out-degree of node j. The second term describes the uniform injection of probability into the network in which each node receives a contribution d/N at each step.

Here d is a free parameter that controls the performance of the Google PageRank algorithm. The prefactor $(1 - d)$ in the first term gives the fraction of random walks that continue to propagate along the links; a complementary fraction d is uniformly re-injected into the network, as embodied by the second term.

We suggest that the Google number G_i of paper i, defined by Eq. (1), is a better measure of importance than the number of citations alone in two aspects: (i) being cited by influential papers contributes more to the Google number than being cited by unimportant papers; (ii) being cited by a paper that itself has few references gives a larger contribution to the Google number than being cited by a paper with hundreds of references. The Google number of a paper can be viewed as a measure of its influence that is then equally exported to all of its references. The parameter $d > 0$ prevents all of the influence from concentrating on the oldest papers.

In the original Google PageRank algorithm of Brin and Page (1988), the parameter d was chosen to be 0.15. This value was prompted by the anecdotal observation that an individual surfing the web will typically follow of the order of 6 hyperlinks, corresponding to a leakage probability $d = 1/6 \simeq 0.15$, before becoming either bored or frustrated with this line of search and beginning a new search. In the context of citations, we conjecture that entries in the reference list of a typical paper are collected following somewhat shorter paths of average length 2, making the choice $d = 0.5$ more appropriate for a similar algorithm applied to the citation network. The empirical observation justifying this choice is that approximately 50% of the articles[1] in the reference list of a given paper A have at least one citation $B \to C$ to another article C that is also in the reference list of A. Assuming that such "feed-forward" loops result from authors of paper A following references of paper B, we estimate the probability $1 - d$ to follow this indirect citation path to be close to 0.5.

To implement the Google PageRank algorithm for the citation network, we start with a uniform probability density equal to $1/N$ at each node of the network and then iterate Eq. (1). Eventually a steady state set of Google numbers for each node of the network is reached. These represent the occupation probabilities at each node for the random-walk-like process defined by Eq. (1). Finally, we sort the nodal Google numbers to determine the Google rank of each node. It

[1] The actual fraction of "followed citations" (such as B in an A → B, A → C, and B → C loop) is 42% for the entire dataset and 51% for papers published during the last 4 years.

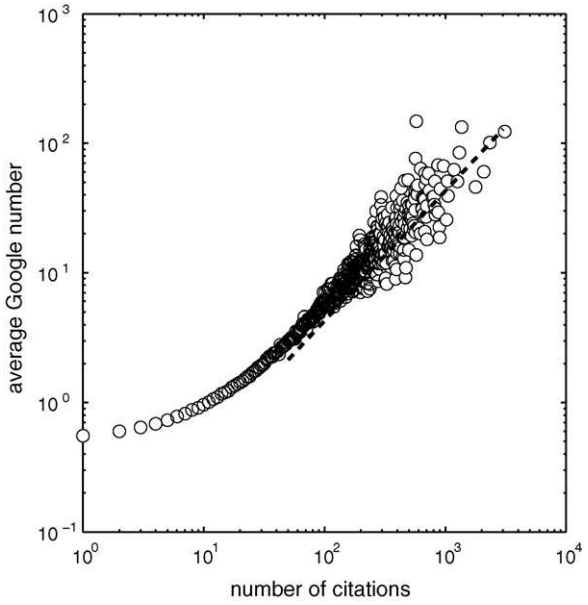

Fig. 1. Average Google number $\langle G(k)\rangle$ vs. number of citations k. The dashed line of slope 1 is a guide for the eye.

is both informative and entertaining to compare the Google rank with the citation (in-degree) rank of typical and the most important publications in Physical Review.

3. Google's PageRank for Physical Review

Fig. 1 shows the average Google number $\langle G(k)\rangle$ for publications with k citations as a function of k. For small k, there are many publications with the same number of citations and the dispersion in $G(k)$ is small. Correspondingly, the plot of $\langle G(k)\rangle$ versus k is smooth and increases linearly with k for $k \gtrsim 50$. Thus, the average Google number and the number of citations represent similar measures of popularity, a result that has been observed previously (Fortunato, Boguna, Flammini, & Menczer; Fortunato, Flammini, & Menczer). In fact, the citation and Google number distributions are

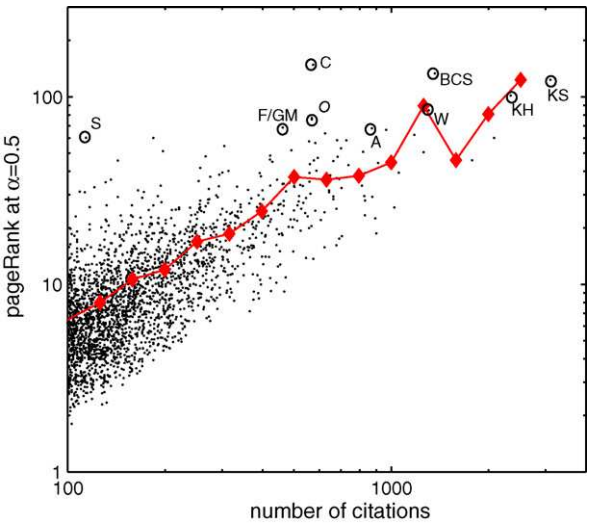

Fig. 2. Individual outlier publications. The scatter plot of the Google number vs. the number of citations. The top-10 Google-ranked papers are identified by author(s) initials (see Table 1). As a guide to the eye, the solid curve is logarithmically binned average of the data of $\langle G(k)\rangle$ vs. k in Fig. 1.

Table 1
The top-10 Google-ranked publications when $d = 0.5$

Google rank	Google # ($\times 10^{-4}$)	Cite rank	# cites	Publication				Title	Author(s)
1	4.65	54	574	PRL	10	531	1963	Unitary symmetry and leptonic...	N. Cabibbo
2	4.29	5	1364	PR	108	1175	1957	Theory of superconductivity	J. Bardeen, L. Cooper, and J. Schrieffer
3	3.81	1	3227	PR	140	A1133	1965	Self-consistent equations...	W. Kohn and L.J. Sham
4	3.17	2	2460	PR	136	B864	1964	Inhomogeneous electron gas	P. Hohenberg and W. Kohn
5	2.65	6	1306	PRL	19	1264	1967	A model of leptons	S. Weinberg
6	2.48	55	568	PR	65	117	1944	Crystal statistics I	L. Onsager
7	2.43	56	568	RMP	15	1	1943	Stochastic problems in...	S. Chandrasekhar
8	2.23	95	462	PR	109	193	1958	Theory of the Fermi interaction	R.P. Feynman and M. Gell-Mann
9	2.15	17	871	PR	109	1492	1958	Absence of diffusion in...	P.W. Anderson
10	2.13	1853	114	PR	34	1293	1929	The theory of complex spectra	J.C. Slater

qualitatively similar, further indicating that citations and Google numbers are, on the average, similar measures of importance.

However, for large k, much more interesting behavior occurs. When k is sufficiently large, there is typically only one publication with k citations. Of particular interest are the extreme outliers with respect to the linear behavior of Fig. 1. The 10 articles with the highest Google numbers are marked with large circles in Fig. 2 on the background of Google number versus the number of citations for all publications (dots). The top-10 papers are identified by author initials (see Table 1). Table 1 also lists the number of citations and the citation rank of these publications. While several of the highest-cited Physical Review papers appear on this list, there are also several more modestly-cited papers that are highly ranked according to the Google algorithm.

The disparity between the Google rank and citation rank arises because, as mentioned in the previous section, the former involves both the in-degree as well as the Google PageRank of the neighboring nodes. According to the Google algorithm of Eq. (1), a citing publication ("child") j contributes a factor $\langle G_j/k_j\rangle$ to the Google number of its parent paper i. Thus, for a paper to have a large Google number, its children should be important (large G_j), and also each child should have a small number of parents (small out-degree k_j). The latter ensures that the Google contribution of a child is not strongly diluted.

With this perspective, let us compare the statistical measures of the two articles "Unitary Symmetry and Leptonic Decays", Phys. Rev. Lett. 10, 531 (1963) by N. Cabibbo (C) and "Self-Consistent Equations Including Exchange and Correlation Effects", Phys. Rev. 140, A1133 (1965) by W. Kohn & L. J. Sham (KS). The former has the highest Google number of all Physical Review publications, while the latter is the most cited. The high Google rank of C stems from the fact that that value of $\langle G_j/k_j\rangle = 1.52 \times 10^{-6}$ for the children of C is an order of magnitude larger than the corresponding value $\langle G_j/k_j\rangle = 2.31 \times 10^{-7}$ for the children of KS. This difference more than compensates for the factor 5.6 difference in the number of citations to these two articles (3227 for KS and 574 for C as of June 2003). Looking a little deeper, the difference in $\langle G_j/k_j\rangle$ for C and KS stems from the denominator; the children of C have 15.6 citations an average, while the children of KS are slightly "better" and have 18.4 citations on average. However, the typical child of C has fewer references than a child of KS and a correspondingly larger contribution to the Google number of C.

The remaining research articles on the top-10 Google-rank list but outside the top-10 citation list are easily recognizable as seminal publications. For example, Onsager's 1944 paper presents the exact solution of the two-dimensional Ising model; both a calculational *tour de force*, as well as a central development in the theory of critical phenomena. The paper by Feynman and Gell-Mann introduced the *V-A* theory of weak interactions that incorporated parity non-conservation and became the "standard model" of weak interactions. Anderson's paper, "Absence of Diffusion in Certain Random Lattices" gave birth to the field of localization and is cited by the Nobel prize committee for the 1977 Nobel prize in physics.

The last entry in the top-10 Google-rank list, "The Theory of Complex Spectra", by J. C. Slater (S) is particularly striking. This article has relatively few citations (114 as of June 2003) and a relatively low citation rank (1853th),

Table 2
The remaining top-100 Google-ranked papers when $d = 0.5$ in which the ratio of Google rank to citation rank is greater than 10

Google rank	Google # ($\times 10^{-4}$)	Cite rank	# cites	Publication				Title	Author(s)
1	4.65	54	574	PRL	10	531	1963	Unitary symmetry and leptonic...	N. Cabibbo
8	2.23	95	462	PR	109	193	1958	Theory of the fermi interaction	R.P. Feynman and M. Gell-Mann
10	2.13	1853	114	PR	34	1293	1929	The theory of complex spectra	J.C. Slater
12	2.11	712	186	PR0	43	804	1933	On the constitution of...	E. Wigner and F. Seitz
20	1.80	228	308	PR0	106	364	1957	Correlation energy of an ...	M. Gell-Mann and K. Brueckner
21	1.69	616	198	PRL	58	408	1987	Bulk superconductivity at ...	R.J. Cava, et al.
25	1.58	311	271	PRL	58	405	1987	Evidence for superconductivity ...	C.W. Chu, et al.
30	1.51	1193	144	PRL	10	84	1963	Photon correlations	R.J. Glauber
35	1.42	12897	39	PR0	35	509	1930	Cohesion in monovalent metals	J.C. Slater
49	1.21	1342	136	PR0	60	252	1941	Statistics of the two- ...	H.A. Kramers and G.H. Wannier
58	1.17	1433	135	PR0	81	440	1951	Interaction between the ...	C. Zener
59	1.17	5196	66	PR0	45	794	1934	Electronic energy bands in ...	J.C. Slater
60	1.16	2927	108	PRB	28	4227	1983	Electronic structure of ...	L.F. Mattheiss & D. R. Hamann
64	1.12	642	199	PR0	52	191	1937	The structure of electronic ...	G.H. Wannier
70	1.08	1653	130	PRL	10	518	1963	Classification of two-electron ...	J. Cooper, U. Fano, and F. Prats
72	1.06	1901	118	PR0	46	509	1934	On the constitution of ...	E. Wigner and F. Seitz
73	1.05	876	180	PR0	75	486	1949	The radiation theories of ...	F.J. Dyson
78	1.03	1995	119	PR0	109	1860	1958	Chirality invariance and ...	E. Sudarshan and R. Marshak
85	1.00	201853	3	PRB	22	5797	1980	Cluster formation in ...	H. Rosenstock and C. Marquardt
87	0.99	10168	48	PRL	6	106	1961	Population inversion and ...	A. Javan, W. Bennett, and D. Herriott
90	0.98	3231	86	PR0	79	350	1950	Antiferromagnetism ...	P.W. Anderson
92	0.97	1199	149	PR0	76	749	1949	The theory of positrons	R.P. Feynman

but its Google number 2.13×10^{-4} is only a factor 2.2 smaller than that of Cabibbo's paper! What accounts for this high Google rank? From the scientific standpoint, Slater's paper introduced the determinant form for the many-body wavefunction. This form is so ubiquitous in current literature that very few articles actually cite the original work when the Slater determinant is used. The Google PageRank algorithm identifies this hidden gem primarily because the average Google contribution of the children of S is $\langle G_j/k_j \rangle = 3.51 \times 10^{-6}$, which is a factor 2.3 larger than the contribution of the children of C. That is, the children of Slater's paper were both influential and Slater loomed as a very important father figure to his children.

The striking ability of the Google PageRank algorithm to identify influential papers can be seen when we consider the top-100 Google-ranked papers. Table 2 shows the subset of publications on the top-100 Google rank in which the ratio of Google rank to citation rank is greater than 10; that is, publications with anomalously high Google rank compared to their citation rank. This list contains many easily-recognizable papers for the average physicist. For example, the publication by Wigner and Seitz, "On the Constitution of Metallic Sodium" introduced Wigner-Seitz cells, a construction that appears in any solid-state physics text. The paper by Gell-Mann and Brueckner, "Correlation Energy of an Electron Gas at High Density" is a seminal publication in many-body theory. The publication by Glauber, "Photon Correlations", was recognized for the 2005 Nobel prize in physics. The Kramers-Wannier article, "Statistics of the Two-Dimensional Ferromagnet. Part I", showed that a phase transition occurs in two dimensions, contradicting the common belief at the time. The article by Dyson, "The Radiation Theories of Tomonaga, Schwinger, and Feynman", unified the leading formalisms for quantum electrodynamics and it is plausible that this publication would have earned Dyson the Nobel prize if it could have been shared among four individuals. One can offer similar rationalizations for the remaining articles in this table.

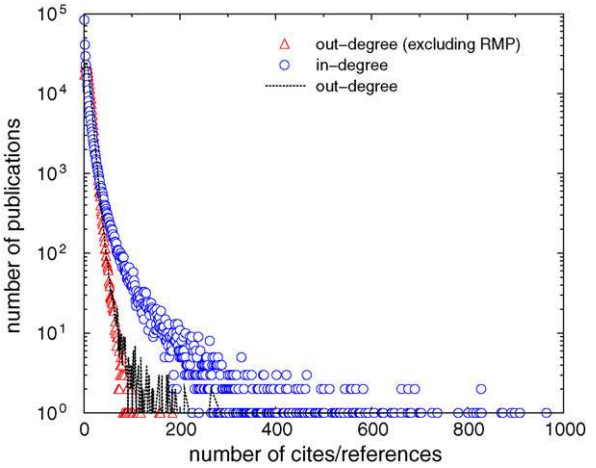

Fig. 3. The in-degree distribution (citations to) and out-degree distribution (the length of the reference list) for all Physical Review publications. The out-degree distribution is shown with and without the contribution of Reviews of Modern Physics. Eleven papers with more than 1000 citations are beyond the upper limit of the x-axis.

On the other hand, an apparent mistake is the paper by Rosenstock and Marquardt, "Cluster formation in two-dimensional random walks: Application to photolysis of silver halides" (RM). Notice that this article has only three citations! Why does RM appear among the top-100 Google-ranked publications? In RM, a model that is essentially diffusion-limited aggregation is introduced. Although these authors had stumbled upon a now-famous model, they focused on the kinetics of the system and apparently did not appreciate its wonderful geometrical features. This discovery was left to one of the children of RM—the famous paper by T. Witten and L. Sander, "Diffusion-Limited Aggregation, a Kinetic Critical Phenomenon" Phys. Rev. Lett. **47**, 1400 (1981), with 680 citations as of June 2003. Furthermore, the Witten and Sander article has only 10 references; thus a substantial fraction of its fame is exported to RM by the Google PageRank algorithm. The appearance of RM on the list of top-100 Google-ranked papers occurs precisely because of the mechanics of the Google PageRank algorithm in which being one of the few references of a famous paper makes a huge contribution to the Google number.

A natural question to ask is whether the Google rankings are robust with respect to the value of the free parameter d in the Google algorithm. As mentioned above, we believe that our *ad hoc* choice of $d = 0.5$ accounts in a reasonable way for the manner in which citations are actually made. For $d = 0.15$, as in the original Google algorithm, the Google rankings of highly-cited papers locally reorder to a considerable extent compared to the rankings for the case $d = 0.5$, but there is little global reordering. For example, all of the top-10 Google-ranked papers calculated with $d = 0.5$ remained among the top-50 Google-ranked papers for $d = 0.15$. Thus, up to this degree of variation, Google rankings are a robust measure. To further quantify it we measured the Spearman rank correlation of PageRank(d) with our choice of PageRank(0.5). Overall the correlation is very high. It varies between 0.98 and 1 for $0.1 < d < 0.9$. The rank correlation 0.91 with the number of citations is also quite high.

One can show that for $d \to 1$ Google rank almost exactly reduces to the citation rank. Indeed, in the extreme case of $d = 1$, the Google number of each node equals $1/N$. For $d \to 1$, we therefore write $d = 1 - \epsilon$, with $\epsilon \ll 1$, and also assume that there is a correspondingly small deviation of the Google numbers from $1/N$. Thus we write $G_i = \frac{1}{N} + \mathcal{O}(\epsilon)$. Substituting these into Eq. (1), we obtain

$$G_i = \epsilon \sum_j \frac{G_j}{k_j} + \frac{1-\epsilon}{N} \approx \frac{1}{N}\left[1 + \epsilon\left(\sum_j \frac{1}{k_j} - 1\right)\right] \tag{2}$$

To estimate the sum in Eq. (2), we use the fact that the out-degree distribution is relatively narrow (Fig. 3), especially if we exclude the broad tail that is caused by the contributions of review articles that appear in the Reviews of Modern Physics. While the mean in-degree and out-degrees are both close to 9 (and should be exactly equal for the complete citation network), the dispersion for the in degree is 23.15, while the dispersion for the out degree (excluding Reviews of Modern Physics) is 8.64.

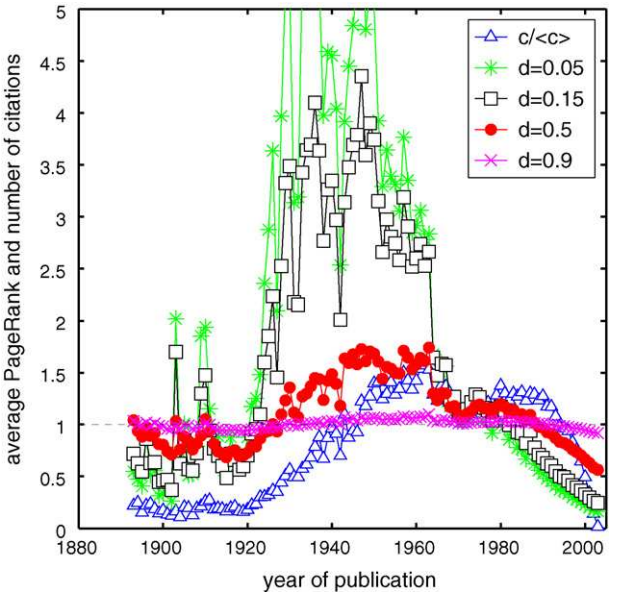

Fig. 4. Average value of PageRank vs. year of publication. Different symbols correspond to different values of d: 0.05—green stars, 0.15—black open squares, 0.5—red filled circles, 0.9—purple x's. Open blue triangles show the number citations c_i which, for proper comparison with the PageRank, is normalized by its average value ($\langle c \rangle = 8.08$). Dashed line at 1 gives the global average value of PageRank.

As a result of the sharpness of the out-degree distribution, the sum $\sum_{j \text{nn} i} \frac{1}{k_j}$ for nodes with high in-degree is approximately equal to the in-degree c_i of node i times $\langle \frac{1}{k} \rangle$. With this assumption, Eq. (2) becomes

$$G_i = \frac{1}{N}\left[1 + \epsilon\left(c_i \left\langle \frac{1}{k} \right\rangle - 1\right)\right] . \tag{3}$$

That is, the leading correction to the limiting $d = 1$ result that $G_i = \frac{1}{N}$ is proportional to the in-degree (the number of received citations) c_i of each node. Thus, as $d \to 1$, the Google rank of each node is identical to its citation rank under the approximation that we neglect the effect of the dispersion of the out-degree in the citation network.

One final aspect to consider is whether PageRank gives an "unfair" advantage to older papers. Indeed long random walks on time-directed networks inevitably drift towards older papers (Xie et al., in preparation). Since the average length of the walk in the PageRank algorithm is given by $1/d$, this effect is especially pronounced for small values of d. To investigate this question, we plot in Fig. 4 the average value of normalized PageRank as a function of the year of publication for different values of d. For comparison we also include the same plot for the normalized number of citations. As one can see from this figure, for $d = 0.5$ the age variation is not great and is comparable to that in the number of citations. However, in agreement with our qualitative understanding, the relative weight assigned to older papers (those published in 1920–1960) increases for smaller values of d. Interestingly, the oldest papers in our dataset (those published before ~1920 overall have below average PageRank as well as the number of citations.

4. Conclusions

We believe that protocols based on the Google PageRank algorithm hold a great promise for quantifying the impact of scientific publications. They provide a meaningful extension to traditionally-used importance measures, such as the number of citation of individual articles and the impact factor for journals as a whole. The PageRank algorithm implements, in an extremely simple way, the reasonable notion that citations from more important publications should contribute more to the rank of the cited paper than those from less important ones. Other ways of attributing a quality for a citation would require much more detailed contextual information about the citation itself.

The situation in citation networks is not that dissimilar from that in the World Wide Web, where hyperlinks contained in popular websites and pointing to your webpage would bring more Internet traffic to you and thus would contribute

substantially to the popularity of your own webpage. Scientists commonly discover relevant publications by simply following chains of citation links from other papers. Thus, it is reasonable to assume that the popularity or "citability" of papers may be well approximated by the random surfer model that underlies the PageRank algorithm. One meaningful difference between the WWW and citation networks is that citation links cannot be updated after publication, while WWW hyperlinks keep evolving together with the webpage containing them. Thus, scientific papers and their citations tend to age much more rapidly than active webpages. These differences could be taking into account by explicitly incorporating the effects of aging into the Page Rank algorithm (Xie et al., in preparation).

Acknowledgments

Two of us, (PC and SR) gratefully acknowledge financial support from the US National Science Foundation grant DMR0535503. SR also thanks A. Castro Neto, A. Cohen, and K. Lane for literature advice. Work at Brookhaven National Laboratory was carried out under Contract No. DE-AC02-98CH10886, Division of Material Science, U.S. Department of Energy.

References

Bollen, J., Rodriguez, M. A., & Van de Sompel, H. (2006). Journal status. *Scientometrics*, *69*(3), 669–687. This work was later put in the context of other relevant publications in: Ball, P. (2006). Prestige is factored into journal ratings. *Nature, 439*(7078), 770–771.

Bonacich, P. F. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*, 113–120.

Bonacich, P. F. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, *92*, 1170–1182.

Brin, S., & Page, L. (1988). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, *30*, 107–117.

Fortunato, S., Boguna, M., Flammini, A., & Menczer, F. How to make the top ten: Approximating PageRank from in-degree. Cs.IR/0511016.

Fortunato, S., Flammini, A., & Menczer, F. (2006). Scale-free network growth by ranking. *Physics Review Letters*, *96*, 218701.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, *46*, 604–632.

Laherrère, J., & Sornette, D. (1998). Stretched exponential distributions in nature and economy: "Fat tails" with characteristic scales. *European Physical Journal B*, *2*, 525–539.

Price, D. J. de Solla. (1965). Networks of Scientific Papers. *Science*, *149*, 510–515.

Price, D. J. de Solla. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*, 292–306.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, *4*, 131–134.

Redner, S. (2005). Citation statistics from more than a century of physical review. *Physics Today*, *58*, 49. See http://arxiv.org/abs/physics/0407137

Xie, H., Walker, D., Yan, K.-K, Maslov, S., & Redner, S. (in preparation). Ranking scientific publications using a simple model of network traffic.