Proceedings of the 4th International Conference on Applied Physics and Mathematical Modeling
DOI: 10.54254/2753-8818/2026.36716

Eigenvector-Based Web Page Ranking with PageRank and Random Walk Models

Shengshu Huang

High School Attached to Northeast Normal University, Changchun, China
catsaki@outlook.com

Abstract. Web page ranking converts the hyperlink structure of the Web into numerical importance scores to help retrieve information. This review examines eigenvector and random walk models, with an emphasis on PageRank, Hyperlink-Induced Topic Search (HITS), Stochastic Approach for Link-Structure Analysis (SALSA), personalised PageRank, and topic-sensitive variations. Directed web graphs, sparse adjacency matrices, Markov chains, damping factors and stationary distributions are the mathematical foundations for link analysis, as shown above. Representative studies are compared based on data source, graph scale, convergence behaviour, personalisation, temporal update, fairness and ranking quality. Based on the above studies, the classical model is still popular due to its explicit assumptions and efficient computation via sparse linear algebra; newer methods have added content semantics, user context and dynamic link updates. The above shortcomings include the assumption that hyperlinks are always endorsements, sensitivity to link manipulation, high computational cost at the web scale, stale rankings in dynamic environments, and limited personalisation of global models. Future work will introduce scalable iterative solvers, temporal models, robust link analysis, contextual personalisation and interpretable hybrid methods to combine topology with content and user behaviour.

Keywords: PageRank, HITS, link analysis, random walk, web graph

1. Introduction

A general way to organise the vast number of web pages that search engines need to help users find is by determining how important each page is. With the development of the World Wide Web, simple keyword matching alone became insufficient, and thus link-based authority evaluation became increasingly important. A web can be modelled as a directed graph where pages are nodes and hyperlinks are edges. PageRank formally introduced the concept of links from high-authority pages as stronger indicators of authority [1]. Transforming the topology of links into computable scores is an application of mathematics in information retrieval.

Research soon expanded from PageRank to Hyperlink-Induced Topic Search (HITS), which models hubs and authorities [2], the Stochastic Approach for Link-Structure Analysis (SALSA) [3], topic-sensitive ranking [4], and personalised web search [5]. Recently, scholars have addressed the fairness problem of PageRank through link recommendation [6], strengthened the evaluation of graph-based ranking [7], developed sublinear local centrality estimation [8], accelerated

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.36716

personalised PageRank computation [9], and improved ranking maintenance in dynamic graphs [10]. Link-based signals have also been combined with topical features in focused crawling [11] and with semantic or collaboration features in scholarly recommendation and graph-based text summarisation [12, 13].

This paper introduces an organised introduction to eigenvector and random-walk-based ranking. Section 2 introduces the web graph and Markov chain foundations formally. Section 3 presents the primary algorithms of PageRank, HITS, SALSA and their customised modifications, reviews recent research achievements, and finally discusses limitations and future research directions. The conclusion presents the main results and extended applications of the mathematical model.

2. Basic concepts of web graphs, link analysis and Markov chains

This section briefly introduces the general ideas of web graphs, link analysis and Markov chains. A directed graph can be used to model the World Wide Web, and its nodes are web pages and directed edges are hyperlinks. A sparse adjacency matrix can represent this structure and is suitable for link-based computations. Link analysis finds the relative strength of pages in the web graph. A hyperlink connecting one page to another can be regarded as a structural feature of high priority, but it does not necessarily indicate approval. The Random-Surfer model of PageRank can be expressed as a Markov chain. At each step, a random surfer will either follow an outgoing link or jump to another page, and the damping factor controls the probability of following outgoing links. After multiple applications of this transition rule, the distribution will be close to stable under normal conditions. The weight given to each page in that distribution is its PageRank score. The first two problems of the model are how to handle dangling pages without outgoing links and how to make the random-surfer process uniquely stationary. Teleportation can avoid the problem of being stuck in disconnected or periodic regions of the graph. Most PageRank implementations have a damping factor of approximately 0.85. Therefore, it gives more weight to hyperlink following and assigns a relatively small proportion to random jumps to stabilise the long-run ranking [1].

Recently, research on web graphs has begun to pay more attention to sparse representations, local estimation and dynamic updates, and ranking has been applied to large, continuously changing networks [8, 10].

3. Eigenvector and random-walk methods for web page ranking

3.1. The pagerank algorithm

PageRank is based on the stationary distribution of a random-surfer process. Given the initial distribution of scores, the algorithm repeatedly spreads importance via outgoing hyperlinks and uses teleportation until the scores converge [1].

With dangling pages handled appropriately, the teleportation component allows all pages to remain reachable and supports a single stable ranking even if the web graph is disconnected [1].

PageRank is attractive because of its random-walk foundation and the exploitation of sparse graph structure in computation. Its global authority scores, however, may favour widely popular pages that are not related to a particular query, and the resulting distribution can replicate the imbalance already present in the link graph [4, 6].

DOI: 10.54254/2753-8818/2026.36716

3.2. HITS and SALSA

HITS assigns separate authority and hub scores based on mutual reinforcement; that is, good authorities are connected to good hubs, and good hubs connect to good authorities [2]. SALSA uses the above distinction and stochastic walks on a bipartite representation of the link graph [3]. Since both operate in the query-focused subgraph, their stability and applicability depend significantly on the construction of this subgraph.

3.3. Personalised and topic-sensitive variants

Personalised PageRank modifies the teleportation distribution to have random jumps favour a specific set of pages for different users [5]. Topic-Sensitive PageRank precomputes the rankings of topic categories and combines them at query time [4]. Recently, work has accelerated high-precision personalised PageRank computation [9] and proposed graph interventions to improve group-level PageRank fairness [6].

3.4. Evaluation metrics

Metrics of Algorithm Performance are listed below. Computational efficiency includes convergence speed, memory consumption and the number of graph elements accessed; recently, local estimation and acceleration methods have reduced these costs directly [8, 9]. Ranking quality is generally evaluated by the precision at a certain cutoff, mean average precision, and normalised discounted cumulative gain. Ranking stability measures the sensitivity of rankings to small graph changes, and credible evaluation compares multiple benchmarks and relevance judgments instead of using a single proxy [7].

3.5. State-of-the-art studies on pagerank and link-based ranking

Research at present differs in terms of data sources, graph sizes and applications. Web-graph studies emphasise scalability, freshness, and fairness [6, 8-11]. Other areas have also applied PageRank-type centrality to scholarly-paper ranking and recommendation [7, 12], and related random-walk ranking can combine graph structure with semantic representations in text summarisation [13]. A typical achievement is time-aware ranking for evolving graphs. Recently, efficient methods for updating PageRank under edge insertion and deletion have been developed [10]. Another direction combines structure and semantics; focused crawlers adjust hyperlink priorities based on topical evidence [11]. Recently, scholars have studied damping sensitivity, convergence, robustness, and the cost of local-versus-global computation in their research. Benchmark studies show that the results differ depending on the selected ground truth or proxy [7]. Local centrality estimators and accelerated personalised PageRank methods reduce the computational cost by not requiring a full-scale global ranking for every query [8, 9], and fairness-aware link interventions investigate the utility cost of altering the score distribution [6]. The computational pipeline is still conceptually the same: build the graph representation, handle dangling nodes, define a stochastic transition process with teleportation, and repeat until the ranking converges. Sparse linear algebra, local estimators and optimised iterative solvers have made this pipeline feasible for large graphs [8, 9].

DOI: 10.54254/2753-8818/2026.36716

3.6. Limitations and prospects

3.6.1. Limitations

Existing ways have some defects. Frequently, the assumption that a hyperlink represents an endorsement is violated by navigational, sponsored or spam links. Dangling nodes and broken links require explicit handling. Given the large scale and sparsity of web graphs, storage and computation costs are high; thus, local, accelerated, and dynamic methods have been proposed to alleviate this problem [8-10]. Static rankings will be outdated after changes to pages and links, and a single global score may not meet all the different needs of users. Fairness is also a problem because structural imbalances in the graph may be reflected in the resulting ranking [6].

3.6.2. Future directions

Five Directions for Future Research are connected. First, session-aware personalisation should update rankings based on the current context, not fixed profiles. Second, temporal models should include link freshness and page-update rates [10]. Third, scalable local and accelerated solvers should reduce latency while maintaining convergence guarantees [8, 9]. Fourth, hybrid approaches should combine hyperlink topology with text, user behaviour and other contextual cues [5, 11-13]. Ranking systems need to improve their immunity to manipulation and ensure fairness and interpretability in web communities [6].

4. Conclusion

Eigenvector and random-walk models are mathematically intuitive ways to rank pages in large directed networks. PageRank is based on the stationary distribution of a random-surfer process, HITS divides authority and hub roles, and SALSA combines mutual reinforcement with stochastic transitions. Personalised and topic-sensitive variants modify the teleportation or transition mechanism to show rankings based on user interests, query context or preset topics. All these methods have their own advantages in terms of construction of the web graph, handling of dangling nodes and disconnected components, selection of damping factor, and which relevance and stability measures are used. Classical link analysis is still convenient because it can use sparse matrix operations and iterative solvers, but its assumptions are not ideal. Hyperlinks may be navigation, advertising or manipulation, not endorsement; static scores will become outdated as pages and links change. The global authority score cannot capture diverse user intentions. Therefore, recent work has moved towards incremental updates, fairness-aware link analysis, semantic feature integration and contextual personalisation. The extensions should be judged based on clearly stated graph scales, update frequencies, computing budgets and relevance judgments, rather than by standalone accuracy claims. Future systems will probably integrate link topology with textual, behavioural, temporal and multimodal information, and keep interpretable components that explain why a page has a high score. In short, the long-term value of the eigenvector and random walk models is that they can convert complex network structure into reproducible ranking rules. Their future applications will depend on scalable computing, transparent evaluation, and careful optimisation for the diverse and constantly changing environments of the present-day Web.

DOI: 10.54254/2753-8818/2026.36716

References

[1] Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems, 30*(1–7), 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X

[2] Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM, 46*(5), 604–632. https://doi.org/10.1145/324133.324140

[3] Lempel, R., & Moran, S. (2001). SALSA: The stochastic approach for link-structure analysis. *ACM Transactions on Information Systems, 19*(2), 131–160. https://doi.org/10.1145/382979.383041

[4] Haveliwala, T. H. (2003). Topic-sensitive PageRank: A context-sensitive ranking algorithm for Web search. *IEEE Transactions on Knowledge and Data Engineering, 15*(4), 784–796. https://doi.org/10.1109/TKDE.2003.1208999

[5] Jeh, G., & Widom, J. (2003). Scaling personalized web search. In *Proceedings of the 12th International Conference on World Wide Web* (pp. 271–279). ACM. https://doi.org/10.1145/775152.775191

[6] Tsioutsiouliklis, S., Pitoura, E., Semertzidis, K., & Tsaparas, P. (2022). Link recommendations for PageRank fairness. In *The ACM Web Conference 2022* (pp. 3541–3551). ACM. https://doi.org/10.1145/3485447.3512249

[7] Zhang, Y., Wang, M., Saberi, M., & Chang, E. (2022). Analysing academic paper ranking algorithms using test data and benchmarks: An investigation. *Scientometrics, 127*, 4045–4074. https://doi.org/10.1007/s11192-022-04429-z

[8] Bressan, M., Peserico, E., & Pretto, L. (2023). Sublinear algorithms for local graph-centrality estimation. *SIAM Journal on Computing, 52*(4), 968–1008. https://doi.org/10.1137/19M1266976

[9] Chen, Z., Guo, X., Zhou, B., Yang, D., & Skiena, S. (2023). Accelerating personalized PageRank vector computation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 262–273). ACM. https://doi.org/10.1145/3580305.3599251

[10] Pashikanti, R. P., & Kundu, S. (2022). FPPR: Fast pessimistic (dynamic) PageRank to update PageRank in evolving directed graphs on network changes. *Social Network Analysis and Mining, 12*(1), Article 141. https://doi.org/10.1007/s13278-022-00968-8

[11] Liu, J., Yang, Z., Yan, X., & Chen, D. (2024). Applying particle swarm optimization-based dynamic adaptive hyperlink evaluation to focused crawler for meteorological disasters. *Complex & Intelligent Systems, 10*, 233–255. https://doi.org/10.1007/s40747-023-01121-4

[12] Kanwal, T., & Amjad, T. (2024). Research paper recommendation system based on multiple features from citation network. *Scientometrics, 129*, 5493–5531. https://doi.org/10.1007/s11192-024-05109-w

[13] Alselwi, G., & Taşcı, T. (2024). Extractive Arabic text summarization using PageRank and word embedding. *Arabian Journal for Science and Engineering, 49*, 13115–13130. https://doi.org/10.1007/s13369-024-08890-1