

Social Networks, Incentives, and Search

Jon Kleinberg*
Dept. of Computer Science
Cornell University, Ithaca NY
kleinber@cs.cornell.edu

Categories and Subject Descriptors: F.2.2 Analysis of Algorithms and Problem Complexity: Non-numerical Algorithms and Problems

General Terms: Algorithms, Theory.

Keywords: social networks, decentralized search, game-theoretic analysis

1. ABSTRACT

The role of network structure has grown in significance over the past ten years in the field of information retrieval, stimulated to a great extent by the importance of link analysis in the development of Web search techniques [4]. This body of work has focused primarily on the network that is most clearly visible on the Web: the network of hyperlinks connecting documents to documents. But the Web has always contained a second network, less explicit but equally important, and this is the social network on its users, with latent person-to-person links encoding a variety of relationships including friendship, information exchange, and influence. Developments over the past few years — including the emergence of social networking systems and rich social media, as well as the availability of large-scale e-mail and instant messaging datasets — have highlighted the crucial role played by on-line social networks, and at the same time have made them much easier to uncover and analyze. There is now a considerable opportunity to exploit the information content inherent in these networks, and this prospect raises a number of interesting research challenges

Within this context, we focus on some recent efforts to formalize the problem of searching a social network. The goal is to capture the issues underlying a variety of related scenarios: a member of a social networking system such as MySpace seeks a piece of information that may be held by a friend of a friend [27, 28]; an employee in a large company

*Supported in part by a John D. and Catherine T. MacArthur Foundation Fellowship, a Google Research Grant, and NSF grants CCF-0325453, IIS-0329064, CNS-0403340, and BCS-0537606.

searches his or her network of colleagues for expertise in a particular subject [9]; a node in a decentralized peer-to-peer file-sharing system queries for a file that is likely to be a small number of hops away [2, 6, 16, 17]; or a user in a distributed IR or federated search setting traverses a network of distributed resources connected by links that may not just be informational but also economic or contractual [3, 5, 7, 8, 13, 18, 21]. In their most basic forms, these scenarios have some essential features in common: a node in a network, without global knowledge, must find a short path to a desired “target” node (or to one of several possible target nodes).

To frame the underlying problem, we go back to one of the most well-known pieces of empirical social network analysis — Stanley Milgram’s research into the small-world phenomenon, also known as the “six degrees of separation” [19, 24, 25]. The form of Milgram’s experiments, in which randomly chosen starters had to forward a letter to a designated target individual, established not just that short chains connecting far-flung pairs of people are abundant in large social networks, but also that the individuals in these networks, operating with purely local information about their own friends and acquaintances, are able to actually find these chains [10]. The Milgram experiments thus constituted perhaps the earliest indication that large-scale social networks are structured to support this type of decentralized search. Within a family of random-graph models proposed by Watts and Strogatz [26], we have shown that the ability of a network to support this type of decentralized search depends in subtle ways on how its “long-range” connections are correlated with the underlying spatial or organizational structure in which it is embedded [10, 11]. Recent studies using data on communication within organizations [1] and the friendships within large on-line communities [15] have established the striking fact that real social networks closely match some of the structural features predicted by these mathematical models.

If one looks further at the on-line settings that provide the initial motivation for these issues, there is clearly interest from many directions in their long-term economic implications — essentially, the consequences that follow from viewing distributed information retrieval applications, peer-to-peer systems, or social-networking sites as providing marketplaces for information and services. How does the problem of decentralized search in a network change when the participants are not simply agents following a fixed algorithm, but strategic actors who make decisions in their own self-interest, and may demand compensation for taking part

in a protocol? Such considerations bring us into the realm of algorithmic game theory, an active area of current research that uses game-theoretic notions to quantify the performance of systems in which the participants follow their own self-interest [20, 23]. In a simple model for decentralized search in the presence of incentives, we find that performance depends crucially on both the rarity of the information and the richness of the network topology [12] — if the network is too structurally impoverished, an enormous investment may be required to produce a path from a query to an answer.

2. REFERENCES

- [1] L. Adamic, E. Adar. How to search a social network. *Social Networks*, 27(3):187-203, July 2005.
- [2] J. Aspnes, G. Shah. Distributed data structures for P2P systems. in *Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless and Peer-to-Peer Networks* (Jie Wu, ed.), CRC Press, 2005.
- [3] J. Callan. Distributed information retrieval. In W.B. Croft, editor, *Advances in information retrieval*, chapter 5, pages 127-150. Kluwer Academic Publishers, 2000.
- [4] S. Chakrabarti, *Mining the Web: Discovering Knowledge from Hypertext Data*, Morgan Kaufmann, 2002.
- [5] N. Craswell. Methods for distributed information retrieval. Ph. D. thesis, The Australian Nation University.
- [6] A. Crespo, H. Garcia-Molina. Routing indices for peer-to-peer systems. *Proc. of the International Conference on Distributed Computing Systems (ICDCS)*, July 2002.
- [7] E. A. Fox, M. Goncalves, M. Luo, Y. Chen, A. Krowne, B. Zhang, K. McDevitt, M. Perez-Quinones, R. Richardson, L. Cassel. Harvesting: Broadening the Field of Distributed Information Retrieval. *SIGIR 2003 Workshop on Distributed Information Retrieval*. J. Callan, F. Crestani, M. Sanderson (eds.).
- [8] L. Gravano, P. Ipeirotis, and M. Sahami, QProber: A System for Automatic Classification of Hidden-Web Databases, *ACM Transactions on Information Systems*, vol. 21, no. 1, Jan. 2003.
- [9] H. Kautz, B. Selman and M. Shah. ReferralWeb: Combining Social Networks and Collaborative Filtering. *Communications of the ACM*, 1997.
- [10] J. Kleinberg. The small-world phenomenon: An algorithmic perspective. *Proc. 32nd ACM Symposium on Theory of Computing*, 2000.
- [11] J. Kleinberg. Complex Networks and Decentralized Search Algorithms. *Proceedings of the International Congress of Mathematicians (ICM)*, 2006.
- [12] J. Kleinberg, P. Raghavan. Query Incentive Networks. *Proc. IEEE Symposium on Foundations of Computer Science*, 2005.
- [13] C. Lagoze, H. Van de Sompel. The open archives initiative: building a low-barrier interoperability framework. *Proc. ACM/IEEE Joint Conference on Digital Libraries*, 2001.
- [14] C. Li, B. Yu and K. Sycara. An Incentive Mechanism for Message Relaying in Peer-to-Peer Discovery. *2nd Workshop on Economics of Peer-to-peer systems*, 2004.
- [15] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins. Geographic routing in social networks. *Proc. Natl. Acad. Sci. USA*, 102(Aug 2005).
- [16] J. Lu, J. Callan. Federated search of text-based digital libraries in hierarchical peer-to-peer networks. *Proc. 27th European Conference on Information Retrieval Research (ECIR)*, 2005.
- [17] E-K Lua, J. Crowcroft, M. Pias, R. Sharma and S. Lim. A Survey and Comparison of Peer-to-Peer Overlay Network Schemes, *IEEE Communications Surveys and Tutorials*, 7(2005).
- [18] W. Meng, C.T. Yu and K.L. Liu. (2002). Building efficient and effective metasearch engines. *ACM Comput. Surv.* 34(1).
- [19] S. Milgram, The small world problem. *Psychology Today* 1(1967).
- [20] C. H. Papadimitriou. Algorithms, Games, and the Internet. *Proc. 33rd ACM Symposium on Theory of Computing*, 2001.
- [21] L. Si, J. Callan. Modeling search engine effectiveness for federated search. *Proc. 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2005.
- [22] O. Simsek and D. Jensen. Decentralized search in networks using homophily and degree disparity. *Proc. 19th International Joint Conference on Artificial Intelligence*, 2005.
- [23] É. Tardos. Network Games. *Proc. 36th ACM Symposium on Theory of Computing*, 2004.
- [24] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry* 32(1969).
- [25] Duncan J. Watts. Six Degrees: The Science of a Connected Age, W. W. Norton, 2003.
- [26] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature* 393(1998).
- [27] B. Yu and M. P. Singh. Searching Social Networks. *Proc. 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems*, 2003.
- [28] J. Zhang and M. Van Alstyne. SWIM: fostering social network based information search. *Proc. ACM SIGCHI Conf. on Human Factors in Computing Systems*. 2004.