Referral Web: Combining Social Networks and Collaborative Filtering.

Article *in* Communications of the ACM · March 1997

DOI: 10.1145/245108.245123 · Source: DBLP

3 authors, including:

Henry Kautz
University of Rochester
123 PUBLICATIONS **10,646** CITATIONS
SEE PROFILE

Bart Selman
Cornell University
277 PUBLICATIONS **15,935** CITATIONS
SEE PROFILE

This paper appears in the *Communications of the ACM*,
vol. 40 no. 3, March, 1997.

ReferralWeb: Combining Social Networks and Collaborative Filtering

Henry Kautz, Bart Selman, and Mehul Shah
AT&T Laboratories
600 Mountain Avenue, Room 2a-402
Murray Hill, NJ 07974
(908) 582-2815
{kautz, selman}@research.att.com
shah@mit.edu

1. The Power of Referral Chaining

Numerous studies have shown that one of the most the most effective channels for dissemination of information and expertise within an organization is its informal network of collaborators, colleagues, and friends (Granovetter 1973; Kraut 1990; Wasserman and Galaskiewicz 1994). Indeed, the social network[1] is as least as important as the official organizational structure for tasks ranging from immediate, local problem solving (for example, fixing a piece of equipment) to primary work functions, such as creating project teams. Part of the success of social networks can be attributed to the "six degrees of separation" phenomena, which means that the distance between any two individuals in terms of direct personal relationships is relatively small. An equally important factor is that there are limits as to the amount and kinds of information that a person is able or willing to make available to the public at large. For example, a expert in a particular field is almost certainly *unable* to write down all he knows about the topic, and is likely to be *unwilling* to make letters of recommendation he has written for various people publicly available. Searching for a piece of information in this situation thus becomes a matter of searching the social network for an expert on the topic together with a *chain* of personal referrals from the searcher to the expert. The *referral chain* serves two key functions: it provides a reason for the expert to agree to respond to the requester by explicating their relationship (*e.g.*, they have a mutual collaborator), and it provides a criteria for the searcher to use in evaluating the trustworthiness of the expert.

Nonetheless, manually searching for a referral chain can be a frustrating and

[1] Note that by "social network" we explicitly include groups of people linked by professional activities; we do not use "social" in the more limited sense of "recreational".

time consuming task. One is faced is with the trade-off of contacting a large number of individuals at each step, and thus straining both the time and good will of the possible respondents, or of contacting a smaller, more focused set, and thus being more likely to fail to locate an appropriate expert. In response to these problems we are building ReferralWeb, an interactive system for reconstructing, visualizing, and searching social networks on the World Wide Web. Simulation experiments we ran before we began construction of ReferralWeb showed that automatically generated referrals can be highly successful in locating experts in a large network (Kautz, Selman, and Milewski 1996). The higher degree of responsiveness of an automated system can be effectively traded-off for the (potentially) lower accuracy of individual automatic referrals over those generated by a person.

2. Reconstructing and Querying Social Networks

A social network is modeled by a graph, where the nodes represent individuals, and an edge between nodes indicates that a direct relationship between the individuals has been discovered. There are many possible sources for determining direct relationships. At one extreme, which we reject as too burdensome, users could be required to enter lists of close colleagues. Analysis of email logs provides a rich source of relationships, as shown by Schwartz and Wood (1993). In fact, the initial version of our system derived its network by analyzing mail archives (Kautz, Selman, and Milewski, 1996). However, the use of such information raises concerns of privacy and security that are hard to allay. The current ReferralWeb system uses the co-occurrence of names in close proximity in any documents publicly available on the World Wide Web as evidence of a direct relationship. Such sources include:

- Links found on home pages;
- Lists of co-authors in technical papers and citations of papers;
- Exchanges between individuals recorded in netnews archives; and
- Organization charts (*e.g.*, for university departments).

The network model is constructed incrementally. When a user first registers with the system, it uses a general search engine to retrieve Web documents that mention him or her. The names of other individuals are extracted from the documents; this can be done with high accuracy (better than 90%), using techniques such as those described in (Sundheim and Grishman, 1995). This process is applied recursively for one or two levels, and the result merged into the global network model.

The network is then used to guide the search for people or documents in response to user queries. Most simply, a person may ask to find the chain between himself and a named individual (e.g., "What is my relationship to Marvin

Minsky?"). To search for an expert, the user may specify both a topic and an effective social radius: for example, a query might be, "What colleagues of mine, or colleagues of colleagues of mine, know about simulated annealing?" Another kind of query takes advantage of a designated, known expert to control the search. For instance, one might ask to "List documents on the topic 'annealing' by people close to Scott Kirkpatrick."

It is important to emphasize that ReferralWeb does not *replace* generic search engines such as AltaVista, but instead uses the social network to make their search more focused and effective. For example, in the last example above, the requirement of proximity to (the computer scientist) Kirkpatrick helps disambiguate the query between the computer science use of the term "annealing" and the use of the same term in metallurgy. The social network also prioritizes the answers, in that the user retrieves hits on people that are closest to him or herself, rather than simply a long list of hundreds of names or documents.

ReferralWeb has a number of features that sets it apart from many other collaborative filtering and recommender projects:

1. ReferralWeb attempts to uncover *existing* social networks, rather than providing a tool for creating new communities. While new communities may be appropriate for recreational uses of the Web, we emphasize helping individuals make more effective use of their large, existing networks of professional colleagues.

2. While recommender systems are often designed to provide *anonymous* recommendations, ReferralWeb is based on providing referrals via chains of *named* individuals. This is critical, because not all sources of information are equally desirable. For example, as we will describe below, a user may choose to search for referrals to people who are closely associated with some other known, trusted expert.

3. Some recommender systems require the user to manually enter a personal profile of interests, preferences, or expertise. Recommendations are generated by matching profiles that exist within the system. By contrast, ReferralWeb primarily builds its model of its users' social network by data mining public documents found on the World Wide Web. This model includes many more individuals than those who explicitly register with the service.

4. Users of ReferralWeb are not limited to any set of topic areas determined in advance. ReferralWeb uses a general full Web indexing engine (currently, AltaVista) to match individuals to topic areas.

The ReferralWeb system is part of our effort to develop agent-based programs that address practical communication needs (Kautz, Selman, and Coen, 1994). We have developed several internal prototypes of the system, and are currently building a version that will be made accessible externally on the Web.

Finally, the ReferralWeb system expands users awareness of their existing communities. Typically a user is only aware of a portion of the social network to which he or she belongs. By instantiating the larger community, the user can discover connections to people and information that would otherwise lay hidden over the horizon.

References

Granovetter, M. Strength of Weak Ties. *American Journal of Sociology* 78 (1973), 1360–1380.

Kautz, H., Selman, B., and Coen, M. Bottom-up Design of Software Agents. *Communications of the ACM*, 37,7 (1994) 143–146.

Kautz, H., Selman, B., and Milewski, A. Agent Amplified Communication. *Proceedings of AAAI-96* (Portland, Oreg.). MIT Press, Cambridge, 1996, 3–9.

Galegher, J., Kraut, R., and Edigo C. *Intellectual Teamwork: Social and Technological Bases for Cooperative Work.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1990.

Schwartz, M. F. and Wood, D. C. M. Discovering shared interests using graph analysis. *Comm. ACM* 36, 8 (1993), 78–89.

Sundheim, B. and Grishman, R., Eds. *Proceedings of the Sixth Message Understanding Conference (MUC-6).* Morgan Kaufmann, San Francisco, Calf., 1995.

Wasserman, S. and Galaskiewicz, J., Eds. *Advances in Social Network Analysis.* Sage Publications, Thousand Oaks, Calf., 1994.