Social Search

Marc Najork

Microsoft Research
1065 La Avenida, Mountain View, CA 94043, USA

Abstract. "Social Search" refers to two aspects of the integration of web search with social networks: how queries to a search engine may surface (socially) relevant content from social networks, and how signals from social networks may influence the (personalized) ranking of search results. The first part of the talk surveys the integration of Bing with Facebook, Twitter, Quora, Foursquare, LinkedIn, Klout, and other social platforms. The second part focuses on two technical details of this integration: a measure for quantifying the "affinity" between two users of a social network and an efficient algorithm for computing that measure, and a method for efficiently surfacing pages "liked" by your friends from a document-sharded index. The final part discusses limitations of social search, such as skewed demographics and weak homophily.

Keywords: Web search, social networks, social search

1 Introduction

In this talk, I discuss the integration of web search with social networks. "Social Search" refers to two distinct aspects of this integration: how a user's queries to a search engine may surface content from social networks (possibly authored by the searcher's connections on that network), and how signals from social networks (say, the fact that a friend "liked" a web page) may influence the personalized ranking of algorithmic search results.

The talk is divided into three parts. In the first part, I survey the integration of Microsoft's Bing Search engine with various social platforms, including Facebook, Twitter, Quora, Foursquare, LinkedIn, Klout, and others. Bing surfaces relevant content from multiple social networks, whether it is public or authored by the searcher's connections on each network. It also promotes algorithmic search results that were endorsed by the user's friends, or that are trending social media platforms.

Bing pays particular attention to "people search", queries meant to retrieve relevant information about a person. Celebrity search is a well-studied problem, and retrieval precision is high, but this is less true for non-celebrity people search, due to the ambiguity of common names. Bing uses separation in social networks to surface individuals in the searcher's extended social circle. Furthermore, it allows registered users to claim content related to them, thereby making it possible to cluster people search results by individual. Finally, it shows summaries

of LinkedIn profiles directly on the results page. For celebrity searches, it will similarly show their presence in social media prominently on the results page together with related postings.

The second portion of the talk focuses on two technical details of Social Search. First, I describe an "affinity" measure [3] for quantifying how robustly connected two nodes in a graph (or two users of a social networks) are, or more precisely, what fraction of the graph's edges can be deleted before the nodes become disconnected. The affinity measure can be efficiently estimated by a randomized, sketch-based algorithm. The off-line phase of that algorithm computes a fixed-size sketch for each node of the graph, capturing a representative of its connected component at various levels of edge deletion. The online phase consists of retrieving the sketches of two nodes and performing a pointwise comparison on them to compute the affinity. The space complexity of the algorithm is $O(n)$, the time complexity of the off-line phase is $O(\alpha(n))$ (the complexity of union-find with path compression), while the time complexity of the online phase is $O(1)$.

Second, I discuss an approach for efficiently retrieving web pages "liked" by a user's friends. While seemingly trivial, it is challenging to integrate this functionality into a document-sharded distributed search index [2]. In such a setting, queries are distributed from a front-end to many index servers (each holding a part of the index), and results are sent back. Because of network constraints, both query and result transfers should be small; in particular, it is neither feasible to send the full set of the searcher's friends down the distribution tree, nor to send the full set of results up the aggregation tree. Moreover, social graphs can be very large and change continuously, making it impractical to maintain a copy of the graph on each index server.

The final part of the talk confesses to some of the limitations of Social Search; namely, that many social networks have skewed demographics [4] (in terms of gender, race, age, education and income), making it dangerous to generalize trends in networks to the overall population; and that while a user's actions on social networks *is* predictive of their proclivities [1], it is not clear that these preferences transfer to their "virtual" friends. Finally, it is challenging to "separate the wheat from the chafe" – to identify salient posts in a sea of the mundane.

References

1. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, pp. 5802–5805. National Academy of Sciences (2013).
2. Najork, M.A., Panigrahy, R., Shenoy, R.K.: Considering document endorsements when processing queries. US Patent App. 13/218,450 (filed 2011)
3. Panigrahy, R., Najork, M., Xie, Y.: How user behavior is related to social affinity. In: 5th ACM International Conference on Web Search and Data Mining, pp. 713–722. ACM, New York (2012)
4. Rainie, L., Brenner, J. , Purcell, K.: Photos and videos as social currency online. Pew Research Center (2012).