

The Role of Network Distance in LinkedIn People Search

Shih-Wen Huang

Computer Science & Engineering
University of Washington

wenhuang@cs.washington.edu

Daniel Tunkelang

Query Understanding
LinkedIn

dtunkelang@linkedin.com

Karrie Karahalios

Computer Science
University of Illinois

kkarahal@cs.uiuc.edu

ABSTRACT

LinkedIn is the world's largest professional network, with over 300 million members. One of the primary activities on the site is people search, for which LinkedIn members are both the users and the corpus. This paper presents insights about people search behavior on LinkedIn, based on a log analysis and a user study. In particular, it examines the role that network distance plays in name searches and non-name searches. For name searches, users primarily click on only one of the results, and closer network distance leads to higher click-through rates. In contrast, for non-name searches, users are more likely to click on multiple results that are not in their existing connections, but with whom they have shared connections. The results show that, while network distance contributes significantly to LinkedIn search engagement in general, its role varies dramatically depending on the type of search query.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – search process, selection process

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Social network, exploratory search, people search

1. INTRODUCTION

People search is one of the most popular types of online search. It is estimated that about 11% to 17% of web queries contain a person's name. [5] Therefore, understanding how people search for people is a critical issue in information retrieval.

People search has been studied in several different contexts. Weerkamp et al. [5] analyzed the query log of a people search engine on the web. They observed many differences between people search and web document search (e.g., higher percentage of one-query sessions and lower click-through rates). Also, Guy et al. [3] presented a study of the people search application (Faces) at IBM. They showed that the users behaved quite differently from the users who performed people search on the web. These results suggest that people searches vary based on their contexts,

and more studies are necessary to better understand and support various online people search.

The online professional network is another context where people search is used extensively. For instance, LinkedIn, which has established itself as the largest professional network, has over 300 million members [7]. LinkedIn has reported that its members performed over 5.7 billion searches on the platform in 2012 [7]. The emergence of LinkedIn and other professional networks has given rise to a new class of search problems. Specifically, the graph structure of professional networks allows the users search for people with different network distances, which makes it essentially different from people search on the web or in an enterprise.

The people search systems in online social networks (e.g., LinkedIn and Facebook) mostly focus on supporting name search. At LinkedIn, there is a very long tail of search queries, and the largest class of queries is name queries [8]. In addition, previous research [4] suggests it is difficult to deliberately set out to discover specific information from Facebook's search and browsing features — especially in the case where one searches for an unknown person or expert. However, upon analysis of the search log of LinkedIn people search, we found that when users enter certain type of queries (e.g., Job Title), their intentions and consequent actions are very different from those of name searches. This suggests that the type of search performed on LinkedIn is query dependent, and we need to rethink how to better support these different types of search.

The rest of this paper focuses on understanding the differences between how LinkedIn searchers behave in name searches and non-name searches. We first present a log analysis of LinkedIn people search that characterizes click behavior for different query tags. We find that there are significant differences between the click behaviors of name queries and non-name queries. Moreover, we find that network distance plays a key role in influencing click behavior, but that it acts differently for name and non-name searches. We then present the results of a follow-up user study that we conducted to better understand the intentions motivating searchers to select a given search result. We conclude with the discussion of the design implications of our research for improving people search in the context of a professional network like LinkedIn.

2. LOG ANALYSIS

We analyzed the search logs of LinkedIn people searches conducted from 6/19/2013 to 7/18/2013. The search log contains millions of unique people searches. We automatically tagged each query using a hidden Markov model that segmented the keywords and tagged each segment as an entity type (e.g., First Name and Company Name). For proprietary reasons, we could not discuss the details of the tagger, but the internal human evaluation shows

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGIR '14, July 6–11, 2014, Gold Coast, Queensland, Australia.
Copyright © 2014 ACM 978-1-4503-2257-7/14/07...\$15.00.

Category	Query Tag	Percentage of multiple-click search	CTR on 1 st -degree	CTR on 2 nd -degree	CTR on 3 rd -degree or more	Connect rate after click on 2 nd -degree	Connect rate after click on 3 rd -degree or more
Non-name	Job Title	13.92	1 (normalizer)	2.08	1.57	2.68	0.36
Non-name	Skill	13.69	1 (normalizer)	1.68	1.37	2.40	0.27
Non-name	Company Name	8.01	1 (normalizer)	0.99	0.90	1.28	0.21
Name	Last Name	5.28	1 (normalizer)	0.75	0.51	1.41	0.50
Name	First Name	2.98	1 (normalizer)	0.49	0.43	1.21	0.33
Name	First Name, Last Name	1 (normalizer)	1 (normalizer)	0.27	0.09	1.00 (normalizer)	0.55

Table 1. Summary of the search log of LinkedIn people search. All the numbers were normalized for proprietary reasons. Column 2 shows the top 6 most popular query tags in the search log. Column 3 shows the percentage of multiple-click search of the 6 query tags (Normalized by the percentage of the query tagged as First Name, Last Name). Column 4, 5, and 6 display the CTR of search results with different network distances (Normalized by CTR on 1st-degree results of each query tag). Column 7 and 8 show the fraction of searches that lead to new connections after clicks on results. (Normalized by the connect rate of First Name, Last Name search after a click)

that our tagger has a precision of over 90%, and over 99% for identifying first and last names.

We first analyzed the percentage of the searches that resulted in multiple clicks when at least one click is performed based on their query tags (see Column 3 in Table 1). The results suggest that the sequence of tags describing a query was highly predictive of whether a search would result in single click or multiple clicks. Queries tagged as Job Title (e.g., “software engineer”) were about 14 times more likely to have multiple clicks than queries tagged as First Name, Last Name. A Chi-squared test found that all the difference between the groups was significant with $p < 0.001$. This suggests that we can use the query tags to infer the intent of the searchers. When users issue name queries (e.g., First Name, Last Name), they are more likely to be interested in a single target. Therefore, most of the times these searches result in only one click. On the other hand, when users issue non-name queries (e.g., Job Title and Skill), the users often want to explore the results and click on more than one result.

To better understand the targets of the searches, we computed the aggregated click-through rate (CTR) over search results, broken down by the network distance between the searcher and person in the search result.

We found the patterns were very different for searches with different query tags. For name searches, 1st-degree connections yield the highest CTR, and the CTR decreases as the network distance increased. In contrast, for non-name searches, the CTR for the results of 1st-degree connections is comparatively much lower. For Title and Skill queries, the CTR for the results of 1st-degree connections are the lowest. The CTR is maximal for 2nd-degree connections, and then drops for higher network distances – but is still higher than or at least similar to the CTR for direct connections. (see Columns 4, 5, and 6 in Table 1) The Chi-squared tests show that all the differences of the CTR on results with different network distances were statistically different ($p < 0.001$). These results suggest that, while name searches often target people that are already in the searcher’s connections, non-name searches usually prefer results that are not in the searcher’s existing connections. More specifically, non-name searches prefer 2nd-degree results – that is, the people with whom the searcher shares connections.

We further analyzed whether the users who search for people beyond their 1st-degree connections connect with them after the searches. Our analysis found that the rate of building a new connection (an invitation to connect that is accepted) dropped when there were no shared connections between the user and their search results. This pattern is especially strong for non-name searches. The Chi-squared tests suggest that there are significant differences between the connect rates after clicks on 2nd-degree results and the connect rates after clicks on 3rd-degree results ($p < 0.001$).

We found that users who issue non-name queries were much more likely to connect with the 2nd-degree results they clicked on than those who issued name searches. But this pattern was reversed when there was no shared connection: users who issued non-name queries were much less likely than those who issued name queries to connect to the 3rd-degree (or more distant) results they clicked on.

In summary, our log analysis shows that query tags are highly associated with the click behavior and connection rate for people search results with different network distances. For name searches, searchers often click on one result, and network distance negatively correlates to both CTR and connection rate. Users who search for non-name queries, however, are most interested in 2nd-degree connections, as evidenced by both CTR and connection rate.

3. USER STUDY

Our log analysis shows that the users who search for non-name queries behave much differently from users who search for name queries. Specifically, their click behaviors suggest they are performing exploratory search [6]. This means they are not searching for a specific target. Instead, they are exploring a set of people search results with desired characteristics. This is surprising because current people search systems focus on supporting name searches with specific targets. To gain more insight into how and why users conduct exploratory people search on LinkedIn, we conducted a user study. We surveyed 20 LinkedIn members who had performed non-name searches on LinkedIn. The participants were 60% male, and 65% were US-

based. Their stated motives for using LinkedIn broke down as follows: 60% were recruiters looking for candidates, 25% were job seekers looking for new career opportunities, and 15% of them were sales representatives or consultants looking for clients.

We provided the participants with the following definition of an exploratory search: “searching for people with some set of characteristics, as opposed to searching for a specific person.” We then asked them a set of questions that included the following:

- How do you use LinkedIn for exploratory search?
- When you perform exploratory searches, are you more interested in finding people you already know, (1st degree connections); people you don't know, but with whom you share connections (2nd-degree connections); or anyone, regardless of whether you know them or share connections with them.

We also asked participants about the actions they performed after finding people through search, as well as the general strategies they used to achieve their goals.

3.1 Exploratory searchers target people that are not their 1st-degree connections

The results of our log analysis suggest that in exploratory people search, users are more interested in the results that are not their direct connections. Our user study also supports this finding. Here are some excerpts from participant responses:

“If the person is someone I know, I can send them email directly and don't need to use the search function at all.” (P9, job seeker)

“The people in my first-degree connections are either the ones that are already working in my company or those who didn't get [an] offer after the interview process, so I would not send invitations to them.” (P2, recruiter)

“I am more interested in growing my network and finding little gems who don't have many connections. It's more likely they've not already been contacted by my competitors.” (P6, sales)

Exploratory searchers target people that are not in their existing connections for two main reasons:

1. Searchers contact people they already know (1st degree) through other methods, e.g., phone, email, face-to-face meeting.
2. People they don't know are more likely to be useful toward satisfying the needs of an exploratory search, e.g., finding candidates to fill a job position.

These results suggest that exploratory people search on LinkedIn allows people to locate and connect to people that are not in their existing connections. These people can be more important for their career, and usually are hard to connect through earlier traditional methods (e.g., phone and email).

3.2 Shared connections increase the chances for users to connect with people

Both our log analysis and our user study indicated that shared connections lead searchers to make new connections after exploratory searches.

The user study found two strategies that searchers employed in reaching out to 2nd-degree connections.

The first is to mention shared connections:

“I usually write something like: ‘Hi, I saw you also know X. He worked with us as an intern last summer.’ This greatly increases the response rate.” (P2, recruiter)

“Most of the time, people do not reply to people without shared connections, so I seldom try to contact people in my third degree connections.” (P11, job seeker)

The second is to ask the shared connection for an introduction:

“When I find the people in the company that I'm interested in, I will not contact them directly. I will first send a message to one of our common friends for introducing us.” (P12, job seeker)

“I feel like the most important thing for me to get a response from my target is whether my friend is willing to introduce us or not.” (P13, job seeker)

Searchers also said that common characteristics (e.g., education, previous employment) also increase the likelihood of a positive response:

“I would say the response rate for ordinary recruiters is just around 20%. For alumni, the response rate will increase to about 50%” (P16, job seeker)

“When I try to connect to someone, I will write in the message that says: it's weird that we haven't met since we have so much in common. Let's connect!” (P19, recruiter)

In sum, even though exploratory search users target people that are not in their existing connections, they still rely on the shared connections and other common characteristics to build up trust between each other. Hence, they place the highest value on 2nd-degree connections.

4. IMPLICATIONS

Our results have several design implications for people search in online professional networks.

First, the ranking function in the people search system – and perhaps the overall search experience – should adjust based on the query type. Our analysis suggests that users who perform non-name searches prefer results that are not in their 1st-degree connections, which is the opposite of what we see for name searches. The search engine should assign a lower rank to 1st-degree connections for non-name searches. Since we also found that the query tagger can be very accurate, we advocate an

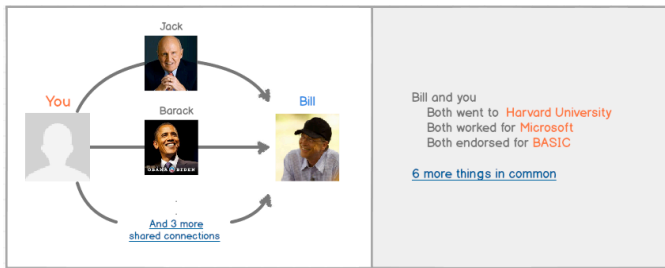


Figure 1. An interface prototype that lists all the common characteristics between the users and their indirect ties.

approach where the search engine adjusts the ranking model and interface when it detects a different search intent.

In addition, the search interface should clearly show how the 2nd-degree connections are related to the users. Since people use shared connections and characteristics to build new connections, providing this information when the users explore the results can help them decide which results are most useful. The current LinkedIn interface shows users their number of shared connections. This abstract number, however, may not be detailed enough for detailed exploration. Figure 1 shows an alternative interface (based on a visualization from the LinkedIn profile page) that presents more information about the common traits between users.

Moreover, our study also provides implications for social networking research. In Granovetter's seminal work on the strength of weak ties [2], he showed that in face-to-face environments people use their weak ties to connect to 2nd-degree connections. These weak ties, which serve as bridges between strongly connected groups, created the most opportunities for social mobility. While more work is necessary to study the parallels of Granovetter's findings in online spaces, to date, our findings confirm Granovetter's work that people search various opportunities through their 2nd-degree connections. Professional online networks such as LinkedIn are changing the job search and recruit landscape. Unlike in face-to-face job interaction, they enable members to explicitly identify and access their 2nd-degree connections without direct help from weak ties. Perhaps it is time to create new tie theories for a world where people have richer access to their extended professional networks.

Finally, the system should help the users to evaluate their chances to get positive response by calculating the strength of the path to the indirect ties. Previous research has shown that the strength of direct ties [1] can be accurately predicted. However, calculating the strength of indirect ties is more complicated than the strength of direct ties. It involves at least three main aspects:

1. The tie strength between the user and the user's direct tie which served as the bridge.
2. The tie strength between the direct tie and the indirect tie.
3. The common characteristics between the user and the indirect tie.

Understanding how these three aspects contribute to the strength of indirect ties allows us to find the best candidates and the optimal path that can reach to them. This also lets us reexamine

the weak tie hypothesis to see whether weak or strong ties can be better bridges to reach out to their targets.

5. CONCLUSIONS

We found two key contributions in our study of people search on LinkedIn. First, Our study suggests that search intention can be detected using automatic query tagging. We found that users who perform non-name searches on LinkedIn target their 2nd-degree connections because they are unexplored, but still have the shared connections increase their chances of making a new connection. This is significantly different from the user behavior of searchers who issue name queries, where both CTR and connection rates drop as the network distance increases.

In addition, our study shows that LinkedIn people search allows searchers to engage their 2nd-degree connections directly. How could the system better support the interactions between the people who don't know each other in the real world? And how does this change the way people seek new career opportunities? These questions represent opportunities for subsequent research, and are compelling research directions in the areas of online social and professional networking.

6. ACKNOWLEDGMENTS

We would like to thank our survey participants for their insight and time. In addition, we thank Igor Perisic, Pei-Fen Tu, Nikita Spirin, and the LinkedIn Query Understanding Team and Search Relevance Team for valuable feedback.

7. REFERENCES

- [1] Gilbert, E. and Karahalios, K. Predicting tie strength with social media. ACM CHI'09 (2009), 211
- [2] Granovetter, M.S. The strength of weak ties. American journal of sociology, (1973), 1360–1380.
- [3] Guy, Ido, Sigalit Ur, Inbal Ronen, Sara Weber, and Tolga Oral. "Best faces forward: a large-scale study of people search in the enterprise." In Proceedings of ACM CHI'12 (2012).
- [4] Scale, Mark-Shane. "Facebook as a social search engine and the implications for libraries in the twenty-first century." Library Hi Tech 26.4 (2008): 540-556.
- [5] Weerkamp, Wouter, Richard Berendsen, Bogomil Kovachev, Edgar Meij, Krisztian Balog, and Maarten de Rijke. "People searching for people: Analysis of a people search engine log." In Proceedings of ACM SIGIR'11 (2011).
- [6] White, R.W. and Roth, R.A. Exploratory Search: Beyond the Query-Response Paradigm, (2009), 1–98.
- [7] <http://press.linkedin.com/about>.
- [8] <https://www.linkedin.com/today/post/article/20131007130344-50510-optimize-for-the-short-snout>