Unearthing People from the SaND: Relationship Discovery with Social Media in the Enterprise

Adam Perer
IBM Research
adam.perer@us.ibm.com

Ido Guy, Erel Uziel, Inbal Ronen, Michal Jacovi
IBM Research Haifa
{ido, erelu, inbal, jacovi}@il.ibm.com

Abstract

The popularity of social media across the Internet results in people-centric data sources that users can potentially leverage for relationship discovery tasks. In this paper, we describe SaNDVis, a visual system that supports tasks like expertise location, team building, and team coordination in the enterprise. By integrating social position, evidence, and facets into SaNDVis, we demonstrate how users can reflect on existing relationships as well as build new relationships in an enterprise setting.

Introduction

As the use of online social media continues to grow, rich data sources become available that can potentially be leveraged to support tasks in the enterprise. For instance, people may wish to locate experts or build teams seek to find people relevant to their interests so they can build new relationships. Furthermore, people may wish to reflect on existing relationships to understand how information flows through their company and how much people collaborate. We refer to these people-centric tasks as *relationship discovery* tasks because they are tasks in which users are examining or creating new relationships.

In this paper, we focus on extracting a social graph from data inside the enterprise. Such data includes online social media, like blogs, bookmarks, and communities, as well as traditional media like papers, patents and organizational charts. In order to help users manage this complex multi-dimensional information, we present SaNDVis, a novel relationship discovery tool that helps users accomplish real enterprise tasks. SaNDVis not only represents the social graph, but also highlights evidence for why relationships exist as well as linking to related documents.

Related Work

Relationship discovery is related to various research areas previously studied. *Social matching* systems have various motivations for people searching for other individuals: dating, pursuing shared interests, addressing community issues, solving technical problems, or even just having a good conversation (Terveen and Mcdonald 2005).

Expertise location systems are typically approached as social matching systems with respect to a certain topic. These systems allow the user to enter a search query and output a list of potential experts e.g., (Reichling et al. 2005). There are a variety of motivations for expertise location in a large organization, including the most common ones: “getting answers to technical questions” and “finding people” (Ehrlich and Shami 2008). The visualization in this work is based on retrieving the top people related to a search query, and thus can be viewed as an enhanced expertise location system.

Several expertise location systems incorporate social network information in addition to matching person profiles to queries e.g., (Jarvenpaa and Leidner 1999). Expertise Recommender filters expert search results based on two elements of the user's social network: organizational relationships and social relationships gathered through ethnographic methods such as interviews (McDonald et al. 2001).

Visualizations have been used as aids for understanding social networks since the 1930s, typically represented as node-link diagrams (Freeman, 2000). There is a body of research focusing on visualization of personal networks e.g., (Fisher and Dourish 2004), (Heer and boyd, 2005), (Perer and Shneiderman, 2006), (Viegas et al. 2004).

There have been several works studying social networks around a specific topic. For example, an algorithm for detecting the most authoritative and sociable individuals in social networks has been described in (Chen et al. 2009). SmallBlue is a social networking application, which allows searching for experts and analyzing social paths among them (Lin et al. 2008).

SaNDVis

SaNDVis is a people-centric visual interface for supporting relationship discovery tasks. The SaNDVis UI was implemented to run in a standard web browser using Adobe's Flash framework. To begin using the interface, users enter a topic in the textbox at the top of the interface. After a topic is entered, three components of the interface are populated with results: 1) a social graph visualization of top-ranking people that match the query, 2) an evidence overview of the documents and tags associated with the people who match the topic, and 3) a facet overview of the attributes of the matching people.

The Social Graph View

As people are the focus of relationship discovery, the largest component of SaNDVis is the social graph view. The top *n* people related to the user's topic are displayed (by default, *n*=25). However, in this view, people are not simply represented as a textual list but instead displayed using social graph visualization. While such a display is more complex to comprehend than a list, the visualization highlights a pivotal type of information relevant to relationship discovery: *social position*.

Social position is important because users will typically be unfamiliar with most of the people who match their queries. However, by seeing how those people connect to themselves, their peers, or known individuals, users can gauge which people are better suited for their relationship tasks. Social position can also be a barometer for judging whether or not a matched person might be willing to communicate with the user. For instance, prior work shows that 'social software participation' is a significant signal of likelihood of contact (Shami et al. 2009). Finding a matched person with few social connections may be adequate but finding a well-connected individual might better meet the user's needs.

Social position, as shown in Figure 1, is conveyed via a social graph visualization. Nodes represent each of the top people matching the user's topic, and edges represent the types of relationships that connect various people. Each node features the person's name and image. As there can be multiple categories of relationships connecting two individuals, bands are added for each edge representing each

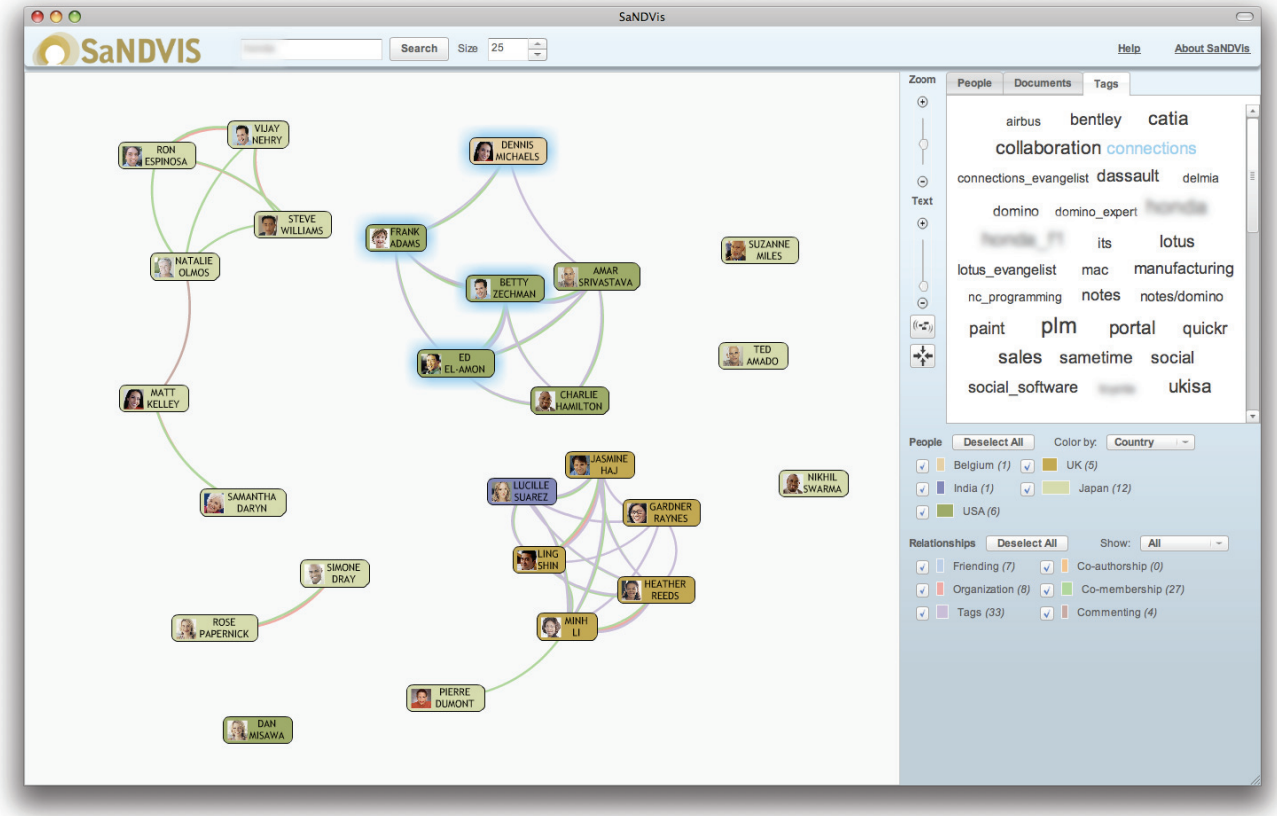

Figure 1. The SaNDVis UI contains three views relevant to relationship discovery task.
A) On the left, a social graph view allows users to interpret the social position of the results of a topic query.
B) On the top right, an evidence view allows users to examine the documents, tags, and people associated with the query.
C) On the bottom right, a facet view allows users to get an overview of the categories with histograms and allows users to filter out any irrelevant categories.

category – producing a "rainbow" when multiple categories are present. Thus, the thickness of an edge is relative to the overall relationship score.

While social graph visualizations have a tendency to be complex, sometimes derisively compared to hairballs or spaghetti, SaNDVis's design attempts to maximize visual legibility. Nodes and links are positioned using an advanced force-directed, stress majorization algorithm to minimize node overlaps and edge crossings. The number of nodes is, by default, kept to only 25 so the visualization is optimized to the design guidelines for achieving "NetViz Nirvana" (Dunne and Shneiderman 2009), but users can increase this number.

The Evidence View

In order for users to understand why people are connected in the social graph, users need access to all available information, including people, documents, and tags. In SaNDVis, it is possible for users to freely pivot from one data type to another to find the information they need. This functionality is powerful, as providing only a list of names is often not enough to be of practical use when results include people the user is unfamiliar with. By providing coordinated evidence that the user can explore, users can become acquainted enough to judge whether or not the person is a useful result. We organize the three types of information into three separate tabs in the evidence view.

Initially, the tag tab is shown which presents a tag cloud of all of the tags related to the people in the social graph. The size of each tag is proportional to the number of people on the left that are associated with that tag. As shown in Figure 1, users can mouse-over a tag (in this case, 'connections') and the people associated with that tag are highlighted in blue on the left. Conversely, users can mouse-over a specific person in the social graph view and see all of the tags associated with that person. Users can also click on a tag to filter the view to include only those people that are highlighted. This allows users to drill-down to interesting subcomponents of their initial results.

Clicking the documents tab, users are provided with a table of the top documents associated with the matched people and the query. Users can inspect these documents by double-clicking to navigate to them in the web browser. Users can inspect the documents associated with a specific person and the topic query by clicking on the corresponding node. These documents serve as "evidence" for why the top people were associated with the query. Additionally, users can inspect the documents connecting two individuals by clicking on the edge that connects them. These documents serve as the "evidence" for the corresponding relationship and may include papers or wikis that both individuals have co-authored, communities they are co-members of, web pages they have both bookmarked, and so on.

The people tab gives users the option of viewing the list of people in a more traditional table. While such a table view does not directly show relationships, the table is coordinated with the social graph so users can highlight and pivot to specific people in either view. This view can be useful if users want to sort the list of people by an attribute to quickly find a node of interest.

The Facet View

SaNDVis allows users to use facets to filter down results to the types of people they care about. Attributes of the people relevant to users' tasks should be accessible and filterable. For instance, in an enterprise scenario, a user may wish to build a local team in China and analyzing matched people in the USA would not be relevant to the task at hand. Similarly, if the user is looking for experts to answer a technical question, sales people may not be appropriate.

SaNDVis currently allows filtering nodes by two facets related to relationship discovery in the enterprise: location and division. For each unique type of category (e.g., USA or Research) in the results, a histogram and checkbox appear, as shown in Figure 1. The histogram represents how many nodes of that type appear in the search result. Users can remove nodes of this type by unselecting the checkbox. Each type of facet is also given a unique color, and all of the nodes that fall into that category are colored the same in the social graph view.

Furthermore, users can also filter edges, where filtering can be done according to the two dimensions. The user can choose to filter to familiarity-only or similarity-only relationships (default includes both), and independently use the checkboxes next to each of the six categories of relationships to include only categories of interest. For example, in cases where the original graph is very dense, the user can choose to focus on familiarity relationships based on org chart and friendship only.

When users filter according to facets, animation is used to maintain the user's mental model for the transition between states. The system also optimizes screen real estate by automatically fitting the nodes. So if the user filters out a certain class of nodes, the system will zoom-in automatically on the remaining nodes. By default, filtering edges will not rerun the force-directed layout procedure as this can easily disrupt user's mental model, even with animation. Users can manually initiate a re-layout at any time.

Data Sources

SaNDVis uses SaND to mine, aggregate and crawl public information from multiple services within the enterprise (Ronen et al. 2009). SaND generalizes and extends SONAR, a system previously introduced to aggregate social network information (Guy et al. 2008). The fact that the mined information is public allows for the presentation of all the underlying relationships without exposing information to which the user does not have access in their original service. Moreover, the fact that SaND is based on public information allows a high level of transparency, by presenting "evidence" for each relationship.

In this work, the following enterprise services are mined:

- A blogging system with 16,300 blogs, 144,200 entries, 121,750 comments, 70,000 overall users and 357,000 tags (Huh et al. 2007)

- A wiki system with 6,150 public wikis, 13,000 editors and 24,450 tags
- A social bookmarking system with 1.1M bookmarks by 68,000 users with 3M tags (Millen et al. 2006)
- A file sharing system with 46,700 public files used by 31,800 users and tagged with 86,000 tags (Muller et al. 2009)
- A community system with 9,400 online communities, each including resources such as feeds and forums, with an overall of 226,000 members and 32,500 tags
- An organizational chart including nearly 450,000 employees
- An enterprise SNS that allow users to reciprocally connect to each other, with an overall of 250,000 connections between 99,000 users (DiMicco et al. 2008)
- A patent database with 132,000 patents authored by 31,500 users
- A publication database that includes 28,950 papers authored by 3,200 users
- A projects wiki that includes 1,980 projects with 1,260 members and 2,450 tags (Danis and Singer 2008)
- An open source project system with 1,860 projects and 11,850 total members
- A forum system with 2,590 forums, 466,300 threads and 53,000 users
- A people tagging application that allows users to tag each other, with 9,300 users who tagged 50,000 other individuals with 160,000 public tags (Farrell et al. 2007)

From these sources, almost 450,000 people and over 73 million relationships were mined.

To allow users to better leverage this data, we categorize the relationships into six categories: organizational, friending, tagging, commenting, co-authorship, and co-membership. We also distinguish between relationships that are likely to reflect familiarity between two individuals (e.g., tagging each other or having a common manager) and relationships that are likely to reflect similarity between the individuals (e.g., using the same tag or commenting on the same blog entry).

Conclusion

In this paper, we have demonstrated a system that supports relationship discovery in the enterprise. By integrating social position, evidence, and facets into a visual interface, SaNDVis, users can leverage existing social media behavior to assist them in their relationship discovery tasks, including expertise location, team building, and personal reflection.

References

Chen, D., Tang, J., Li, J., and Zhou, L. 2009. Discovering the staring people from social networks. *Proc. WWW '09*, 1219-1220.

Danis, C. and Singer, D. 2008. A wiki instance in the enterprise: opportunities, concerns and reality. *Proc. CSCW '08*, 495-504

DiMicco, J., Millen, D.R., Geyer, W., Dugan, C., Brownholtz, B., and Muller, M.J. 2008. Motivations for social networking at work. *Proc. CSCW '08*, 711-720

Dunne, C. and Shneiderman, B. 2009. Improving graph drawing readability by incorporating readability metrics. *University of Maryland, Technical Report* HCIL-2009-13.

Ehrlich, K. and Shami, N. 2008. Searching for expertise. *Proc. CHI '08*, 1093-1096.

Farrell, S., Lau, T., Nusser, S., Wilcox, E., and Muller, M. 2007. Socially augmenting employee profiles with people-tagging. *Proc. UIST '07*, 91-100.

Fisher, D. and Dourish, P. 2004. Social and temporal structures in everyday collaboration. *Proc. CHI '04*, 551-558.

Freeman, L. Visualizing Social Networks. *Journal of Social Structure*, 1, 2000.

Guy, I., Jacovi, M., Shahar, E., Meshulam, N., Soroka, V., and Farrell, S. 2008. Harvesting with SONAR: The value of aggregating social network information. *Proc. CHI '08*, 1017-1026.

Heer, J. and Boyd, D. 2005. Vizster: Visualizing online social networks. *Proc. InfoVis '05*.

Huh, J., Jones, L., Erickson, T., Kellogg, W. A., Bellamy, R. K., and Thomas, J. C. 2007. BlogCentral: the role of internal blogs at work. *Proc. CHI '07*.

Jarvenpaa, S. L. and Leidner, D.E. 1999. Communication and trustinglobal virtual teams. *Organization Science*, 10(6), 791–815.

Lin, C., Ehrlich, K., Griffiths-Fisher, V., and Desforges, C. 2008. SmallBlue: People mining for expertise search. *IEEE MultiMedia* 15, 1 (Jan. 2008), 78-84.

McDonald, D. W. 2001. Evaluating expertise recommendations. *Proc. GROUP '01*, 214-223.

Millen, D.R., Feinberg, J., & Kerr, B. 2006. Dogear: Social bookmarking in the enterprise. *Proc. CHI'06*, 111-120.

Muller, M.J., Millen, D.R., and Feinberg J. 2009. Information curators in an enterprise file-sharing service. *Proc. ECSCW '09*, 403-410.

Perer, A. and Shneiderman, B. Balancing Systematic and Flexible Exploration of Social Networks. In *IEEE Transactions on Visualization and Computer Graphics*. 12(5), 693-700.

Reichling, T., Schubert, K., and Wulf, V. 2005. Matching human actors based on their texts: design and evaluation of an instance of the ExpertFinding framework. *Proc. GROUP '05*, 61-70.

Ronen, I., et al. 2009. Social networks and discovery in the enterprise (SaND). *SIGIR '09 Demos*.

Shami, N. S., Ehrlich, K., Gay, G., and Hancock, J. T. 2009. Making sense of strangers' expertise from signals in digital artifacts. *Proc. CHI '09*, 69-78.

Terveen, L. and McDonald, D. W. 2005. Social matching: A framework and research agenda. *ACM Trans. Comput.-Hum. Interact.* 12, 3 (Sep. 2005), 401-434.

Viegas, F. B., Boyd, D., Nguyen, D. H., Potter, J., and Donath, J. 2004. Digital artifacts for remembering and storytelling: PostHistory and Social Network Fragments. *Proc. HICSS '04*.