Visual Social Network Analytics for Relationship Discovery in the Enterprise

Adam Perer[*], Ido Guy[†], Erel Uziel[†], Inbal Ronen[†], Michal Jacovi[†]

[*]IBM Research, Cambridge   [†]IBM Research, Haifa

1 ABSTRACT

As people continue to author and share increasing amounts of information in social media, the opportunity to leverage such information for relationship discovery tasks increases. In this paper, we describe a set of systems that mine, aggregate, and infer a social graph from social media inside an enterprise, resulting in over 73 million relationships between 450,000 people. We then describe SaNDVis, a novel visual analytics tool that supports people-centric tasks like expertise location, team building, and team coordination in the enterprise. We also provide details of a 12-month-long, large-scale deployment to almost 1,800 users from which we extract dominant use cases from log and interview data. By integrating social position, evidence, and facets into SaNDVis, we demonstrate how users can use a visual analytics tool to reflect on existing relationships as well as build new relationships in an enterprise setting.

KEYWORDS: information discovery, social networks, social data mining, social visualization

INDEX TERMS: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle;

2 INTRODUCTION

There is a growing need for visual analytics in the workplace as information overload continues to grow. Countless new sources of knowledge appear in the public everyday, ranging from blog posts and wiki pages to papers and patents. Many people aim to keep track of all data sources that produce information relevant to their interests, but there is too much information being produced and too little time to manage it all. Instead, users are forced to rely on sophisticated information discovery tools to locate data on-demand (i.e., web search engines). Such tools crawl, mine, and rank data sources and organize the data in a manner that allows users to reach large amounts of information. However, most systems that support information discovery are document-centric: their databases are indexed by documents and their user interfaces focus on documents as well.

Nevertheless, not all users performing information discovery tasks are interested in documents. For instance, users trying to locate experts or build teams seek to find people, not documents, relevant to their interests so they can build new relationships. Another example are users who wish to reflect on existing relationships to understand how information is flowing through their company and how much people are collaborating. We refer to these people-centric tasks as *relationship discovery* tasks because they are tasks in which users are examining or creating new rela-

tionships.

Relationship discovery is not trivial to support, as the backbone of a people-centric discovery system is a social graph, not a traditional document-indexed database. However, it is possible to extract a social graph from documents with social attributes. In this paper, we focus on extracting a social graph from documents in the enterprise. Such documents include traditional media like papers, patents and organizational charts, as well as online social media, like blogs, bookmarks, and communities.

There are several differences between document discovery tasks and relationship discovery tasks. For instance, in order to judge the relevance of a result set of people, users may require access to features rarely present in document discovery interfaces. As an example, access to the social structure of relationships is critical to understanding who the people are, how well they communicate with others, and which ties they may share in common with the user. Furthermore, a list of people's names may not be enough if they are strangers, so access to evidence of why particular people were chosen is crucial. Also, users should be able to filter by meaningful facets, as not all candidates can fulfill all users' needs equally. For instance, an enterprise team-building task might require finding people with certain types of job roles at particular locations. We take these considerations into the design of SaNDVis, a visual analytics tool that better supports relationship discovery than typical information discovery interfaces.

The contribution of this work is an end-to-end analytics system that supports relationship discovery in the enterprise. Such a system requires several components: SaND mines and aggregates social media from dozens of data sources, SaNDGraph organizes the data into a people-centric database to support fast social graph queries, and SaNDVis provides a visual analytics UI to help users manage this complex, multi-dimensional information. SaNDVis not only represents the social graph, but also highlights evidence for why the relationships exist as well as linking to related documents. SaNDVis receives most of the focus in this paper, as it is the visual analytics interface to which users can leverage the rich data collected by the other components for real enterprise tasks.

Our paper begins with a reflection on related work, where we discuss social matching, expertise location, and visual social network analytics research. Then, we explain SaND, quantify all of the data sources it mines, and describe the implementation of SaNDGraph, which infers a social graph from enterprise documents. We follow this description with SaNDVis, a novel visual analytics tool for relationship discovery. We then describe an evaluation based on logging system data, followed by interviews to further understand how users utilized the system. Finally, we discuss positive and negative aspects of our work, and then conclude.

3 RELATED WORK

We first describe research that aims to support relationship discovery tasks in the enterprise without visual analytics. We then describe recent visual analytics work that aims to help users make sense of social networks.

---

[*]email: adam.perer@us.ibm.com

[†]email: {ido, erelu, inbal, jacovi}@il.ibm.com

3.1 Social Matching and Expertise Location

Relationship discovery is related to various research areas previously studied. Terveen and Mcdonald [41] define a framework for *social matching* systems and enumerate different motivations for people searching for other individuals: dating, pursuing shared interests, addressing community issues, solving technical problems, or even just having a good conversation.

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. For example, Reichling et al. [36] describe an expertise location system based on text analysis - user profiles are generated and matched by extracting keywords from related text documents. Vivacqua et al. [45] present Expert Finder, a tool that matches novices and experts by analyzing java documents related to them. Ehrlich et al. [9] study the motivations for expertise location in a large organization and find that the most common are "getting answers to technical questions" and "finding people" (the latter especially common for people in client facing roles). The visualization in this work is based on retrieving the top people related to a search query, and hence 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. ReferralWeb [25] combines social networks and collaborative filtering and allows the user to specify both a search topic and a social criterion (e.g., people who are related by up to two degrees to John Smith). Expertise Recommender [30][29] 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. In this work we do not filter results based on social network information, but rather present the relationships among the returned "experts" to allow better judgment by the user. Campbell et al. [4] show that considering both content-based expertise and the social graph, as reflected in one's email, is more effective in expertise identification.

In recent years, as social media has evolved, more means for mapping people to topics have become available. In particular, people tagging allows the crowd to assign a person with descriptive terms. Collabio [3] is a Facebook application that encourages friends to tag each other through a game. Collabio tags are shown to be accurate and augment information that could have been scraped online. Farrell et al. [11] present people tagging in the enterprise and show that tags given by others are accurate descriptors of the employee's interests and expertise. People tagging is one of many data sources used in this work for mining expertise and other relationships.

3.2 Visual Social Network Analytics

Visualizations have been used as aids for understanding social networks since the 1930s [13], using a variety of visualization techniques [23]. Typically, node-link diagrams, matrix visualizations [17], or hybrids [21][22] are used to represent the linkage structure. Another alternative is to visualize networks according to attributes of the nodes using semantic substrates [39] or rolled-up PivotGraphs [42]. NetLens [26] provides multiple coordinated views of node attributes using simple data graphics, and ManyNets [14] enables comparison of multiple networks using tabular views of summary statistics. There are also a variety of network analysis tools that combine visualization and statistics inside an interactive visual analytics system [1][16][34][40].

There is a body of research focusing on visualization of personal networks. PersonalMap [10] presents to the user a social map of her email social network, highlighting the different groups within the network and the intensity of communication with different individuals. ContactMap [43] provides a social desktop representation of the user's contacts in a way that makes demographic attributes such as group affiliation or geographic location more prominent. Soylent [12] as well as Post History and Social Network Fragments [44] are visualizations designed to reveal social patterns in user activity (primarily based on email), with substantial focus on temporal aspects. While these tools help users manage their own contact lists, they do not expose them to individuals beyond their egocentric network. Vizster is an example of an interactive visualization system using node-link diagrams, which supports end-user exploration of online social networks [20]. Like Vizster and TouchGraph[1], we represent social networks using a node-link diagram as it is the representation of social networks most familiar to users and is visually comprehensible for the size of networks we display [22].

There have been several works studying social networks around a specific topic. For example, Chen et al. [5] present an algorithm for detecting the most authoritative and sociable individuals in social networks. To this end, they create co-authorship graphs around domains such as "data mining" or "databases". SmallBlue [28] is a social networking application, which allows searching for experts and analyzing social paths among them. One of its components, SmallBlue Net, presents the social network around a certain topic based on email communication data. While the SmallBlue visualization is completely focused on people and their relationships, in this work we present a more comprehensive view of enterprise relationships that include people, documents, and tags. Moreover, as email data is private and sensitive, no explanations can be presented for why a person is related to a topic or why two people are related to each other, in contrast to our work that provides evidence we argue is critical to users tasks.

4 SOCIAL NETWORKS & DISCOVERY (SAND)

4.1 Relationship Aggregation

Social **N**etworks & **D**iscovery (SaND) is a relationship aggregation analytics system that mines multiple services within the enterprise. SaND models the relationships between three core entities – people, documents, and tags – as reflected in the different services. The fact that SaND is used within the enterprise facilitates handling user identities, as each employee has a single ID used across all organizational services. SaND generalizes and extends SONAR, a system previously introduced to aggregate social network information [17]. While SONAR aggregates solely people-to-people relationships, SaND aggregates any of the 9 types of relationships between two sets: [people, documents, tags] and [people, documents, tags]. This generalization allows the combination of social and content analysis to optimally unleash the value of information unlocked throughout different enterprise sources.

Figure 1 depicts the basic direct relationships modeled in SaND. A person can be related to a document (a wiki, a shared file, a bookmarked web page, etc.) in a variety of ways (e.g., as an author, a commenter, or a sharer). Similarly, a person can be related to tags (e.g., used or was tagged with a tag). Finally, a person can be related to another person (e.g., a friend on an SNS, connected on an enterprise org chart). Additionally, a document is directly related to a tag if it includes it within its content (as in a

---

[1] http://www.touchgraph.com

Figure 1. Direct Relationships Modeled in SaND

search index) or is tagged with it. In addition to direct relationships, two entities can be indirectly related to each other through another entity. For example, a person can be related to a tag applied on a document s/he authored or to another person if they share a common friend; a document can be related to another document if they are both tagged with the same tag.

SaND uses a weighting scheme to calculate a relationship score between two given entities. For example, an authorship relationship is of higher weight than a commenting relationship and friendship is considered stronger than tagging. Several other parameters are considered, such as the date in which a relationship was formed, or the numbers of other co-authors or co-members. Ultimately, the overall relationship score between two entities is determined by summing over all relationships – direct and indirect – that connect them. More details about relationship weighting scheme can be found in [1] and [18].

SaND exposes its aggregated relationships through an API that allows other applications to query for entities. For example, an application can query for all tags and documents related to a certain person, ranked by their relationship score to that person. Some parameters, such as a date range or the number of desired results can be added to the query request.

The implementation of SaND is based on the unified approach [1]. While a regular index maps terms to documents, the unified index maps people, documents, and tags to people, documents, and tags. This allows querying for each subset of entities given another subset of entities.

4.2   Data Sources

SaND mines public information from different services within a large organization. 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 an explanation, or "evidence" for each relationship. For example, it can be shown that a user is related to a tag as s/he was tagged with it by 3 other individuals; or that two individuals are related to each other because they co-authored two wikis and are co-members in three communities.

In this work, the following enterprise services are mined by SaND by scraping web content or using provided APIs, where available:

- A blogging system [24] with 16,300 blogs, 144,200 entries, 121,750 comments, 70,000 overall users and 357,000 tags

- A wiki system with 6,150 public wikis, 13,000 editors and 24,450 tags
- A social bookmarking system [31] with 1.1M bookmarks by 68,000 users with 3M tags
- A file sharing system [32] with 46,700 public files used by 31,800 users and tagged with 86,000 tags
- 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
- Two enterprise SNSs [7][11] that allow users to reciprocally connect to each other, with an overall of 250,000 connections between 99,000 users
- 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 [6] that includes 1,980 projects with 1,260 members and 2,450 tags
- 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 [11] that allows users to tag each other, with 9,300 users who tagged 50,000 other individuals with 160,000 public tags

Figure 2 illustrates SaND's social search UI, similar to the unified interface described by Amitay el al. [1]. The user can search for a textual query or for a person (or any combination of the two) and the results include the related documents (on the right) as well as the related people and tags (on the left). While found very useful for enterprise search [1], the UI is document-centric, focusing on document results (as most web search engines), while the people and tags are presented as facets[2] of the returned documents. Many of the relationships described in Figure 1, such as those among the people in the result or among the people and tags, are not revealed in the document-centric UI. There are several differences between SaND and the unified search described by Amitay et al [2]. Notably, SaND adds a more comprehensive relationship model (direct people-to-people relationships did not exist in [2]) and the other system only mined three data sources. Further information about SaND is described briefly in [37].

In the section that follows, we describe a different, people-centric, visual analytics UI, which exposes a wider portion of the relationships among the entities in a result.

4.3   From Documents to People: Supporting Relationship Discovery

The document-centric interface of SaND's social search UI is quite powerful for finding textual information across the intranet. For relationship discovery tasks, new features in the database as well as the interface are needed. In order to support fast queries of relationships between people, a document-based index is no longer adequate. To resolve this, we built a people-centric database that we call SaNDGraph. SaNDGraph's database is indexed by people, so the server can retrieve queried relationships of peo-

---

[2] In a faceted search [19], the user can refine the results based on each of the categories (facets) chosen to characterize the data.
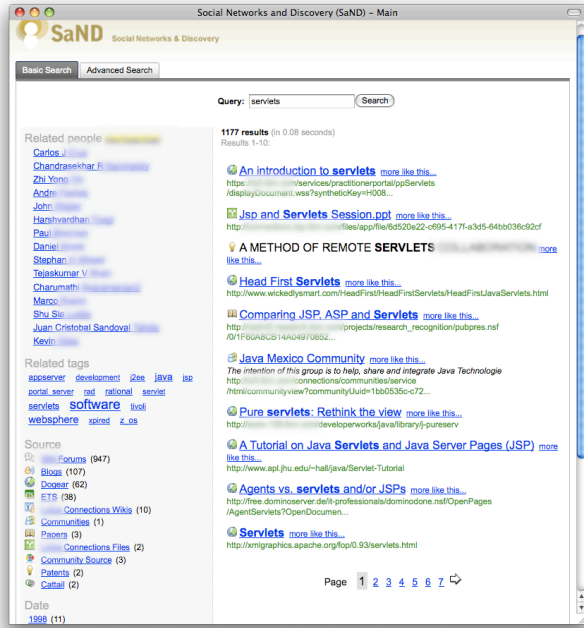
Figure 2. SaND's Social Search UI.

ple in real time, despite containing almost 450,000 people and over 73 million relationships.

SaNDGraph creates the people-indexed database by analyzing SaND's documents and then inferring a social graph based upon people's mutual activities across the social documents. As the people-to-people relationships that can be extracted from all these sources are complex, we allow two dimensions of abstraction for such relationships. First, we categorize the relationships into six categories: organizational, friending, tagging, commenting, co-authorship, and co-membership. Second, we 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). Table 1 illustrates the complete set of person-to-person relationships modeled by SaNDGraph.

5   SaNDVIS: A System for Relationship Discovery

SaNDVis is a people-centric visual analytics tool for supporting relationship discovery tasks. It utilizes SaNDGraph to allow efficient retrieval of results for social graph queries.

The SaNDVis UI, briefly introduced as a poster in [33], 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.

5.1   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 based on similarity 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 a social graph

	Familiarity	**Similarity**
Organizational Chart	Being manager or employee Sharing a 1st- or 2nd-level manager	
Friending	Being friends on an SNS	Having a common friend on an SNS
Tagging	Tagging a person Being tagged by a person	Co-tagging the same document Co-tagging the same person Co-usage of the same tag Being tagged by the same person Being tagged with the same tag
Commenting	Commenting to a person's document (blog entry or file) Comment by a person to own document (blog entry or file)	Co-commenting on the same document (blog entry or file)
Co-authorship	Co-authorship of a patent, paper, file, or wiki	Being shared a file by the same person Corresponding on the same forum thread
Co-membership	Co-membership in a small community (<=20 total members) or project	Co-membership in a large community (>20 total members)

Table 1. A classification all of the people-to-people relationships indexed by SaNDGraph. Familiarity relationships indicate the connected people probably know each other. Similarity relationships indicate the connected people behave similarly across social media.

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 analytic 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 [38]. 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 3, 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 (i.e., each of the relationship categories listed in Table 1). 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 category – producing a "rainbow" when multiple categories are present. Thus, the thickness of an edge is relative to the overall relationship score, determined by SaND's weighting scheme described in [18].

While social graph visualizations have a tendency to be complex, sometimes derisively compared to hairballs or spaghetti,
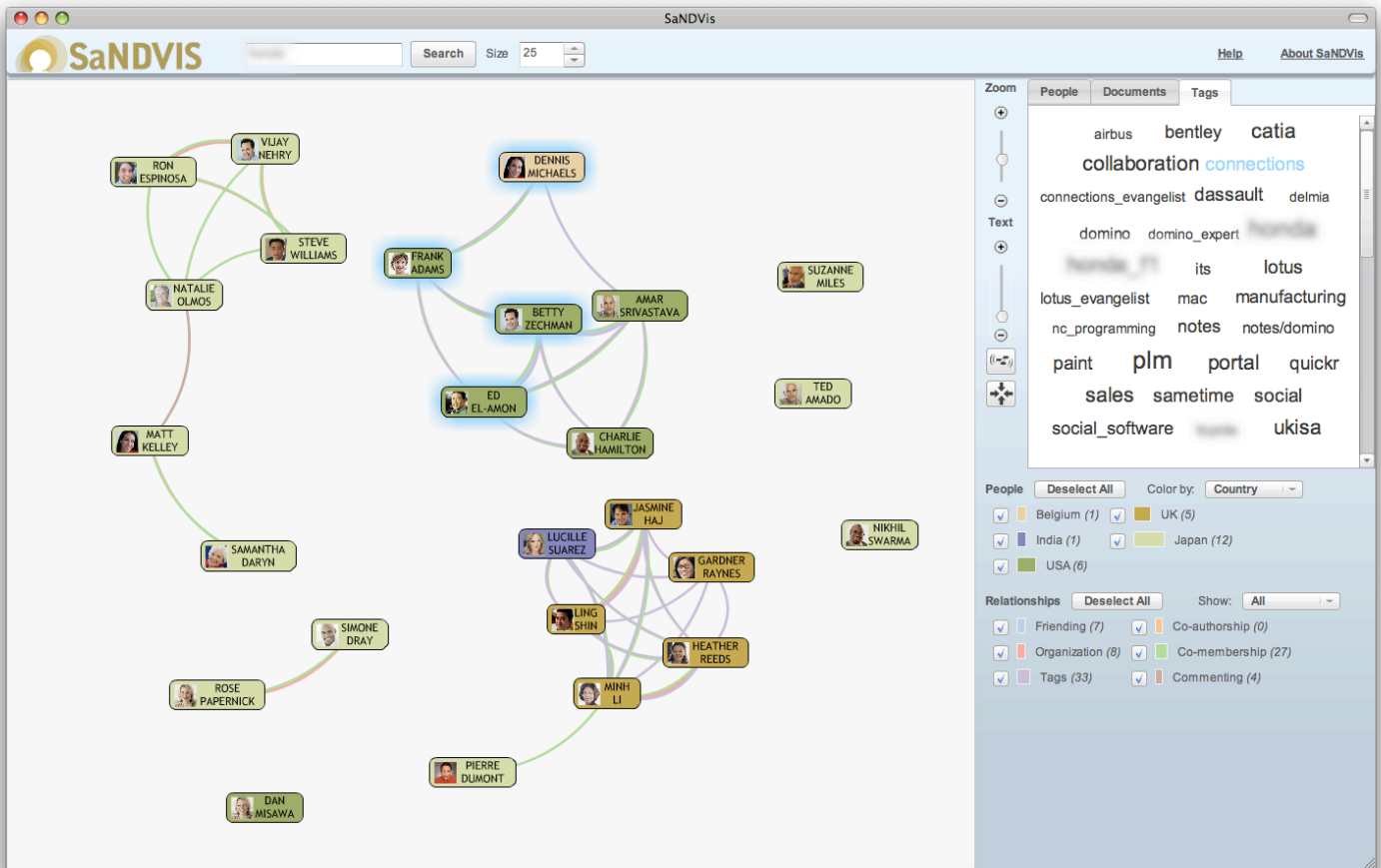
Figure 3. The SaNDVis UI contains three views relevant to relationship discovery task. A) On the left, a social graph view allows users to interpret the social position of the results of a topic query. B) On the top right, an evidence view allows users to examine the documents, tags, and people associated with the query. C) On the bottom right, a facet view allows users to get an overview of the categories with sparklines and allows users to filter out any irrelevant categories.

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 [15]. The number of nodes is, by default, kept to only 25 so the visualization is optimized to the design guidelines for achieving "NetViz Nirvana" [8], but users can increase this number.

5.2 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. As these three types of information are connected in analytics system, 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 an 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 3, users can mouse-over a tag (in this case, 'connections') and the people associated with that tag are highlighted in blue on the left. Con-

versely, 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.

5.3   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 sparkline and checkbox appear, as shown in Figure 3. The sparkline 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 depicted in Table 1. 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.

6   EVALUATION

SaNDVis has been deployed using viral promotion across a large global IT company. 1,790 users from 50 countries have performed 15,674 relationship discovery queries over a period of twelve months between March 2010 and March 2011.

In order to understand how SaNDVis was used, all queries were logged on the server. On a first visit to SaNDVis, users were asked to provide their user-id that was stored in a cookie.

6.1   Categories of Relationship Discovery Queries

As visual analytic tools are uncommon to many users, we were curious how SaNDVis would be utilized across the enterprise. We analyzed user queries by using two independent coders who hand-coded each of the 6,510 unique search terms to one of 8 categories. If a term was unfamiliar, the coders searched for the term on SaND to uncover more evidence. For any terms that were still ambiguous, we classified them in the 'Other' category (4.45% of all terms). As Figure 4 illustrates, the most popular scenario involved users analyzing a specific person's relationships across the enterprise (40.39%). This was followed by queries about relationships surrounding specific technologies or products (22.41%). Next were queries about internal projects or groups (14.41%), followed by queries containing more general terms (domains) (9.25%). The least popular queries were terms about companies (3.91%), locations or events (3.38%), and character traits or roles of people (1.78%). The large percentage of person queries intrigued us, and through follow-up interviews described
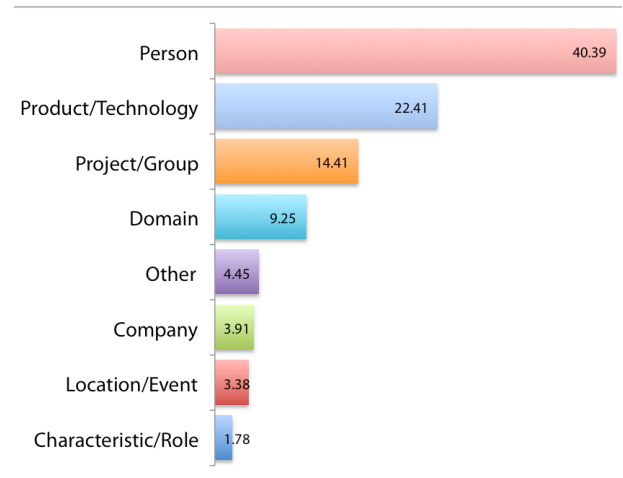

Figure 4. Percentages of the different categories of user queries using SaNDVis.

later in the paper we demonstrate the use cases for these popular queries.

7   USE CASES

While the patterns generated from the logs provided interesting clues about the behaviors of the users, we also sought to understand more specific scenarios of how they were utilizing the visual analytic tool for enterprise tasks. We solicited interviews via email from the same set of users analyzed in logs. Based upon their comments, we demonstrate three use cases of how they used SaNDVis.

7.1   Expertise Location and Team Building

One of the most common scenarios reported by users was the task of finding an expert. Expertise location is a relationship discovery task because users may not simply want an expert, but rather an expert who is well-positioned in the company. One such participant summarizes their experience:

"We needed someone with a background in Agile to present to our customer. SaNDVis results were right on the money, with the top person being the expert for Agile in [a large software company]. This gave us great leads in finding the right person to respond to our customer's needs."

The social graph displayed in Figure 5a is a network of the top Agile experts, who happen to span 12 countries. The participant filtered the social graph according to countries and tags of relevance and was then able to recommend an Agile expert for a meeting with his customers, based on the expert's central position. The participant also mentioned that the coordination of the people with documents was critical to his information discovery needs. He was able to track down various previously unknown web resources associated with the experts, including an internal Agile e-community and other resources for Agile development.

SaNDVis can also be used when users have a specific question that needs answered:

"Recently I wanted to know who used a [specific tool]. I searched for this term, and then used SaNDVis to know which people are close to me in the UK and then pinged a guy directly to get the help I needed! Else, I would have searched in google, post in an unknown forum, and wait for some random answers which I might have already tried myself."

A related scenario described in the interviews was the exercise of team-building. When assembling a team of experts, it is useful
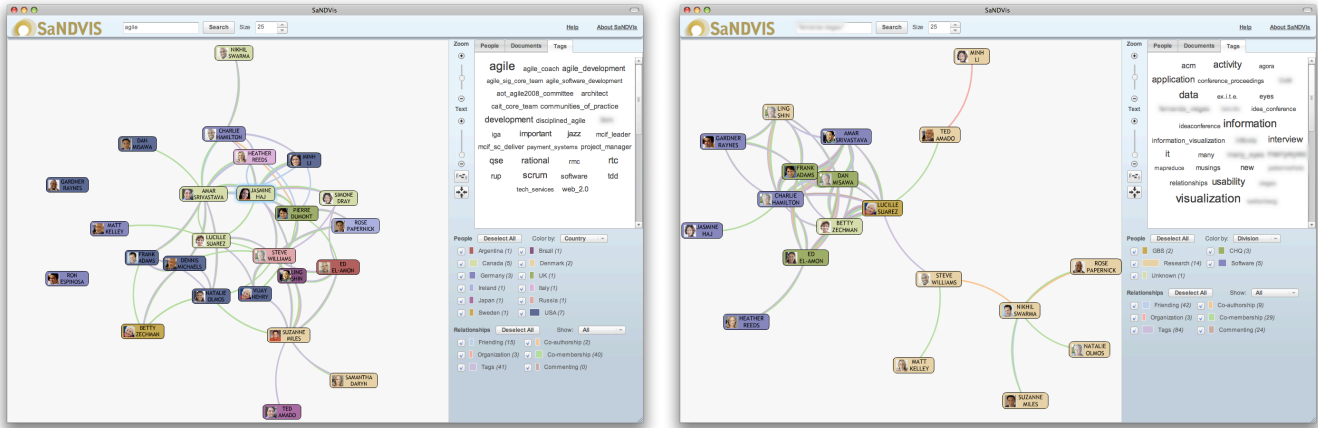
Figure 5. The SANDVis UI demonstrating two use cases. A) On the left, results from a relationship query seeking Agile experts across the enterprise, with nodes colored by their country. B) On the right, results from a relationship query about a specific person, anonymized as Lucille Suarez, with nodes colored by division in the company.

to know if their relationships suggest they have a history of successful collaboration. The team builder does not typically need the most experts, but instead people that can work well together. SaNDVis enables users to examine the relationships and choose experts that can collaborate:

"I used SaNDVis to find who was in charge of the customer account and who were the potential legal and sales contacts who would be able to help us set up the right conditions to start a collaboration."

Another participant adds about the value of visualizing social position in the results:

"SaNDVis is key when interviewing new team members - I can follow the chain to someone I know to get a trusted reference."

7.2 Team Coordination and Reflection

Relationship discovery tasks can also be useful in scenarios that involve analyzing relationships between people already familiar to the user, for instance when joining a team. This is particularly true for virtual teams, where many challenges exist and providing structural views can improve trust [25][33]. As one participant explained:

"I used [SaNDVis] to get more information on some [colleagues] I was meeting with. It was interesting to see what patents they were associated with, etc. It helped me to better understand the people I'm working with."

A related scenario involves trying to understand how familiar people fit into the ecosphere of a broader group of people who may be performing similar tasks. For instance, understanding the relationships between people associated with a similar customer. We illustrate this scenario by showing results from a user's query examining a customer in the automobile industry (Figure 3). From the visualization, it is clear that there are three clusters of people not interacting with each other. The Asian cluster, colored in light green, does not interact with the European cluster, colored in orange, and both do not interact with the American cluster, colored in dark green. This gives users a clue that relevant information may not be spreading between these groups, and the user can act on this knowledge to bridge new connections between these disparate groups. As another participant explains:

"I was looking into how the user experience people within [my new company] are connected. The interesting thing was that I

found only one person from my team (we have 20 people). The reason I came up with was that that person is very active [in the new company], and the rest of the team is still acting like they are working for [the acquired company]. What I mean is that social networking within [my new company] is not very common in our team, and I personally think that's a pity."

SaNDVis can be a powerful tool for management to understand how information is flowing in the company and how much people collaborate. Users can identify where other categories of expertise exist across the group. As one participant noted:

"Actually 3 years ago I started as a newbie with [a technology] and I would have loved to have SaNDVIS back then. Instead, I had to dig through blogs, and [corporate directories] to find the relevant persons, or establish a connection or see what they shared."

7.3 Personal Reflection

Finally, we illustrate the most common scenarios in our logs: discovery tasks with queries referring to terms related to the user. Often users would issue a query of their own name in SaNDVis. A participant acknowledges:

"It is very useful to understand what your network and extended network is doing and following … it has allowed me to find interesting and surprising information."

Figure 5b shows the results of a query for 'Lucille Suarez' (name anonymized for privacy). Lucille appears in the center of the social graph, which contains nodes colored according to the division of the person they represent. From the visualization, it is clear that Lucille acts a bridge between the research division (oranges nodes on the top center and bottom right) to a variety of other divisions (multi-colored nodes on the left). By understanding her unique position in the network, Lucille realizes her value in fostering collaboration between research and the rest of the company. However, she also notices that while she is a bridge, the number of her connections to research is small. To increase her social capital, she can connect with several of the other contacts to become an even greater asset for collaboration.

Many other participants also noticed strangers in their network visualized by SaNDVis, which then prompted them to serendipitously connect to them based on their similiarity. As one participant explains:

"My best experience is to find people co-tagging the same things as I do, suggesting we have the same interests. That gives a good hint for a connection that I didn't have so far."

7.4 Need for a People-Centric Approach

One of the common themes for all three use cases described above was the notion that SaNDVis puts people as the focus of the search results, rather than documents. A participant highlighted the value of a people-centric discovery tool rather than a document-centric tool:

"People author documents, so it makes sense to me to locate the people involved with a topic. You cannot collaborate with documents!"

As another participant stated:

"If you are looking for a technical solution, a focus on documents works great. If you are looking for people that are interested in an issue or subject, [focusing on] people is much better."

Yet another participant stated:

"Most often finding people who are experts in some area is a lot more beneficial than finding documents of that topic. Also seeing the relationships helps to find more relevant people and to identify the ones who are most central and most likely know a lot on the topic, or at least can point me to a person who could reply to my questions."

While online social networks like Facebook, LinkedIn, and Beehive offer some aspects of people-centric search, they often only support navigation and analysis of explicit ties. SaNDVis supports explicit and implicit ties aggregated from many data sources. Furthermore, SaNDVis provides access to tags and evidence for the ties, allowing users to understand exactly why people are connected, providing full transparency.

8 DISCUSSION AND LIMITATIONS

We have shown that there are a variety of compelling uses of a visual analytics tool for relationship discovery tool in the enterprise. However, as SaNDVis is a novel tool, some users probably did not utilize the tool for actual work-related questions during their initial explorations, and therefore may have simply inputted queries relating to themselves in order to evaluate the accuracy of the results.

Furthermore, the evidence displayed by SaNDVis is only based on social media use. Thus, there may be some relationship discovery scenarios that will be of limited success if certain key individuals are less active on social media. However, by being able to easily analyze one's traces of social information, additional people may be motivated to produce more social media information to ensure they become associated with topics related to their expertise and interests.

From our interviews, we learned that users do not always wish to use keyword queries. For instance, many users expressed interest in visualizing members of predefined groups such as members of an e-community, an activity, or a wiki. Furthermore, as a large portion of the queries involved specifying an individual person, enabling search for explicit groups may also turn out to be productive.

We also acknowledge that building a similar tool outside an enterprise firewall poses new challenges, such as mapping the multiple identities of users and dealing with massive amounts of people, facets, and relationship types.

Finally, while each of the three views in SaNDVis is arguably not a UI advancement on its own, we propose that the people-centric integration of these three views in a visual analytics tool is novel. The system makes over 73 million relationships manageable for users via social graph visualizations with smart, compre-hensible defaults coordinated with evidence views and filterable facets. With such a system, it is possible to utilize the vast relationships in enterprise social media to support common user tasks. Our interviews suggest that a key reason that SaNDVis is able to support such tasks is the ability for users to examine why relationships exist in the social graph representation as well as finding related documents. Such coordination to evidence is often missing from social network visualizations.

9 CONCLUSION

In this paper, we have demonstrated a visual analytics system that can support relationship discovery in the enterprise. By building an end-to-end system that can mine, aggregate, and assemble a social graph from documents, it is possible to support tasks including expertise location, team building, and personal reflection. These tasks are typically poorly supported by traditional information discovery interfaces. By integrating social position, evidence, and facets into a visual analytics interface, users can leverage existing social media behavior to assist them in their relationship discovery tasks. Our initial deployment of SaNDVis suggests that there is great value in providing these capabilities inside the enterprise.

10 REFERENCES

[1] Adar, E. GUESS: a language and interface for graph exploration. In ACM SIGCHI, pages 791–800, 2006.

[2] Amitay, E., Carmel, D., Har'el, N., Soffer, A., Golbandi, N., Ofek-Koifman, S., and Yogev, S. 2009. Social search and discovery using a unified approach. *Proc. HT '09*, 199-208.

[3] Bernstein, M., *Tan, D., Smi*th, G., Czerwinski, M., and Horvitz, E. 2009. Collabio: a game for annotating people within social networks. *Proc. UIST '09*, 97-100.

[4] Campbell, C. S., Maglio, P. P., Cozzi, A., and Dom, B. 2003. Expertise identification using email communications. *Proc. CIKM '03*, 528-531.

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

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

[7] 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

[8] Dunne, C. and Shneiderman, B. 2009. Improving graph drawing readability by incorporating readability metrics: a software tool for network analysts. *University of Maryland, Technical Report* HCIL-2009-13, May 2009.

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

[10] Farnh*am, S., P*ortnoy, W., Turski, A., Cheng, L., and Vronay, D. Personal Map: *Automatically m*odeling the u*ser's online social network. *Proc. INTERACT '03*.

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

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

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

[14] Freire, M., Plaisant, C., Shneiderman, B., and Golbeck, J. ManyNets: an interface for multiple network analysis and visualization. In *ACM CHI*, pages 213–222, 2010.

[15] Gansner, E., Koren, Y., and North, S. 2004. Graph drawing by stress majorization, *Proc. GD '04*, 239–250.

[16] Gephi. http://gephi.org/.

[17] Ghoniem, M., Fekete, J.D., and Castagliola, P. On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. Information Visualization, 4:114–135, July 2005.

[18] 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.

[19] Hearst, M.A. 2006. Design recommendations for hierarchical faceted search interfaces. *Proc. SIGIR '06 Workshop on Faceted Search*, 26–30.

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

[21] Henry, N. and Fekete, J.D. Visually exploring large social networks. In INTERACT, pages 604–610, 2007.

[22] Henry, N., Fekete, J.D., and McGuffin M. 2007. NodeTrix: a hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 1302-1309.

[23] Herman, I., Melancon, G., and Marshall, M.S. Graph visualization and navigation in information visualization: A survey. *IEEE Trans Vis and Comp Graphics*, 6:24–43, January 2000.

[24] 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*.

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

[26] Kang, H., Plaisant, C., Lee, B., and Bederson, B. B. NetLens: iterative exploration of content-actor network data. *Information Visualization*, 6:18–31, March 2007.

[27] Kautz, H., Selman, B., and Shah, M. 1997. Referral Web: combining social networks and collaborative filtering. *Commun. ACM* 40, 3 (Mar. 1997), 63-65.

[28] 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.

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

[30] McDonald, D. W. and Ackerman, M. S. 2000. Expertise recommender: a flexible recommendation system and architecture. *Proc. CSCW '00*, 231-240.

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

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

[33] Perer, A., Guy, I., Uziel, E., Ronen, I., and Jacovi, M. 2011. Unearthing People from the SaND: Relationship Discovery with Social Media in the Enterprise. To be presented as a poster at *AAAI Conference on Weblogs and Social Media*. ICWSM 2011.

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

[35] Pieper, J. H., Grace, J., and Dill, S. 2009. Team analytics: understanding teams in the global workplace. In *Proc. CHI '09*

[36] 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.

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

[38] 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.

[39] Shneiderman, B. and and Aris, A. Network visualization by semantic substrates. *IEEE Trans Vis and Comp Graphics*, 12:733–740, 2006.

[40] Smith, M. A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V. Dunne, C., Capone, T., Perer, A., and Gleave. E. Analyzing (social media) networks with NodeXL. In Communities and Technologies (C&T), pages 255–264, 2009.

[41] 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.

[42] Wattenberg, M. Visual exploration of multivariate graphs. In *ACM SIGCHI*, pages 811–819, 2006.

[43] Whittaker, S., Jones, Q., Nardi, B., Creech, M., Terveen, L., Isaacs, E., and Hainsworth, J. 2004. ContactMap: Organizing communication in a social desktop. *ACM Trans. Comput.-Hum. Interact.* 11, 4 (Dec. 2004), 445-471.

[44] 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*.

[45] Vivacqua, A. and Lieberman, H. 2000. Agents to assist in finding help. *Proc. CHI '00*, 65-72.