The Longitudinal Use of SaNDVis: Visual Social Network Analytics in the Enterprise

Adam Perer, Ido Guy, Erel Uziel, Inbal Ronen, and Michal Jacovi

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 provide details of a 22-month-long, large-scale deployment to over 2,300 users from which we analyze longitudinal usage patterns, classify types of visual analytics queries and users, and 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.

---

1 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 is 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 relationships.

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, 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, multidimensional 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

---

- *A. Perer is with the Healthcare Analytics Research Group, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598. E-mail: adam.perer@us.ibm.com.*
- *I. Guy, I. Ronen, and M. Jacovi are with the Social Technologies Group, IBM Research, Haifa, 31905, Israel. E-mail: {ido, inbal, jacovi}@il.ibm.com.*
- *E. Uziel is with Anywhere Software, Moshav Yodfat, 20180, Israel. E-mail: erel@basic4ppc.com.*

to which users can leverage the rich data collected by the other components for real enterprise tasks.

Another contribution of this work is the longitudinal analysis of SaNDVis over a period of 22 months with over 2,300 users. This paper extends the analysis of SaNDVis first described in [37] by analyzing 10 additional months of user data as well as several new approaches to analyzing the data, such as a deeper examination of when and what users were analyzing with the tool.

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.

2 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.

2.1 Social Matching and Expertise Location

Relationship discovery is related to various research areas previously studied. Terveen and McDonald [50] 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. [42] describe an expertise location system based on text analysis—user profiles are generated and matched by extracting keywords from related text documents. Vivacqua and Lieberman [55] present Expert Finder, a tool that matches novices and experts by analyzing Java documents related to them. Ehrlich and Shami [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 [27] 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 [33], [32] 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.

2.2 Visual Social Network Analytics

Visualizations have been used as aids for understanding social networks since the 1930s [13], using a variety of visualization techniques [25]. Typically, node-link diagrams, matrix visualizations [17], or hybrids [23], [24] are used to represent the linkage structure. Another alternative is to visualize networks according to attributes of the nodes using semantic substrates [46] or rolled-up PivotGraphs [51]. NetLens [28] 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], [37], [48].

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 [53] 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 [54] 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 [22]. 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 [24].

---

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

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 coauthorship graphs around domains such as "data mining" or "databases." SmallBlue [31] 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.

2.3 Longitudinal Analysis of Visual Analytics

There have been relatively few papers that study long-term usage patterns of visual analytic tools. Lam et al. provide a taxonomy of evaluation scenarios to understand and evaluate visual analytic systems [30]. Their survey notes that only 5 out of 850 recent visualization papers even consider the "long-term to assess the adoption of a visualization in a real usage scenario" [30].

There are several exceptions. Multidimensional in-depth long-term case studies (MILCs) [47] suggest multiple techniques for observing the usage of a visualization tool of a period of time, including observations, interviews, survey, and logging. Seo and Shneiderman used the MILCs methodology for a knowledge discovery tool over a period of 4 to 6 weeks with domain experts [44]. Perer and Shneiderman extended this approach by integrating a visual social network analysis tool into domain experts' workplaces for eight weeks and reporting on their results [39]. fSaraiya et al. conducted a longitudinal study using existing visual analytic tools with biologists to understand how each tool led to insights in microarray data over a three-month period [49]. To our knowledge, the user study in our paper is the first that analyses a visual analytics tool over such a long period of time (22 months) with thousands of users.

3 SOCIAL NETWORKS AND DISCOVERY (SaND)

3.1 Relationship Aggregation

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 nine types of relationships between two sets: [people, documents, tags] and [people, documents, tags]. This generalization allows the combination of social and

Fig. 1. Direct relationships modeled in SaND.

content analysis to optimally unleash the value of information unlocked throughout different enterprise sources.

Fig. 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 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 she/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 coauthors or comembers. 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 schemes 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 [2]. 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.

3.2 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 she/he was tagged with it by three other individuals; or that two individuals are related to each other because they coauthored two wikis and are comembers 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 [26] 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 [34] with 1.1M bookmarks by 68,000 users with 3M tags.
- A file sharing system [36] 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.

Fig. 2 illustrates SaND's social search UI, similar to the unified interface described by Amitay et al. [2]. 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 Fig. 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
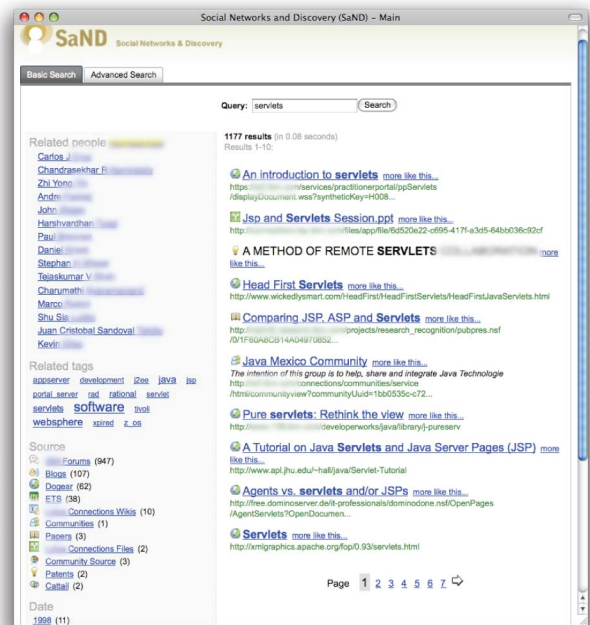
2. In a faceted search [21], the user can refine the results based on each of the categories (facets) chosen to characterize the data.

Fig. 2. SaND's social search UI.

relationships did not exist in [2]) and the other system only mined three data sources. Further information about SaND is described briefly in [43].

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.

3.3   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. 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 people 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, coauthorship, and comembership. 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.

TABLE 1
A Classification All of the People-to-People Relationships
Indexed by SaNDGraph

	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)

Familiarity relationships indicate the connected people probably know each other. Similarity relationships indicate the connected people behave similarly across social media.

4 SaNDVis: A System for Relationship Discovery

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

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.

4.1 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 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 [45]. 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 Fig. 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, 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.

4.2 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 an 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 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 Fig. 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. 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.
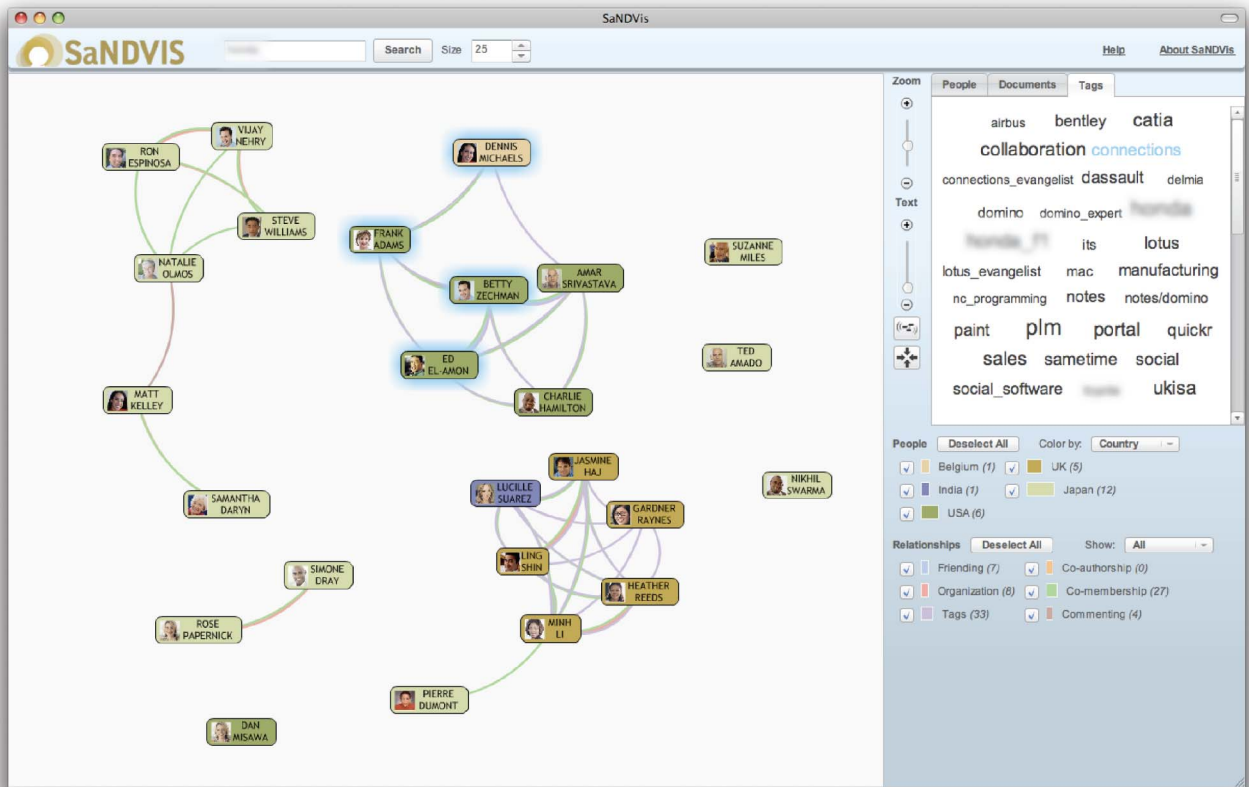
Fig. 3. The SaNDVis UI contains three views relevant to relationship discovery task. A) On the left, a social graph view allows users to interpret the social position of the results of a topic query. B) On the top right, an evidence view allows users to examine the documents, tags, and people associated 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.

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 coauthored, communities they are comembers 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.

4.3 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 Fig. 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.

The distinction between familiarity and similarity relationships has been extensively explored in our previous work [19], [20]. The two relationship types were shown to be useful for different scenarios: familiarity relationships were useful, for instance, for recommending people to

TABLE 2
A Summary of User Activities in SaNDVis

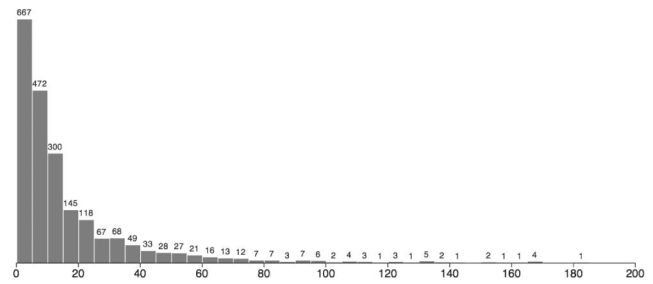| User Action	Total Count	Count Average per User Session
Topic Query (Search)	79,952	11.02
Mouseover Node	72,331	9.97
Click Node	24,206	3.38
Filter Edges by Facet Type	7,621	1.05
Filter Nodes by Facet Type	7,546	1.04
Tag Selection in Tag Cloud	4,856	0.67
Person Selection in People Table	3,674	0.51
Document Selection in Document Table	1,092	0.15
Change Node Facet Category	889	0.12
Click Edge	694	0.10
Change Edge Facet Category	669	0.09

Fig. 4. Histogram of users binned by frequency of searches in SaNDVis. There were 13 users who had performed more than 200 searches who are not depicted in this histogram due to the long tail distribution.

query (averaging over 11 queries per user session). Since there were more queries than any other type of user interaction in the SaNDVis interface, this suggests that users were often able to interpret the visualization without any need to interact with the visualization. We speculate that some of this may be due to users need to research immediately, such as noticing a typo in their search term or getting an empty result set. We also note that the high number of queries compared to other user interactions might imply that in some cases users were disappointed with the query's outcome to a degree that they did not bother to further explore it.

Although queries were the most popular activity among users, the distribution of queries per users has a long tail. Fig. 4 shows a histogram of all 2,341 SaNDVis users based on their total number of queries across all 22 months of deployment. The largest number of users (667, or 28.5 percent) performed five queries or less. As the histogram highlights, the number of users drops significantly as the number of queries increases. The distribution had an extremely long tail, so the 13 users who performed over 200 queries are not depicted in Fig. 4.

The next most common tasks involve gaining a better understanding of the people in the graph visualization. Users frequently inspected people in the social graph visualization by hovering their mouse over a node to get more information about them (9.97 hovers per session). They also regularly clicked on a person to make their selection persistent when inspecting other aspects of the visualization and evidence (3.38 clicks per session). However, edges were rarely clicked (0.10 clicks per session). We believe this may be due to the difficulty in clicking on thin edges in the visualization and suggests an enhanced selection technique may be needed.

To a lesser extent, users also felt the need to filter the social graph visualization to either make the visualization more readable or to restrict the results to specific facets of interest. Users used filters on edges (1.05 per session) and filters on nodes (1.04 person) almost equally. However, users rarely chose to recolor the social graph to other available facet categories of nodes or edges (0.10 and 0.09 per session, respectively).

Users interacted with the evidence views less than the facet or social graph views. Tags in the tag cloud were selected the most (0.67 per session), then people in the person table (0.51 per session), followed by documents in the document table (0.15 per session). We hypothesize the document and people tables were used less due to the fact

connect with on an SNS, finding a social path to another individual, or building a buddylist of people to exchange instant messaging with [18], [20]; similarity relationships were used to recommend strangers, to apply collaborative filtering for content recommendation, or for building global diverse teams [19].

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 relayout at any time.

5 EVALUATION

SaNDVis has been deployed using viral promotion across a large global IT company. 2,341 users from 51 countries have performed 79,952 relationship discovery queries over a period of 22 months between March 2010 and December 2011.

5.1 Usage of SaNDVis Features

To understand how SaNDVis was used, all user actions 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. Each unique visit by user was also given a unique session key, so we could track actions by session. There were a total of 7,253 unique user sessions spanning SaNDVis's deployment.

A summary of user activities in the tool is shown in Table 2. The most common activity is performing a topic

Fig. 5. Percentages of the different categories of user queries using SaNDVis.

TABLE 3
The Presence of Users and People with Whom They are Familiar in Search Results

User Action	Initial Deployment (first 3 months)	Entire Deployment (all 22 months)
User in Search Results	35.90%	29.49%
User Familiar with a Person in Search Results	41.40%	28.14%

that their tabs were not selected by default and potentially went unnoticed by some users.

5.2 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 four independent coders who hand-coded each of the 8,601 unique search terms to descriptive categories. If a term was unfamiliar, the coders searched for the term on SaND to uncover more evidence. Of these, 1,432 terms (16.6 percent) were still unable to be deciphered and were removed from the analysis.

Each of the remaining 7,169 terms were assigned to one of eight categories. As Fig. 5 illustrates, the most popular scenario involved users analyzing a specific person's relationships across the enterprise (63.5 percent). This was followed by queries about relationships surrounding specific technologies or products (12.3 percent). Next were queries about internal projects or groups (11.0 percent), followed by queries containing more general terms (domains) (3.8 percent). The least popular queries were terms about companies (3.3 percent), locations or events (2.7 percent), and character traits or roles of people (2.0 percent). Other terms that did not fit in above categories were classified as "Other" (1.4 percent of all terms).

In general, it is apparent from these results that SaNDVis is used mainly to search for specific queries—be it a person, a technology, a product, or a group. The tool is used much less frequently to search for more general subjects or themes, such as "visualization" or "social media." The huge percentage of queries using a person's name indicates that users often prefer to seed the visualization using a specific person. This may be due to the fact that the main visualization in SaNDVis centers on people and their relationships, thus leading users to more naturally starting their search using a person, who would then often appear in the center of the main visualization.

The overwhelming percentage of person queries intrigued us, and through follow-up interviews described later in the paper we demonstrate the use cases for these popular queries.

5.3 Exploration versus Reflection

In addition to inspecting the types of queries, we set out to explore whether users were querying for new topics or for topics they were already familiar with. Since we could unify users' searches with the enterprise activities mined by SaND, it was possible to compare actions on SaNDVis to users' past behavior.

The main question we sought to answer was whether users were examining known topics to monitor existing relationships or seeking new topics to discover new people. Our first method of calculating this was to see if the user was present in the social graph returned by their query. When looking at data from the initial three months of SaNDVis's deployment, we found that in 35.9 percent of the queries, the user was present in the corresponding social graph displayed by SaNDVis. This suggests that for over a third of queries in SaNDVis, users were analyzing people or topics they were already to some extent familiar with. However, we expected that when SaNDVis was new, 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 to evaluate the accuracy of the results. Indeed, if we analyze the entire duration of SaNDVis's deployment instead, as shown in Table 3, this percentage drops to 29.74 percent. We believe this highlights a change in behavior as their experience with SaNDVis matures and the novelty effect wears off.

As a second technique, we examined whether the query results included people the user is *familiar* with. Familiarity was calculated based upon a weighted aggregation of all of the familiarity relationships listed in Table 1, in a similar manner to [18]. If two people share enough familiarity relationships, we classify them as familiar. This calculation allowed us to determine if the user was examining a topic that their peers were familiar with. Note that the user him/herself is not counted as a familiar person. Our method for calculating familiarity relationships has shown effective in previous work [18]. While traditional SNA typically takes into account one signal for familiarity (e.g., connectivity within an SNS), our method is more comprehensive and captures a larger portion of a person's network of familiarity.

When examining the first three months of SaNDVis logs, 41.4 percent of the queries resulted in people with whom the user was already familiar with in the search results. This percentage drops remarkably to 28.14 percent when examining all 22 months of log data. These results indicate
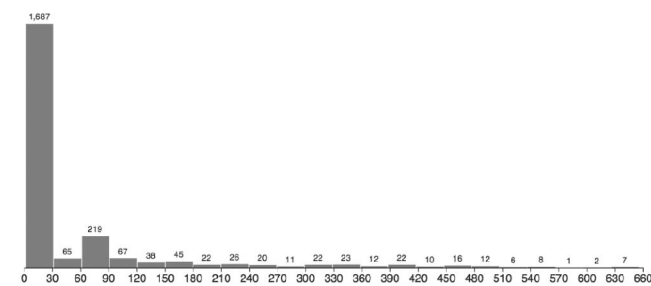
Fig. 6. Histogram of users binned by duration of the number of days they used SaNDVis.

TABLE 4
Population Details of the 172 SaNDVis Users Who Have Used the Tool for over Two Months and Performed At Least 30 Queries in the Tool

Type of SaNDVis User	Number of Users
Frequent, Exploratory User	55
Frequent, Monitoring User	23
Infrequent User	94

that in early stages, more users used SaNDVis to search for familiar topics, for which the results included more people whom they know. It could very well be that they applied such queries to evaluate SaNDVis and better understand how it works and the information it presents. At later stages, more users became more adventurous asking about topics that are less related to them or to their network, which included substantially less people they know. These kinds of queries would potentially teach the users more about people they do not know, their relationship to a topic of interest and the intrarelationships among them.

5.4 Longitudinal Behavior of Users

Since our analysis suggests that users change their use of SaNDVis over time, we decide to investigate this further by examining users who repeatedly used SaNDVis over a lengthy period of time. Unsurprisingly, most users (1,687, or 72.06 percent) used SaNDVis for one month or less. However, this still left a large population of users who chose to return to SaNDVis after a lengthy period of time. Fig. 6 shows a histogram of all users based on their duration of use of SaNDVis, measured in days and binned by months. As the histogram highlights, the number of users drops significantly as duration increases.

To better understand the dominant SaNDVis users, we focused our analysis on users whose use not only spanned a lengthy time, but also performed numerous topic queries. We decided to perform our analysis on the 172 users who have performed at least 30 queries that span at least 60 days over our 22-month corpus of log data.

Over these 172 users, several dominant user types became apparent as shown in Table 4. There were 78 users that exhibited patterns of frequent, regular use of SaNDVis. Of those 78, there were two types of users. Fifty-five were classified as *exploratory* users, who frequently used SaND-Vis and typically performed queries for new topics. The other 23 users were classified as *monitoring* users, as although they used SaNDVis frequently, they typically repeated topic queries from their past sessions. The logs suggest they were interested in seeing how the social structure around their queries of interest changed over time. The remaining 94 users were classified as *infrequent* users. These users typically used SaNDVis only two or three times, and while they had several intense sessions of querying, they did not return to the tool on a consistent basis.

Users were classified by inspecting their usage with an interactive, temporal visualization that displayed their query count over time. In the visualization, users were represented as a horizontal line with circles that represent the number of queries they performed on a given day. The text of the users search queries is visible by moving the mouse over a particular circle. The visualization used is shown in Fig. 7, however due to space limitations, this visualization here is filtered to show only 75 users, each of whom had activity that spanned at least eight months. In this figure, we also color SaNDVis users according to their classification type as defined in Table 4 (exploratory, monitoring, or infrequent).

By analyzing the longitudinal patterns of users, it became clear that SaNDVis was not a repeatedly used tool for most users. While the log data suggests that SaNDVis provided value to our top users, the logs unfortunately do not explain why some users did not come back. SaNDvis was never conventionally marketed inside the enterprise, but instead virally spread from some of it's top users.

Similar to NameVoyager [51], there was a great deal of discussion of our visualization tool in internal blogs and discussion forums. For instance, when two independent, high-profile fans commented "this is exactly the kind of intuitive information presentation that we need" or "I don't know how you guys do it. [SaNDVis has] even tags that relate to my work. Fantastic!" the tool received a surge of new users in June 2010. Some fans even created their own tutorial videos to assist new users in learning SaNDVis. While this viral promotion certainly generated a lot of initial uses, the lack of sustained promotion by the authors, such as a mailing list or reminder emails, may have hindered continued use.

While SaNDVis was never designed with the monitoring use case in mind, the presence of such users in the log data suggests the need to better support them, such as providing regular updates on how relationships are changing with respect to their queries. Such a strategy would improve their user experience in finding critical information, as well as driving return visits to the tool.

Finally, SaNDVis was also never linked to from enterprise's central authoritative homepage of tools where most employees visit for their internal resources. The inability to integrate better in the wider ecosystem of tools limited adoption as well.

These lessons suggest some larger challenges for adoption that are often not discussed by visualization tool builders. Successful research projects like NameVoyager [51] received 900,000 downloads in their first few months of
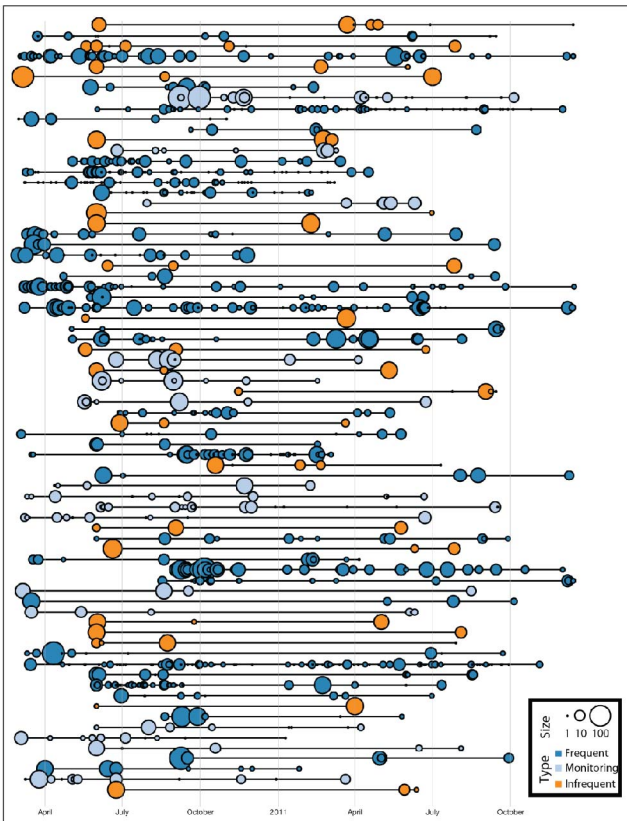
Fig. 7. Temporal visualization of queries performed by 75 SaNDVis users that used the tool for at least eight months. Each line in the visualization represents a SaNDVis user whereas each circle represents the query count for a given day. SaNDVis users are colored according to their classification type.

deployment, but there was no analysis on how and if users came back over time.

6   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.

6.1   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 Fig. 8a 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 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:
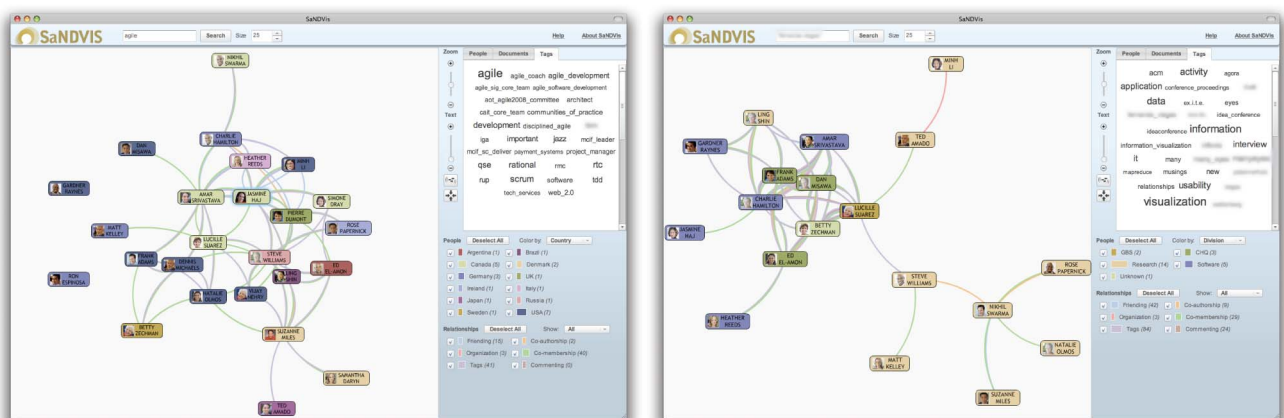

Fig. 8. The SANDVis UI demonstrating two use cases. A) On the left, results from a relationship query seeking Agile experts across the enterprise, with nodes colored by their country. B) On the right, results from a relationship query about a specific person, anonymized as Lucille Suarez, with nodes colored by division in the company.

"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."

6.2 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 [27], [37]. 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 (Fig. 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 three 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."

6.3 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."

Fig. 8b 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 (multicolored 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 similarity. As one participant explains:

"My best experience is to find people cotagging 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."

6.4 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.

7 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 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.

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, comprehensible 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.

8   DIRECTIONS FOR FUTURE WORK

Currently SaNDVis is used inside the organization and allows employees to explore networks related to certain topics. The growing importance of online communities in the enterprise [35] poses more opportunities for the use of SaNDVis as a tool to visualize the members of a community and the relationships among them. Such visualization can be useful both for the community owners and for all the members, to better understand the intrarelationships within the community, the diversity in terms of country or division, or the other main topics of interests of its members. It can help owners build the community in a healthy way and members to judge whom they already know in the community or whom they should get to know. The use of SaNDVis in a community-based setting can also help us further explore the use of such a visualization tool over time and whether it remains handy for owners and members as the community evolves.

Another potential extension of SaNDVis we plan to further explore is its use to explore potential paths between two individuals. While as of now SaNDVis is used to visualize the relationships among a group of individuals, a potential enhancement would allow to focus on all weighted paths between two individuals, explore their edge weights, evidence, and other characteristics, to ultimately infer the optimal path or better understand the potential connection between two strangers in the enterprise. Another extension would allow the user to augment the visualization by brining him or herself into the graph and see how they connect to the group of individuals, using the techniques of [40]

Aggregation of network data can also be relevant outside the enterprise. With the popularity of social media tools, a variety of signals can be used to obtain a comprehensive view on the overall social graph as reflected online. Our future plans include experimentation with SaNDVis outside the enterprise, for visualizing networks that are based on a rich set of relationships. While enterprise-specific person attributes, such as division, may no longer be relevant outside the firewall, other attributes, such as age, gender, or employer, can substitute. Taking SaNDVis outside the firewall will also require dealing with challenges such as multiple identities per individual and large scale of the data.

Finally, while the patterns of longitudinal use have been analyzed in this paper, more evidence is necessary to understand what impeded long-term adoption for most users. Over the 22-month deployment, the authors passively released the tool and did not send any reminders to users about the tool. Future experiments could be designed to see if marketing the tool more effectively would increase adoption and satisfaction with the tool.

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 long-term deployment of SaNDVis suggests that there is great value in providing these capabilities inside the enterprise. While our interviews and analysis suggest that SaNDVis was a useful tool, it was only used longitudinally by a subset of our users. This suggests that building a useful tool is not always enough, but instead effectively integrating it into user's ecosystem of tools is also critical for continued adoption.

REFERENCES

[1] E. Adar, "GUESS: A Language and Interface for Graph Exploration," *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems,* pp. 791-800, 2006.

[2] E. Amitay, D. Carmel, N. Har'el, A. Soffer, N. Golbandi, S. Ofek-Koifman, and S. Yogev, "Social Search and Discovery Using a Unified Approach," *Proc. 20th ACM Conf. Hypertext and Hypermedia (HT '09),* pp. 199-208, 2009.

[3] M. Bernstein, D. Tan, G. Smith, M. Czerwinski, and E. Horvitz, "Collabio: A Game for Annotating People within Social Networks," *Proc. 22nd Ann. ACM Symp. User Interface Software and Technology (UIST '09),* pp. 97-100, 2009.

[4] C.S. Campbell, P.P. Maglio, A. Cozzi, and B. Dom, "Expertise Identification Using Email Communications," *Proc. 12th Int'l Conf. Information and Knowledge Management (CIKM '03),* pp. 528-531, 2003.

[5] D. Chen, J. Tang, J. Li, and L. Zhou, "Discovering the Staring People from Social Networks," *Proc. 18th Int'l Conf. World Wide Web (WWW '09)*, pp. 1219-1220, 2009.

[6] C. Danis and D. Singer, "A Wiki Instance in the Enterprise: Opportunities, Concerns and Reality," *Proc. ACM Conf. Computer Supported Cooperative Work (CSCW '08)*, pp. 495-504, 2008.

[7] J. DiMicco, D.R. Millen, W. Geyer, C. Dugan, B. Brownholtz, and M.J. Muller, "Motivations for Social Networking at Work," *Proc. ACM Conf. Computer Supported Cooperative Work (CSCW '08)*, pp. 711-720, 2008.

[8] C. Dunne and B. Shneiderman, "Improving Graph Drawing Readability by Incorporating Readability Metrics: A Software Tool for Network Analysts," Technical Report HCIL-2009-13, Univ. of Maryland, May 2009.

[9] K. Ehrlich and N. Shami, "Searching for Expertise," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '08)*, pp. 1093-1096, 2008.

[10] S. Farnham, W. Portnoy, A. Turski, L. Cheng, and D. Vronay, "Personal Map: Automatically Modeling the user's Online Social Network," *Proc. Int'l Conf. Human-Computer Interaction (INTERACT '03)*, 2003.

[11] S. Farrell, T. Lau, S. Nusser, E. Wilcox, and M. Muller, "Socially Augmenting Employee Profiles with People-Tagging," *Proc. 20th Ann. ACM Symp. User Interface Software and Technology (UIST '07)*, pp. 91-100, 2007.

[12] D. Fisher and P. Dourish, "Social and Temporal Structures in Everyday Collaboration," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '04)*, pp. 551-558, 2004.

[13] L. Freeman, "Visualizing Social Networks," *J. Social Structure*, vol. 1, 2000.

[14] M. Freire, C. Plaisant, B. Shneiderman, and J. Golbeck, "ManyNets: An Interface for Multiple Network Analysis and Visualization," *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (CHI)*, pp. 213-222, 2010.

[15] E. Gansner, Y. Koren, and S. North, "Graph Drawing by Stress Majorization," *Proc. 12th Int'l Conf. Graph Drawing (GD '04)*, pp. 239-250. 2004.

[16] Gephi, http://gephi.org/, 2012.

[17] M. Ghoniem, J.D. Fekete, and P. Castagliola, "On the Readability of Graphs Using Node-Link and Matrix-Based Representations: A Controlled Experiment and Statistical Analysis," *Information Visualization*, vol. 4, pp. 114-135, July 2005.

[18] I. Guy, M. Jacovi, N. Meshulam, I. Ronen, and E. Shahar, "Public vs. Private: Comparing Public Social Network Information with Email," *Proc. ACM Conf. Computer Supported Cooperative Work (CSCW '08)*, pp. 393-402, 2008.

[19] I. Guy, M. Jacovi, A. Perer, I. Ronen, and E. Uziel, "Same Places, Same Things, Same People?: Mining User Similarity on Social Media," *Proc. ACM Conf. Computer Supported Cooperative Work (CSCW '10)*, pp. 41-50, 2010.

[20] I. Guy, M. Jacovi, E. Shahar, N. Meshulam, V. Soroka, and S. Farrell, "Harvesting with SONAR: The Value of Aggregating Social Network Information," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '08)*, pp. 1017-1026, 2008.

[21] M.A. Hearst, "Design Recommendations for Hierarchical Faceted Search Interfaces," *Proc. SIGIR Workshop Faceted Search*, pp. 26-30, 2006.

[22] J. Heer and D. Boyd, "Vizster: Visualizing Online Social Networks," *Proc. IEEE Symp. Information Visualization (InfoVis '05)*, 2005.

[23] N. Henry and J.D. Fekete, "Visually Exploring Large Social Networks," *Proc. Int'l Conf. Human-Computer Interaction (INTER-ACT)*, pp. 604-610, 2007.

[24] N. Henry, J.D. Fekete, and M. McGuffin, "NodeTrix: A Hybrid Visualization of Social Networks," *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1302-1309, Nov./Dec. 2007.

[25] I. Herman, G. Melancon, and M.S. Marshall, "Graph Visualization and Navigation in Information Visualization: A Survey," *IEEE Trans. Visualization and Comp Graphics*, vol. 6, no. 1, pp. 24-43, Jan. 2000.

[26] J. Huh, L. Jones, T. Erickson, W.A. Kellogg, R.K. Bellamy, and J.C. Thomas, "BlogCentral: The Role of Internal Blogs at Work," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '07)*, 2007.

[27] S.L. Jarvenpaa and D.E. Leidner, "Communication and Trust in global Virtual Teams," *Organization Science*, vol. 10, no. 6, pp. 791-815, 1999.

[28] H. Kang, C. Plaisant, B. Lee, and B.B. Bederson, "NetLens:Iterative Exploration of Content-Actor Network Data," *Information Visualization*, vol. 6, pp. 18-31, Mar. 2007.

[29] H. Kautz, B. Selman, and M. Shah, "Referral Web: Combining Social Networks and Collaborative Filtering," *Comm. ACM*, vol. 3, pp. 63-65, Mar. 1997.

[30] H. Lam, Bertini, E. Isenberg, P. Plaisant, C. Carpendale, and S. Empirical, "Studies in Information Visualization: Seven Scenarios," *IEEE Trans. Visualization and Computer Graphics*, vol. 18, no. 9, pp. 1520-1536, Nov. 2012.

[31] C. Lin, K. Ehrlich, V. Griffiths-Fisher, and C. Desforges, "SmallBlue: People Mining for Expertise Search," *IEEE Multi-Media 15*, pp. 78-84, Jan. 2008.

[32] D.W. McDonald, "Evaluating Expertise Recommendations," *Proc. Int'l ACM SIGGROUP Conf. Supporting Group Work (GROUP '01)*, pp. 214-223, 2001.

[33] D.W. McDonald and M.S. Ackerman, "Expertise Recommender: A Flexible Recommendation System and Architecture," *Proc. ACM Conf. Computer Supported Cooperative Work (CSCW '00)*, pp. 231-240, 2000.

[34] D.R. Millen, J. Feinberg, and B. Kerr, "Dogear: Social Bookmarking in the Enterprise," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '06)*, pp. 111-120, 2006.

[35] M.J. Muller, K. Ehrlich, T. Matthews, A. Perer, I. Ronen, and I. Guy, "Diversity Among Enterprise Online Communities: Collaborating, Teaming, and Innovating through Social Media," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '12)*, pp. 2815-2824, 2012.

[36] M.J. Muller, D.R. Millen, and J. Feinberg, "Information Curators in an Enterprise File-Sharing Service," *Proc. 11th European Conf. Computer Supported Cooperative Work (ECSCW '09)*, pp. 403-410, 2009.

[37] A. Perer, I. Guy, E. Uziel, I. Ronen, and M. Jacovi, "Visual Social Network Analytics for Relationship Discovery in the Enterprise," *Proc. IEEE Conf. Visual Analytics Science and Technology*, 2011.

[38] A. Perer and B. Shneiderman, "Balancing Systematic and Flexible Exploration of Social Networks," *IEEE Trans. Visualization and Computer Graphics*, vol. 12, no. 5, pp. 693-700, Sept. 2006.

[39] A. Perer and B. Shneiderman, "Integrating Statistics and Visualization: Case Studies of Gaining Clarity during Exploratory Data Analysis," *Proc. ACM Conf. Human Factors in Computing Systems (CHI)*, 2008.

[40] A. Perer and F. van Ham, "Integrating Querying and Browsing in Partial Graph Visualizations," IBM Technical Report 12-01, 2012.

[41] J.H. Pieper, J. Grace, and S. Dill, "Team Analytics: Understanding Teams in the Global Workplace," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '09)*, 2009.

[42] T. Reichling, K. Schubert, and V. Wulf, "Matching Human Actors Based on their Texts: Design and Evaluation of an Instance of the Expertfinding Framework," *Proc. Int'l ACM SIGGROUP Conf. Supporting Group Work (GROUP '05)*, pp. 61-70, 2005.

[43] I. Ronen et al., "Social Networks and Discovery in the Enterprise (SaND)," *Proc. 32nd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '09)*, 2009.

[44] J. Seo and B. Shneiderman, "Knowledge Discovery in High Dimensional Data: Case Studies and a User Survey for the Rank-by-Feature Framework," *IEEE Trans. Visualization and Computer Graphics*, vol. 12, no. 3, pp. 311-322, May 2006.

[45] N.S. Shami, K. Ehrlich, G. Gay, and J.T. Hancock, "Making Sense of Strangers' Expertise from Signals in Digital Artifacts," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '09)*, pp. 69-78, 2009.

[46] B. Shneiderman and A. Aris, "Network Visualization by Semantic Substrates," *IEEE Trans Visualization and Comp Graphics*, vol. 12, no. 5, pp. 733-740, Sept./Oct. 2006.

[47] B. Shneiderman and C. Plaisant, "Strategies for Evaluating Information Visualization Tools: Multi-Dimensional In-Depth Long-Term Case Studies," *Proc. AVI Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*, 2006.

[48] M.A. Smith, B. Shneiderman, N. Milic-Frayling, E. Mendes Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, and E. Gleave, "Analyzing (Social Media) Networks with NodeXL," *Proc. Fourth Int'l Conf. Communities and Technologies (C&T)*, pp. 255-264, 2009.

[49] P. Saraiya, C. North, V. Lam, and K.A. Duca, "An Insight-Based Longitudinal Study of Visual Analytics," *Trans. Visualization and Computer Graphics,* vol. 12, no. 6, pp. 1511-1522, 2006.

[50] L. Terveen and D.W. McDonald, "Social Matching: A Framework and Research Agenda," *ACM Trans. Computer-Human Interaction,* vol. 12, no. 3, pp. 401-434, 2005.

[51] M. Wattenberg, "Baby Names, Visualization, and Social Data Analysis," *Proc. IEEE Symp. Information Visualization,* pp. 1-7, 2005.

[52] M. Wattenberg, "Visual Exploration of Multivariate Graphs," *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems,* pp. 811-819, 2006.

[53] S. Whittaker, Q. Jones, B. Nardi, M. Creech, L. Terveen, E. Isaacs, and J. Hainsworth, "ContactMap: Organizing Communication in a Social Desktop," *ACM Trans. Computer-Human Interaction,* vol. 11, no. 4, pp. 445-471. 2004.

[54] F.B. Viegas, D. Boyd, D.H. Nguyen, J. Potter, and J. Donath, "Digital Artifacts for Remembering and Storytelling: Posthistory and Social Network Fragments," *Proc. 37th Ann. Hawaii Int'l Conf. System Sciences (HICSS '04),* 2004.

[55] A. Vivacqua and H. Lieberman, "Agents to Assist in Finding Help," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '00),* pp. 65-72, 2000.

Adam Perer received BS degree in computer science from Case Western Reserve University, in 2002, and the MS and PhD degrees in computer science from the University of Maryland, College Park, in 2005 and 2008, respectively. He is currently a research scientist in the Healthcare Analytics Research Group at IBM T.J. Watson Research Center in New York. He was previously a research scientist in IBM Research's Visual Communication Lab in Cambridge and IBM Research's Social Technologies Group in Haifa. His research focuses on making sense of big data through visualization and interactive analytics. He is a coauthor of more than 20 conference and journal papers in the fields of information visualization, visual analytics, and human-computer interaction. He serves on the organizing and program committees of leading visualization conferences, including InfoVis and EuroVis.

Ido Guy received the BSc degree in software engineering and the MSc degree in computer science both Summa Cum Laude from the Technion, Israel Institute of Technology, in 2005 and 2007, respectively. He currently manages the Social Technologies group, which he established and started to manage in 2007 as part of IBM Haifa Research lab. The group's current activity focuses on the areas of Social Analytics, Social Business Processes, and Crowd Computing. Prior to this, he led and contributed to various projects around collaboration technologies and social network analysis at the IBM Haifa Research lab. He is a coauthor of more than 35 conference and journal papers in the fields of data mining, human-computer interaction, computer supported cooperative work, information and knowledge management, information retrieval, and recommender systems. His research work mostly focuses on social media, especially on social network mining and analysis and social recommender systems. He serves on the organizing and program committees of leading conferences including CHI, WWW, and CIKM. He is currently serving as program cochair for the 2012 ACM Conference on Recommender Systems.

Erel Uziel received the BSc degree in computer science from the Technion, Israel Institute of Technology, in 2009. He was a member of the Social Technologies group in IBM Research Haifa from 2008-2010. He is currently the founder and CEO of Anywhere Software.

Inbal Ronen is received the BSc degree in mathematics and computer science and the MSc degree in computer science from the Hebrew University of Jerusalem. She has been a researcher at the IBM research Lab at Haifa, Israel, for the last 10 years. Previously, she was employed at various startups. She is a member of the Social Technologies group focusing on areas such as social analytics, social recommendations, and online community analysis. She coauthored about 17 papers and several patents.

Michal Jacovi received the MSc degree in computer science from the Technion, Israel, in 1993, when she joined IBM. She belongs to the Social Technologies group in IBM Research Haifa. Her research interests include HCI, CSCW, and social media. She has worked on visualization and awareness on the web, peer support and community building, e-Learning, and various social media research projects around user experience and engagement. She has published more than 20 research papers and filed more than 10 patents.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.