

SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence

Ching-Yung Lin^{#1}, Nan Cao^{*2}, Shi Xia Liu^{*3}, Spiros Papadimitriou^{#4}, Jimeng Sun^{#5}, Xifeng Yan^{#6}

[#]IBM T.J. Watson Research Center
19 Skyline Dr, Hawthorne, NY, USA

¹chingyung@us.ibm.com

⁴spapadim@us.ibm.com

⁵jimeng@us.ibm.com

⁶xifengyan@us.ibm.com

^{*}IBM China Research Lab
8 Dongbeiwang West Road, Beijing, PRC

²caonan@cn.ibm.com

³liusx@cn.ibm.com

Abstract— SmallBlue is a social networking application that unlocks the valuable business intelligence of 'who knows what?', 'who knows whom?' and 'who knows what about whom' within an organization, without requiring explicit involvement of individuals. The aim of SmallBlue is to locate knowledgeable colleagues, communities, and knowledge networks in companies. The suite also helps users manage their personal networks, and reach out to their extended network (the friends of their friends) to find and access expertise and information.

I. INTRODUCTION

Individuals and groups in large companies take on informal roles that can improve innovation effectiveness. Finding the hidden connections in any organization helps people to work together and share social resources to achieve common goals. People rely on a personal network of friends and colleagues to get trusted information; help filter and interpret information; and get referrals to other people. Though personal networks are invaluable for getting quick answers, they aren't always sufficiently large or diverse to reach everyone directly who has the right information. But these informal social networks within formal organizations are a major factor affecting companies' performance [1].

The SmallBlue project [5,6] within IBM (also part of the Atlas software suite) aims to analyze the social network within a corporation, to provide answers to the important questions of "who knows what?" and "who knows whom?". It relies on a small client application that users consent to install on their personal machines, which locally analyzes their email outbox and their outgoing IM chat transcripts. Content is stripped of quoted response and then reduced to stemmed term vectors, before being sent to the central repository for analysis. Thus, we can build growing corpus of corporate emails and chat transcripts, which is similar to a live, dynamically evolving version of the Enron corpus (but with the necessary privacy policies and safeguards in place). As of this writing, the number of IBMers that have installed the software is close to 5,000, and growing fast due to network effects (about 2,000

new subscribers since the beginning of the year). The number of distinct recipients is approaching 350,000 (i.e., close to the entire workforce of IBM). The number of emails and chat transcripts is approximately eight million. The number of keyword stems in the raw dataset, across all languages, also approaches eight million.

SmallBlue extended the capability of our two prior works: ExpertiseNets and CommunityNet. ExpertiseNets are dynamic graphs that model the relational and evolutionary expertise for mining, retrieval, and visualization [2]. CommunityNet integrates automatic email content and social network analysis to find out how experts collaborate and to recommend experts to users [3].

II. SYSTEM OVERVIEW

As a social network tool, SmallBlue is dependant on people contributing information. The more people who contribute, the better and more useful the tools will be. Unlike other social network tools SmallBlue has been designed to minimise the time burden on the practitioner, it automates contribution leaving you more time to use to the tools to help you collaborate.

SmallBlue is made up of a client and a web application. The client is used to analyze the user's *outgoing* email and instant message transcripts to determine the social network and infer expertise. The web application consists of several components.

A. Data Acquisition

Most expertise-locator systems acquire data either by having individuals fill out profile information or by extracting information from existing sources. Those sources could be public, such as coauthored documents; patents; or user-generated blogs, wikis, and social-tagging systems. Private sources, such as email, chat, and calendar entries, can contribute semantic information as well as social network data. In deciding which data source to use for SmallBlue, we ruled

out user-authored profiles because it would have taken too long to build a critical mass of information and would not have included the social network information we needed. In addition, these data sources run into problems when users don't update their information frequently enough.

Private data sources, such as email logs, have the advantage of containing rich information from which data about what one knows and whom one knows can be derived. These sources also address the following issues. **(i) Coverage:** Email use is widespread, so data can be collected from everyone not just the people who have authored documents or other data. **(ii) Maintainability:** New email is constantly being generated. **(iii) Ease of use:** People already use email, so other than asking users for permission to use their data, there is no additional work required by the user.

The disadvantage of using private data is that capturing and using the data might violate privacy. Although there is some evidence suggesting that the majority of users are willing to share some private data under the right circumstances [4], persuading people to provide private data voluntarily is challenging. If privacy issues aren't adequately addressed, users will stop using an expertise-locator system, opt out of volunteering their data, and generate negative word of mouth, all of which would severely impact any ability to have sufficient people in the system to deliver useful search results.

Because of the sensitivity of the data, we developed a strong set of policies that restricted what data we collected, how we used the data, and what information was available to users. In particular, we relied on aggregated and inferred information, which prevented any user from ever seeing a direct relationship between any person in SmallBlue. In other words, we never kept or displayed any information about who communicated with whom about what.

These mechanisms make a strong statement in protecting users' privacy and helped earn their confidence in using SmallBlue.

B. Expertise Search Applications

The first set of applications on the SmallBlue website are related to locating experts on given topics (determined by a set of keywords entered by users in a search box), finding how to reach them based on the personal interactions among individuals, and also analysing the entire communities of people that are highly related to the given topic. In short, the applications here are related to presenting the connections among employees and their areas of expertise.

C. Social Search Applications

The next step forward involves a suite of applications that leverages the connections among employees to make other applications smarter. To that end, we are incorporating social proximity information and combining it with social bookmarking and web search. This allows us to make recommendations of interesting web pages (or, in general, any web-accessible content within the corporate intranet), based on the web pages bookmarked or sought by one's cohorts.

III. PERSONAL SOCIAL CAPITAL

SmallBlue Ego provides a visualization of a user's personal social network as well as the social capital of each contact. Social capital is defined as the number of new contact that a user can be introduced to through their direct contact - including the geographical and organizational distribution of these new contacts. This visualization is private and personalized for each user. This information is only known to the user herself.

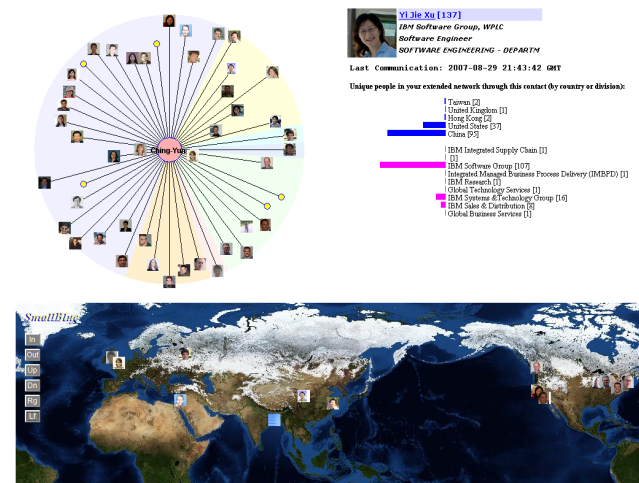


Figure 1. SmallBlue Ego. The circular spider view shows the inferred personal social network of Ching-Yung Lin in August 2007. Different colors represent different business divisions in a company. The largest four portions are research, software, services, and headquarters, respectively. The list next to the spider list shows the social capital of Xu from Lin's viewpoint. The locations of people in Lin's ego network are shown on the map.

On the basis of the aggregated large-scale social network data, the system rates how valuable a person is to the user, in terms of unique social connections. We call this rating the personal social capital. Figure 1a shows Ching-Yun Lin's personal social network. In Figure 1b we see that Yi Jie Xu is Lin's bridge to 137 unique people, including 93 people in China, two in Taiwan, and so on. Figure 1c shows the business division distribution; it indicates Xu can help linking 107 unique software group people and one researcher for Lin.

IV. EXPERTISE SEARCH

Existing expertise-locator systems have focused on search algorithms or methods of acquiring searchable data.⁸ But searching for people is not like searching for documents. People operate in an organizational and social context, which circumscribes what they know and how they know it.

A. People search

The SmallBlue Find search engine, which returns a relevance-ranked list of people, interprets a search string and maps it onto related keywords. The search engine aggregates the results for all the keywords and ranks them according to relevance weighting and aggregated social-network structure. SmallBlue then generates a list of the top 1,000 people who best match the search terms. The list is displayed with each person's picture along with his or her job title, role, and online

status. By default, SmallBlue shows the top matched experts in the entire company. Users can search experts within a business division, country, community, group, or specific social distance. In addition, there is an interface that lets users set several predefined ranges in tabbed pages that facilitate seeing top experts among various search ranges through only one search submission.

Consider the case of someone who has been asked by her or his manager to gather information about Second Life, the 3D virtual environment. In addition to looking at online sources and perhaps consulting internal blogs or social-tagging systems if they exist within the company, the person might also use want to find who in the company might be an expert in Second Life. Entering “second life” into the search window would bring up the display shown in Figure 2.



Figure 2. SmallBlue Find. The list shows the top-10 experts in the company. For each person, it shows division, job title and job role. Moreover, it shows the shortest social path from the user to reach that person.

B. Personalized social path and expertise context

To increase the likelihood of the searcher contacting someone new and receiving a response from that person, we made the social paths to each person visible as a form of six degrees of separation. That is, we display the minimal number of intermediate people to contact to reach the person. Paths of three or fewer are shown in the initial relevance ranked display with the line, “Ask <person>” or “Ask <person1> -> <person2>.” The example of the Find page in Figure 2 shows the social distance (degrees of separation) from the seeker to each of the people on the page, up to three degrees. Of the 10 people displayed in Figure 2, this searcher is directly connected to one of them, shown as “my collaborator or contact,” and indirectly connected to the other nine people, either by two degrees, for example “ask Vicky” or three degrees. Users can also apply a filter to the results so that they only see those people who are within one, two, or three degrees of separation. The default is set to view all people.

Expertise locator systems should provide users with information to help them reach their own determination of the appropriateness of a candidate’s expertise, and their likely responsiveness to being approached. SmallBlue provides this information in two ways. Information about social distance appears in the Reach tool within the system. From the Find page, users can click on a name or picture to navigate to the Reach page (Figure 3), which displays current public information about the selected person, including a list of the shortest paths and up to 16 alternate paths that link the seeker to the expert. In Reach pages, shortest paths are provided up to six degrees.



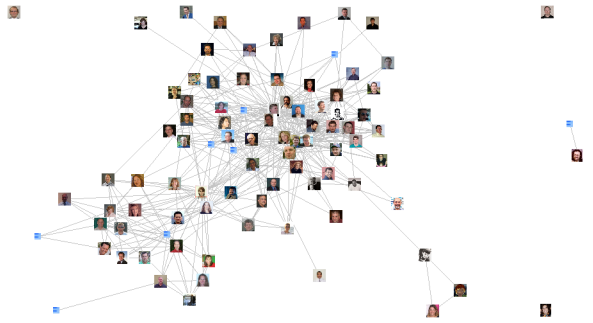
Figure 3. SmallBlue Reach. The left-hand side shows this person’s community information. The middle part shows the shortest social paths from the user to reach another person up to six degrees. The right-hand part shows the personal activities, for example, posting on blogs, forums, social bookmarks of Web pages, and so on. We can see the self-described expertise of the target person.

C. Social network analysis

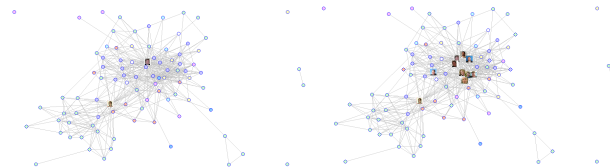
SmallBlue Net displays and analyzes the social network of top experts associated with a topic. Although we think of an expert as someone who has the most knowledge of a particular topic, sometimes we want to find the person who knows the expert rather than the expert him or herself. Or we might want to find the emergent community associated with a topic to see the structure of the community and find who is a hub at the center of the community or who is a connector between communities. For these reasons, it’s useful to see the whole social network. Using our earlier “second life” search example, Figure 4 displays the social network of the same 100 people who were returned.

V. SOCIAL SEARCH

This section briefly describes functionality that is under current development, although we have fully functional prototypes of all applications described here.



(a) Clustering of social network of experts of 'Second Life' in IBM



(b) Key bridges in this network (c) Key hubs in this network



(d) Location and network of experts of 'Second Life' in IBM United States

Figure 4. SmallBlue Net. (a) Clustering of social network of experts on *Second Life* at IBM. (b) Key bridges in this social network. (c) Key hubs in this social network. (d) Location and social network of *Second Life* experts at IBM US.

A. Whisper

SmallBlue Whisper aggregates social network information with Dogear, which is IBM's internal version of social bookmarking applications, such as del.icio.us. Quite often, one's peers are interested in similar topics and may bookmark sites that are also of interest to others in their social cycle. For example, when a group has to complete a task, access to

certain resources may be required. These resources may be hard to find, but once a few members of the group locate them, this information is potentially useful to the other group members.

Whisper presents the user with a list of items bookmarked by others. These items are ranked by how often they have been marked by other people, as well as the social proximity of those people to the user.

B. Web Search

In addition to leveraging social bookmarking information, we are also currently building a web search front end, using the web APIs provided by a popular search engine. This allows us to create mashups, whereby search results are augmented with annotations that have been added by one's peers. Furthermore, it allows us to collect web search clickthrough data, in order to improve the relevance of search results and recommendations.

VI. SUMMARY

SmallBlue (aka. IBM Atlas) is a powerful social network analysis platform, developed within IBM. It is also undergoing development as a product offering. It allows corporations to leverage the power of informal social interactions, in order to locate experts on given topics, find ways to reach them and, eventually, empower every web application with this knowledge.

VII. REFERENCES

- [1] H. Raider and D. Krackhardt, "Intraorganizational Networks," in J. Baum (ed.) *Companion to Organizations* pp. 58-74, Oxford, UK:Blackwell.
- [2] X. Song, X et al., "ExpertiseNet: Relational and Evolutionary Expert Modeling," *Intl. Conf. User Modeling*, Springer, 2005, pp. 24-30.
- [3] X. Song et al., "Modeling and Predicting Personal Information Dissemination Behavior," *Proc. 11th ACM SIGKDD Int'l Conf. Knowledge Discovery in Data Mining*, ACM Press, 2005, pp. 479-488.
- [4] M.S. Ackerman, L.F. Cranor, and J. Reagle, "Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences," *Proc. 1st ACM Conf. Electronic Commerce*, ACM Press, 1999.
- [5] K. Ehrlich, C.-Y. Lin, and V. Griffiths-Fisher. Searching for experts in the enterprise: Combining text and social network analysis. In *GROUP*, 2007.
- [6] Ching-Yung Lin, Kate Ehrlich, Vicky Griffiths-Fisher, and Christopher Desforges, *Smallblue: People Mining for Expertise Search*, IEEE Multimedia, 2008.