SmallBlue: Unlock Collective Intelligence from Information Flows in Social Networks

Dashun Wang
Northeastern University
110 Forsyth Street, Boston, MA 02115

Zhen Wen, Ching-Yung Lin
IBM T. J. Watson Research Center
19 Skyline Drive, Hawthorne, NY 10532

Abstract:
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. Moreover, organizations depend on their members finding and sharing tacit knowledge to be informed, to avoid duplication of effort, and to innovate. Finding the hidden connections in any organization and understanding information flows helps people to work together and share social resources to achieve goals.

This paper describes the SmallBlue system, a social sensing, mining and visualization system designed to unlock this valuable collective intelligence without explicit human involvement. It has been collecting more than 20 millions of email and instant messaging communications, including the time stamps, to whom they communicate with, the subjects, the content statistics, from 10,000+ volunteers in 76 countries for more than 3 years. Besides, more than 2 million of social software data (including bookmarking, file sharing, wiki, blog, etc.) and knowledge and learning activities for more than 30,000 employees were collected. These data have been used for inferring the dynamic social networks and expertise of 400,000 employees. Moreover, we are also collecting the financial performance of 100,000 consultants, including the details of the projects and the billable hours of each individual. After anonymizing the data, these abundant datasets empower our study to examine social capital, human capital, and financial capital simultaneously (http://smallblue.research.ibm.com).

In this paper, we first describe the design of SmallBlue and how it is being used to help people in large organizations to find experts, expand and visualize their social capital. Next, we present SmallBlue’s scientific advances on understanding information diffusion processes. In particular, we focus on two questions regarding to email forwarding: (1) how do the underlying social networks affect the information spreading; and (2) how does the information spreading depend on organizational structures.

Using the SmallBlue datasets, our study quantitatively evaluates multiple factors’ impact on email forwarding processes, including tie strength between people, and organizational structures. The results show the varied influences of these factors on whether and how soon emails are forwarded. We believe that such improved understanding of the information forwarding processes would provide significant insights towards better collective intelligence mining.

Introduction
Individuals and groups in large organizations take on informal roles that can improve operation effectiveness. Take enterprises as an example, business leaders are most interested in establishing a supportive culture and climate to help their companies innovate, including developing new products, services, and markets; creating new business models; or improving existing operations. CEOs care whether they can fully use the hidden potential inside companies—the knowledge in employees’ minds and the relationships that employees have with one another. With a way to find expertise, a knowledgeable colleague might be just one phone call away from helping solve a complex problem. Finding the hidden connections in any organization helps people to work together and share social resources to achieve common goals. SmallBlue, is a people-mining system designed to unlock this valuable business intelligence without explicit human involvement.
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. According to an analyst report, employees get 50 to 75 percent of their information directly from other people [1]. Moreover, companies depend on their employees finding and sharing tacit knowledge to be informed, to avoid duplication of effort, and to innovate. 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 [2]. A tool such as SmallBlue could enable employees to capitalize on networking behaviors, leverage social knowledge and artifact sharing, and enable the business to benefit from the knowledge and experience of practitioners more effectively.

SmallBlue system is a social sensing, mining and visualization system designed to unlock this valuable collective intelligence without explicit human involvement. It achieves this goal by collecting and analyzing 20 million of electronic communication records including email, instant messages, and calendar meetings. These data sources have the advantage of containing rich information from which data about what one knows and whom one knows can be derived. For example, the information diffusion among people (e.g., email forwarding) indicates the tie strength among people. In addition, we are also collecting the financial performance of 100,000 consultants, including the details of the projects and the billable hours of each individual. Such performance data enable large scale study of the relationship between social networks and people’s productivity [7].

This paper is organized as follows: first, we describe the rationale and design of the SmallBlue system. Next, we present SmallBlue’s scientific advances on understanding information diffusion processes. In particular, we focus on email forwarding. Unlike information diffusion in online communities or Twitter [6], email forwarding is a unique process to foster cross organization collaboration, as well as building relationship. For example, by forwarding an email, senders can increase their social capital by demonstrating knowledge of the recipients’ interests or emphasize a connection between sender and recipient. However, it may also damage senders’ social capital if the forwarded content is wrong or harmful [5].

SmallBlue System

The SmallBlue suite 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. Here we provide a high-level overview of SmallBlue. More details can be found in [3, 4]. The SmallBlue suite consists of five components: the SmallBlue Client and four Web-based user applications:

- **SmallBlue Client** is a social-sensing software that resides on a registered user’s machine to capture privacy-protected data for social network and expertise inference. It periodically updates new social activities and extracts features from the captured data.

- **SmallBlue Ego** is a personal social capital management tool. It automatically creates a visualization of a personal social network and shows a friend’s social value by demonstrating what types of people to which this friend can connect (Figure 1a).

- **SmallBlue Find** is a search engine that ranks people according to desired knowledge or skills. (Figure 1b)

- **SmallBlue Reach** is a network-analysis engine that shows users their shortest social paths to reach a person. It also shows the formal organization groups and informal groups to which a person belongs as well as a person’s public activities in blogs, forums, social bookmarks, profiles, and so on (Figure 1c).

- **SmallBlue Net** is a large-scale social network visualization and analysis tool. For a given topic search, it shows the links among experts, and can find alternate experts and identify key influencers and brokers for that particular topic. For a given group search, it visualizes the set of people who have common interests. In addition, it can cluster people to find how they interact, and show the informal group structure (Figure 1d).
**SmallBlue Synergy** is a social network based personalized search tool. Given queries of a user, Synergy leverages the user's social network to re-rank search results (Figure 1e). For example, if certain search results were accessed by the user’s social neighbors, Synergy will promote their ranking.

**SmallBlue Whisper** is a social network based content recommendation tool. For a given user, content accessed by his/her friends may be useful for him/her as well. Therefore, Whisper would recommend such content to the user (Figure 1f).

SmallBlue is based on people’s electronic communication records such as email, which 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:

- **Coverage.** Email use is widespread, so data can be collected from everyone not just the people who have authored documents or other data.
- **Maintainability.** New email is constantly being generated.
- **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.

### Information Forwarding

SmallBlue not only enables users and their organizations to be more effective, it also collects valuable data for conducting scientific research. Here we present a study on how a specific piece of information spreads. In particular, we focus on email forwarding.

As our data set is biased towards the communications within the enterprise, the temporal patterns of the communications are restricted by the working schedules. We show in Figure 2 the number of communications in each hour of the week. We observe a clearly periodic pattern in our data. The load of communication builds up in the morning and decays in the afternoon with a notable dip at noon indicating the lunch time. There are two interesting points we want to make here. Firstly, the forwarding activities are significantly higher than the normal email traffic in the mornings on workdays, especially on Monday, and lower in the afternoon especially on Fridays.
This is a good indicator that forwarded emails are timely and important, representing a special class of the overall email traffic. Secondly, the accessibility to emails is limited by the weekly schedules. This weekly cycle becomes important when we inspect the efficiency of information spreading in the following sections. For example, a delay of two days in the delivery of information, which was received on Friday, could be due to the inability to access the emails during the weekends. Therefore, for any calculation regarding time in the following sections, we also did a check where we remove the off-hours. Yet the results did not change qualitatively.

Our analyses focus on the most fundamental building block of the information spreading process – information pathways, as illustrated in Figure 3. More specifically, user A sent an email to user B at a certain time with a specific title. Then user B waited for some time and passed along the information to user C by forwarding this email.

Factors Impacting Information Forwarding

**Underlying social network.** Indeed, to whom one would pass the information is directly restricted by who he connects to in the social network, and how well he is connected with his social neighbors. Understanding the interplay between the social network and the information spreading process will promote new strategies in various aspects such as viral marketing by taking into account the effect of the network topology. We start by building a social network among our users by aggregating the email communications over a year period of time. We add a link between two users if there has been at least one email between the two. The weight of the link is defined as the number of emails between two nodes. As we are interested in the connectivity between individuals, we focus on the static picture of the network as a whole instead of the dynamics of the network evolutions.

We show in Figure 4, the probability of email forwarding conditional on the weight of the links between user A and B, and between B and C respectively, in the information pathways. We observe the information more likely to spread initially via weak ties and then get passed through strong connections, an evidence of routing information by user B choosing social neighbors of different closeness.

**Organization Structure.** In an enterprise, understanding how information flows within and between different departments and organizational levels is of great importance in various aspects from building a better collaborative environment to controlling information securities. We will
examine impact of organizational context in two directions. One is the influence of departmental restrictions, and the other one is the organizational levels.

In Figure 5, we show the median time delay for different roles of brokerage. There are in total five roles of brokers. We show the illustrative examples for all of them. Each circle represents a department, and users are in the same circle if they are from the same department. Our data set has individuals from as many as 19 departments, and the information pathways which consist of people from different departments are classified into these 5 categories. We observe that the information flows significantly faster in two cases – coordinator and gatekeeper, than other three cases. These are the only two cases where B and C are in the same department. This tells us the bottleneck of the information flow in the departmental context is to get the information out of the department. We further break down the manager and non-manager cases for each role of brokers. We find that the managers are better as a representative while non-managers are better as a liaison. In Figure 4, we show that the probability of email forwarding conditional on the weight of the links AB and BC in the information pathways as shown in Figure 3. We observe the information more likely to spread initially via weak ties and then get passed through strong connections, an evidence of routing information by user B choosing social neighbors of different closeness.

Reference

[6]. Ravi Kumar, Mohammad Mahdian, Mary McGlohon: Dynamics of conversations. ACM Conference on Knowledge Discovery and Data Mining (KDD) page: 553-562, 2010.