International Journal of Scientific & Engineering Research Volume 9, Issue 4, April-2018
ISSN 2229-5518

Social Network Analysis

[1]MohammedAsifRaza, [2]Pradeep Kumar K

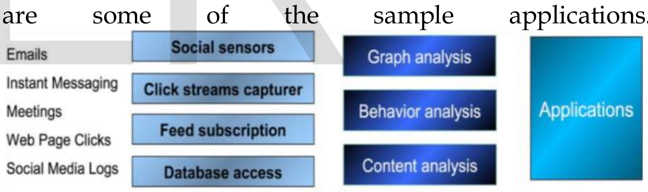
Abstract— Social network analysis (SNA) has been a research focus in multiple disciplines for decades, including sociology, healthcare, business management, etc. Traditional SNA researches concern more human and social science aspects trying to undermine the real relationship of people and the impacts of these relationships. While online social networks have become popular in recent years, social media analysis, especially from the viewpoint of computer scientists, is usually limited to the aspects of people's behavior on specific websites. Thus are considered not necessarily related to the day-to-day people's behavior and relationships. We conduct research to bridge the gap between social scientists and computer scientists by exploring the multifactor existing social networks in organizations that provide better insights on how people interact with each other in their professional life. We describe a comprehensive study on the challenges and solutions of mining and analyzing existing social networks in enterprise. Several aspects are considered, including system issues; privacy laws; the economic value of social networks; people's behavior modeling including channel, culture, and social inference; social network visualization in large-scale organization; and graph query and mining. The study is based on an SNA tool (Small Blue) that was designed to overcome practical challenges and is based on the data collected in a global organization of more than 400 000 employees in more than 100 countries.

1 Introduction

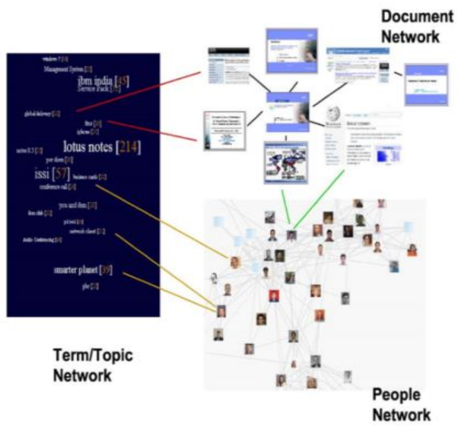
In recent years, we have witnessed a drastic uptick in the growth of information. With the recent advance of social media and the growing use of social networking tools, organizations are increasingly interested in understanding how individuals, teams, and organizations harvest value from their social networks. As estimated in 2006, the amount of digital information created, captured, and replicated is 161 billion GB, about three million times the information in all the books ever 0018-9219/$31.00 2012 IEEE in online social media is its stronger interest in finding the Actual social networks and productivity and security impacts rather than the finding networks. Drawing from the field of economic sociology, social network researchers are written. Thus, the simultaneous explosion of social media, knowledge management, and networking tools is not a mere coincidence, as these technologies have played an important role in sharing and disseminating the vast amount of information recently created. However, before formulating network strategies on how one leverages social networks to achieve superior outcomes, it is crucial to understand how and why networks create advantages. It should be also noted that a major difference of social network analysis (SNA) in en-have long predicted that certain network positions are more advantageous than others. One particular network that has perceived a tremendous amount of attention is structural holes. Actors spanning multiple structural holes are theorized to have more information and control advantage than their peers. For example, bankers with structurally diverse networks are more likely to be recognized as top performers.

2 Data Acquisition and Privacy Issues

Fig. 1(a) describes the fundamental structure of our system. We implemented several methods to collect various aspects of people's activities in enterprise, including: 1) social sensors; 2) click stream capture; 3) feed subscriptions, and 4) database access. Then, we conducted three types of analysis: graph, behavior, and the legislation varying in each of the 27 member states. Fig. 2 shows the current status of privacy laws worldwide. In an organizational setting, other factors related to employment legislation also had to be considered. The employer/employee relationship can compromise the ability to gain free and informed consent of participants in some countries. There are strict limitations around, and in some countries (e.g., Germany and Austria), prohibitions on employee monitoring at work. Together these mean that social software features that present few or no issues in an Internet setting can present significant issues in an enterprise setting. It is order to make the social network mining system a practical and content semantics. Various applications such as expertise search, people and content recommendation, social search, social path access, etc., are some of the sample applications.



(a)



(b)

3 Data Acquisition

Social sensors are based on a distributed front-end analysis mechanism that is installed in individual volunteer's machines. Its usage is twofold. First, this mechanism can distribute the computational workload by placing first level of data gathering and feature extractions. Second, this is an important mechanism for privacy compliance. In several countries, it is illegal to conduct data analysis within the communication channel. Communication providers cannot process data for the

- [1]Mohammed Asif Raza, 2nd –BCA, VIT University.
- [2]Pradeep Kumar K., 2nd –BCA, VIT University.

purpose other than providing communication services. Social sensors solve this legal issue by processing the copy of the data that are stored in an individual's computer, instead of gathering data from communication servers. This mechanism can resolve several legal challenges. Furthermore, via distributed sensors, our system can distribute the first level of feature extraction functions of content analysis such as stop word removal, stemming, one-gram and bi-gram statistics, etc., in an individual's machine to avoid the liability of storing the original communication content in a centralized server. Many features were designed to protect the human rights on privacy and free speech.

4 Privacy Laws

Privacy is a fundamental human right, as described in the United Nations Universal Declaration of Human Rights in 1948. Article 12 specifies: BNo one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such inference or attacks.[A fundamental element of privacy is data privacy, the ability to control one's personal information (PI), where PI is defined as any information that relates to a living individual who can be identified from that data, or from that data plus other information which is in possession of, or is likely to come in possession of the data controller.

5 Value of social network

SmallBlue allows us to track how individuals' networks evolve over time. To evaluate the performance implications of social networks, we also obtained the performance. Design metrics of these individuals. The longitudinal nature of the analyses enables us to explore the potential causal linkage between social networks and performance and observe the micro mechanisms of how networks drive productivity. The detailed recording of electronic communication archives also helps reducing the potential biases derived from using surveys and self-reports.

and the content of these messages is also encoded and stored in archives. The system has a perfect memory of all the electronic communication records. Social networks derived from such data are thus rarely subject to memory errors or recall biases that often mire the validity of survey instruments in earlier network studies. However, social networks instantiated using electronic communications are also not always a perfect representation of a person's overall network.

6 Network Effects on Personal Revenues

To leverage the longitudinal nature of our network data, we created a panel of networks using both three- and six-month intervals with a sliding window of one month. We matched these time- varying network data with consultants' performance as measured by billable revenue. We also gathered information about these consultants such as their gender, division, hierarchy within the firm, seniority, job role as well as the type of work and the industry these consultants typically work for. These factors serve as the control for our econometric analysis to eliminate confounding factors such as more senior consultants are more likely to generate more billable revenue. We leveraged both random- effect and fixed-effect econometric models to eliminate many confounding factors that are unobservable in our data, such as personality traits or inherent abilities.

7 Network Effects on Projects

We explored the implication of structural holes at the project level where each node in the network represents a project and each link in the network represents the communication instances exchanged between the two projects forming the link. Similar to the findings at the individual level, project networks that span structural holes are associated with positive increases in a project revenue, after controlling for the total number and the type of people in each project, temporal and regional shocks such as business cycle at various regions, and the line of business the project is in. We also employed random and fixed- effect specifications to eliminate other time invariant factors.

8 Impact of Social Networking Tool

If social networking tools can facilitate the process of finding the right resources that are critical to the task at hand, they could have tremendous implications for organizations, especially for creating strategies on how to invest and use these technologies to improve firms' bottom line. We studied 2038 anonymized global business consultants for two years. In Fig. 4, we plot the relationship between individual work performance as measured by billable revenue and the number of months since the adoption of SmallBlue. We controlled for factors including temporal shocks, individual characteristics such as job roles and hierarchies within the organizations, and the characteristics of each project such as the line of business and the region when the project was initiated.

Fig 4: Normalized average monthly

The X-axis labels the number of months since a person has adopted SmallBlue; the zero value indicates a person has just adopted SmallBlue; negative values indicate the number of months before a person has adopted SmallBlue; and positive values indicate the number of months after a person has adopted SmallBlue. The Y-axis indicates the extra revenue generated in each month since the adoption. As indicated in the graph, the billable revenue of a person gradually increases after the person has adopted SmallBlue. Prior to the adoption, the coefficient estimates range between \$2300 and \$3300 in monthly billable revenue. These results show that networking tools can play a critical role in facilitating individuals to locate resources and expertise within the firm.

9 Channel, culture, and social inference

Another aspect of the study is to understand how employees behave in terms of the channel, culture, and influence perspectives. In, Yang et al. showed the culture being the most significant factor in shaping perception and behavior via a survey of near 1000 people from four countries in an organization. For instance, Chinese and Indian users are more likely to use online social network tools for Q&A, in comparison to the U.S. and U.K. users. Culture is particularly important for a large organization that operates the channel, culture and inference.

10 VISUALIZATION

Network visualization is the most intuitive way to bring network analytic results to general users. There is no one fit-all network visualization design, as each user may have a rather specified information need. A famous site VisualComplexity.com has accumulated hundreds of network visualizations over diverse network data and scenario. For quite many cases, the standard node-link representations are employed; for others, the matrix visualization is also applied, mainly for networks with dense connections so as to increase the network readability.

11 SOCIAL NETWORK VISUALIZATION

In SmallBlue Net, we depict the fundamental social network of people in enterprise in the traditional nodelink form, as shown in Fig. 10. The major challenge here is how to deal with the potential network size up to half a million nodes. We take a straightforward approach: to filter the network into a smaller size, so that the network layouts can be computed in a reasonable time and the graph complexity can be controlled on a readable scale. We introduce a two-stage filtering strategy for the visualization, including a first- stage in-disk query and a second- stage in-memory filtering. In the query stage, a search interface is provided, as shown in the upper side of which includes the input for subject keywords and drop box selections for the country site and division of the company. More search terms such as the category of connections in the network can also be specified in the advanced search mode

12 EGO NETWORK VISUALIZATION

In SmallBlue Ego, the personal social capital management tool, we present a dynamic ego network view to help the user access his connectivity to the collaborators, their profiles, and his ego network evolvements, hence to increase the user's social situational awareness in the enterprise. In its static mode, the ego network visualization is designed as a disk in the background divided into multiple slices indicating different countries or divisions. The current user is placed to the center of the disk with his direct collaborators scattered out in slices according to their country/division attributes. The closer the collaborators are placed to the center, the stronger they connect to the current user.

13 HUGE GRAPH VISUALIZATION

Although SmallBlue Net is shown to be useful in presenting topic-centric network information, there is still a need to understand the whole picture of the enterprise social network, the communities within it, and the interconnections and structural holes of the network. The simplex methods to filter the huge graph into readable size bring the side effect of losing the overall topology of the network, and more importantly prohibit the access to network details, which could be critical in the user's navigation tasks. In SmallBlue, we introduce a novel technique called HiMap to more effectively visualize huge graphs up to millions of nodes.

14 NETWORK GRAPH MINING

In enterprise, relationships of entities can be mined from various data sources and form networks of millions or billions of nodes and edges. Networks of people can reach millions, if all internal and external contacts are includes. When information content represents the node of a graph, then it is very easy to achieve graphs of billions of nodes and edges. In a newer version of SmallBlue, we addressed the scalability issue mainly
scalable algorithms specific applications, such as anomaly detection and diversity enhancement.

15 GBASE: GRAPH DATABASE FOR HADOOP FRAMEWORK

Numerous applications (e.g., neighborhood search, PageRank, subgraphs, proximity, etc.) are common to network graph analysts. Our goal is to develop a general and scalable graph mining framework for SmallBlue to support a variety of common core operations on large graphs. The design objective is threefold:

1) efficiently store and manage huge graphs in parallel, distributed settings to answer graph queries efficiently; 2) define common, core algorithms to satisfy various graph applications; and 3) exploit the efficient storage and general algorithms to execute queries efficiently.

16 SCALABLE ALGORITHMS DESIGN: CASE STUDIES

Orthogonal to the general graph mining system, we also designed application-specific scalable algorithms. Here, we present two case studies: 1) nonnegative residual matrix factorization for interpretable graph anomaly detection and 2) diversified ranking on large graphs. Non-negative Residual Matrix Factorization: Matrix factorization (i.e., to decompose the adjacency matrix of the graph by the multiplication between two

Query \ Applications	Browsing	Ranking	Finding Community	Anomaly Detection	Visualization
Connected Comp.			✓	✓	
Radius				✓	✓
PageRank, RWR	✓	✓		✓	
Induced Subgraph	✓		✓		✓
(K)-Neighborhood	✓		✓		✓
(K)-Egonet	✓		✓	✓	✓
K**-core**			✓		✓
Cross-edges				✓	✓

Among others, it is now widely recognized that non-negativity is a highly desirable property for interpretation since negative values are usually hard to interpret. For example, for the task of community detection, the so called
improve the low-rank matrices plus a residual matrix) is powerful to find graph patterns. For instance, the two low-rank matrices often capture the community structure

of the graph; and the residual matrix is often a good indicator for anomalies on graphs. A new application of SmallBlue aims at finding anomalies in enterprise. In this scenario, matrix factorization is important.

17 CONCLUSION

We have discussed various challenges and solutions for conducting SNA in enterprise. We considered multimodality aspects of people relationships, including social aspect, financial aspect, and human property aspect. We also discussed various system challenges such as large-scale graph mining and large-scale network visualization. This paper focused on the fundamental research and system issues. It did not discuss the various applications of enterprise SNA such as collaboration (e.g., enterprise location, social proximity access, social recommendation, social search, etc.), cyber security (e.g., anomaly detection, fraud detection, etc.), and commerce (e.g., social marketing and selling). On the scientific aspect, there are many unsolved issues. For instance, despite having collected the largest enterprise data set in literature for employee interactions, we still have not obtained the teleconference data (although it can be approximated by the calendar info) and the face-to-face interaction data in a large setting. Our preliminary studies show that the use of e-mail and IM is so intense that even people who meet face-to-face or on teleconferences spend comparable time in communicating with each other by e- mail and IM.

REFERENCES

[1] Degenne, Alain and Michael Forse, 1999.Introducing Social Networks. London: Sage.

[2] Scott, John. 2000.Social Network Analysis. 2nd edition.

[3] Understanding Social Networks: Theories, Concepts and Findings by Charles Kadushin.