Xjenza Online - Journal of The Malta Chamber of Scientists
www.xjenza.org
DOI: 10.7423/XJENZA.2014.2.02

Malta Chamber of Scientists

Research Article

Dynamics of Private Social Networks

Jonathan Mifsud and Matthew Montebello
Department of Intelligent Computer Systems, University of Malta

Abstract. Social networks, have been a significant turning point in ways individuals and companies interact. Various research has also revolved around public social networks, such as Twitter and Facebook. In most cases trying to understand what's happening in the network such predicting trends, and identifying natural phenomenon. Seeing the growth of public social networks several corporations have sought to build their own private networks to enable their staff to share knowledge, and expertise. Little research has been done in regards to the value private networks give to their stake holders. This is primarily due to the fact as their name implies, these networks are private, thus access to internal data is limited to a trusted few. This paper looks at a particular online private social network, and seeks to investigate the research possibilities made available, and how this can bring value to the organisation which runs the network. Notwithstanding the limitations of the network, this paper seeks to explore the connections graph between members of the network, as well as understanding the topics discussed within the network. The findings show that by visualising a social network one can assess the success or failure of their online networks. The Analysis conducted can also identify skill shortages within areas of the network, thus allowing corporations to take action and rectify any potential problems.

Keywords Linked Data – Social Networks – Social Network Analysis – Semantic Network Analysis

1 Introduction

Social Networks have gathered significant interest, not just by the general population, who got sucked into this whole phenomenon, but also scientists, researchers and corporates. The success of major social networks such as Facebook and Twitter has seen a marked increase in the interest taken by researchers into these networks. Most research has indeed focused on trying to understand what is going on in these Social Networks by doing what is called Semantic Analysis. This allows a machine to identify topics and subjects discussed within the network, and eventually allows machines to reason about the underlying network.

Various corporates have also jumped on the social bandwagon, not just by setting up social profiles on public networks but by actually setting up internal social networks. Indeed 44 % of German SME's interviewed in (Meske and Stieglitz, 2013) indicated that they use Social Media for internal purposes, with 39.06 % having an Internal Social Network. private social networks platforms such as NationalField, Yammer (Yammer, 2014) and Zoho Connect (Z. C. P. Ltd., 2014) have come up and started to cater for enterprises. Companies such as Ebay, XEROX and DHL have successfully made private social networks part of their work, a place where ideas can thrive, and processes streamlined. Such was the effect of these networks that Obama's election campaign's success in 2008 was attributed to their use of a private social network, which led to the creation of NationalField.

Private social networks have become a place where employees of the same company, or group of companies can share information, connect and help each other regardless of the distance. Similarly to their public counterparts, they made the world a smaller place, being a single entry point for communication across the whole company. The information shared on these Private Social Network, could be very rich, allowing machines to understand the topics being discussed, thus understand which members of the network have best understanding of certain subjects.

This paper will look into the unique environment of private social networks, and how the semantics of the content posted on the network can be exploited in order to further understand aspects of the company. The paper is structured in the following manner, first giving a background of the Organisation whose network is being explored. Then understanding and finding influential figures within the company, going to helping individuals finding expertise within the corporation. Whilst Social Networks bring plenty of positives, it doesn't mean that they always work out the way it's intended to be; thus the paper also looks at problems which a Private Social Network comes across. How these problems can be tackled, and identification of when in a company's life using a private social network can be beneficial.

Correspondence to: Jonathan Mifsud (jonathan.mifsud.07@um.edu.mt)

2 Goals

The goal of this research is to explore the phenomenon of private social networks, and their similarity to Public Social Networks. This paper looks with particular interest, towards the network structure and dynamics of the private network. Applying techniques used in public networks in order to understand who the influencers within Private Networks are, and how this relates to the structure of the Company.

The paper goes to explore further the semantics and topics discussed within the organisation, taking particular interest as to whether these can be a reliable source to identify skills within the organisation. All this whilst giving guidance on how to conduct future research on private social networks, and identifying parameters which will increase the chances of having a successful research on private networks.

3 Background Information

Conducting any research on private social networks is not necessarily straight forward, the primary stumbling block is the fact that these networks are private. Apart from the organisation who is running the network, and benefiting from what's being discussed few or none are the people with access to such information. For this research to be possible we had to partner with Impact Hub (Impact Hub, 2014), an organisation who uses a NationalField (National Field, 2014) powered private social network. A background of Impact Hub's organisational structure is provided below to aid one's understanding of the presented work.

Impact Hub, formerly The Hub, is an organisation which was founded in London back in 2005, with the purpose of fostering and promoting social innovation. The success of their model let to an unprecedented increase in their network, and currently have over 54 Impact Hubs world wide, with over 7000 members. This rapid increase both in members and size, meant that Impact Hubs were now not only a local thing, where entrepreneurs and start-ups could learn from each other but rather a global community. In order to bring closer their global community and make better use of skills located within the hub, and further understanding of why some hubs are more successful then others.

Through the platform provided by NationalField, they were partly achieving their goals, of having their whole world-wide network interconnected. This did not give them much information about the network and their users, thus a collaborative partnership was put in place allowing an analysis of their network which gives the opportunity to present this paper.

The Impact Hub Social Network structure provided by NationalField, looks very much like an earlier version of Facebook. In this social network, each member has his / her own profile page and is allowed to post status updates, comment on such updates, as well as like updates done by themselves or other individuals. The network also contained various groups, which are in themselves dependent on either the actual location of the Impact Hub that the user is a member of, or otherwise by interest in a particular group or initiative. All updates are in themselves public, thus anyone member of the network is allowed to read, and the structure of the Home Feed meant that newer posts always come up top of the network significantly increasing the reach within the network. Posts within a particular group automatically trigger an email, sharing the post to all members of the group. This inherent network structure is important to remember when conducting a Social Network Analysis as it does have an impact on the results.

4 Related Work

Social Network Analysis (SNA) is not necessarily done on online social networks (OSN) but can be conducted using various data sources. Previous research has already looked at analysing private enterprise and closed networks within organisations using SNA techniques. Kazienko et al. (2011) use SNA to improve enterprise structures, the research focuses on email to build it's model however recommends overlaying data sources, including emails and logs. This added information would give more dimensions relating to the Social Network, which would be able to map the real social network of the organisation. Zhu et al. (2008) and Golbeck and Hendler (2004) focus on trust within social networks, trust being the basic ingredient to develop relationships within the network. Zhu et al. (2008) use PageRank like technique to give value and weight to email communication, increasing trust through each communication with an individual. Whilst Golbeck and Hendler (2004) discusses how an individual will only seek reputable information through relationships which are in themselves reputable.

Of particular interest are papers on the more sociological side, such as Abreu and Camarinha-Matos (2011), where the authors look at measuring social capital. With this work focusing on an organisation whose members are other Enterprises, this paper gives light on the potential interest in our work. By understanding the skills, and values of the members in a social network or business incubator, one can far easier sell the benefits of being a member in such network. The paper goes on to propose a means of measuring capital within the network, and how easy it is to access the different types of capital available within the network.

Taking a look at online social networks most research has focused on Semantic Analysis of the networks. Twitter was the social network of choice in most research due to various favourable aspects, all updates are limited to 140 characters, and most importantly tweets are public to the whole community. The large volume of tweets, and network structure allows better semantic analysis when compared to other networks such as Facebook. Teufl and Kraxberger (2011) looks at extracting knowledge from /twitter, whilst Asur and Huberman (2010) look into predicting the future using data extracted from Twitter. Of particular interest are the works of Tao et al. (2012) and Abel et al. (2011) where both look at user modelling and semantic user profiles respectively.

Term frequency or TF/IDF could be used along other natural language processing techniques for automatic keyword extraction. This is done by identifying the most used keywords by a particular user and then use these keywords as a means to identify the main user interests (Wang and Jin, 2011). Semantic Enrichment can also be done using techniques such as co-occurrence (Wang and Jin, 2011), relating words which often appear together. Stankovic et al. (2012) look into the effectiveness of expanding the profiles, in addition to co-occurrence (DMSR) they build their own system hyProximity (HPSR). This takes into consideration semantic relatedness between classes such as `rdf:subclassOf` and other links such as objects being part of an industry, service or are a product (Stankovic et al., 2012). Finally they also propose the use of a Pseudo Relevance-Feedback (PRF) which is used to expand searches from the result-set (Stankovic et al., 2012).

It is also possible to use external service providers such as DBPedia Spotlight, Zemanta and OpenCalais (Tao et al., 2012). These services extract semantics from the provided text according to pre-set parameters, and use linked data from well known resources such as DBPedia (Sahnwaldt, 2014), a WikiPedia linked semantic database. The resulting data can then be processed in order to filter and keep only topics for which the user seems to be an expert for building the user profile. Interest Weight to Ontology (Nakatsuji et al., 2006) could be a way to further enhance user profiles by having ontologies pass weight to their parents.

Due to network members discussing various topics, there is also the possibility of using a vector space model to represent the different interests of each of the users (Tao et al., 2012). Vectors could be also used for social connections, where the vectors would represent the number of interactions with different members within the network (Fan and Li, 2008). TUMS presented by Tao et al. (2012) is an interesting project in itself which provides a user modelling service using twitter, and provides both an API and a graphical user interface. This system uses the last 300 tweets and linked articles to obtain a good user representation.

Whilst these papers successfully argue how Semantics can be successfully extracted from Social Networks, there is still a risk of Semantic Attacks on such networks. Still using the example of Twitter Kumar and Geethakumari (2014) argue that "OSN have increased the spread of misinformation in social networks", whereby incorrect information is believed by members of the network and shared again further increasing it's reach. One cannot also discount the possibility of individuals who purposefully manipulate information posted on the social networks. This could lead the machines to extract incorrect information from the systems and understandably give an incorrect analysis. Thus correctness of any semantic information extracted is important to give trust to the underlying system.

5 Methodology

In this section we explain the methodology by which this research was conducted, and is thus essential in view of the outcomes of the research. In the previous sections a background of Impact Hub and their network structure was given, for the purposes of this research access to an API was given. The API allowed the following functions:

1. Obtain a List of Users
2. Request profile information
3. View user stream
4. View comments and likes of updates posted in the stream

With no direct stream of past historical entries, and limited number of hits on the server, it was best to hit the API periodically asking information on a user at a time. This allowed access to whatever appeared on the user's feed, thus picking up any new status updates and comments in the system. It also allowed a fool-proof way of obtaining any updates to each user's personal profiles, as users could update their description at any time. With updates set at 1 minute intervals, and considering the could of total users by the end was of nearly 5800 users, this meant a latency of up to 4 days. Whereby each user's data was updated roughly once every four days.

Whilst admittedly this approach is not practical in realtime applications, due to the restrictions we had in place this was the most appropriate solution. In an ideal world, we would have obtained a Feed which describes the latest changes to the network, then the system could investigate each one separately. Due to the relationship, between us, our partner Impact Hub and NationalField who were providing the social networking platform and API this was not possible.

The information was systematically gathered from July 2012 to January 2014, covering a period of 18 months. During this period the data was semantically analysed and stored in a database. Once we closed the data-set being used for this period we started looking at both the semantics of the system as well as conducting a SNA. Details of the semantic analysis, Social Network Analysis and results will be discussed in the later sections of this paper.

6 Semantic Analysis

For the purposes of this research using an external service provider to parse and analyse data was deemed the most appropriate. The decision was taken in the light of previous research conducted by Tao et al. (2012) which successfully used OpenCalais for semantic annotations, and building user profiles based on these annotations. OpenCalais and DBPedia Spotlight were both considered and tested on sample datasets. The flexibility afforded when using optional parameters in DBPedia Spotlight, made it a better candidate solution for this system. In addition there were two other key points which led to this decision. First all terms are listed in WikiPedia, thus individuals could easily go to topics which are being discussed. Secondly OpenCalais is a propriety system, whilst with DBPedia Spotlight being open source one can load the system on a local server, and possibly extend it should the need arise. Being able to set up locally was considered important as when analysing the data, this would otherwise be sent to a 3rd Party, and would be subject to additional terms and conditions. Working locally ensures that Impact Hub or other enterprises which wish to use similar systems still retain full control of their data.

The Semantic Entity Extraction was then split into parts, depending on the type of content being parsed. Taking into consideration that the Biography, Status Updates and Comments are usually of different lengths and serve different purposes. The parameters used in the analysis were adjusted accordingly to increase recall whilst trying to ensure correctness. Each of the extracted entities were then stored in a database and linked to the original source, be it the biography, status update or a comment.

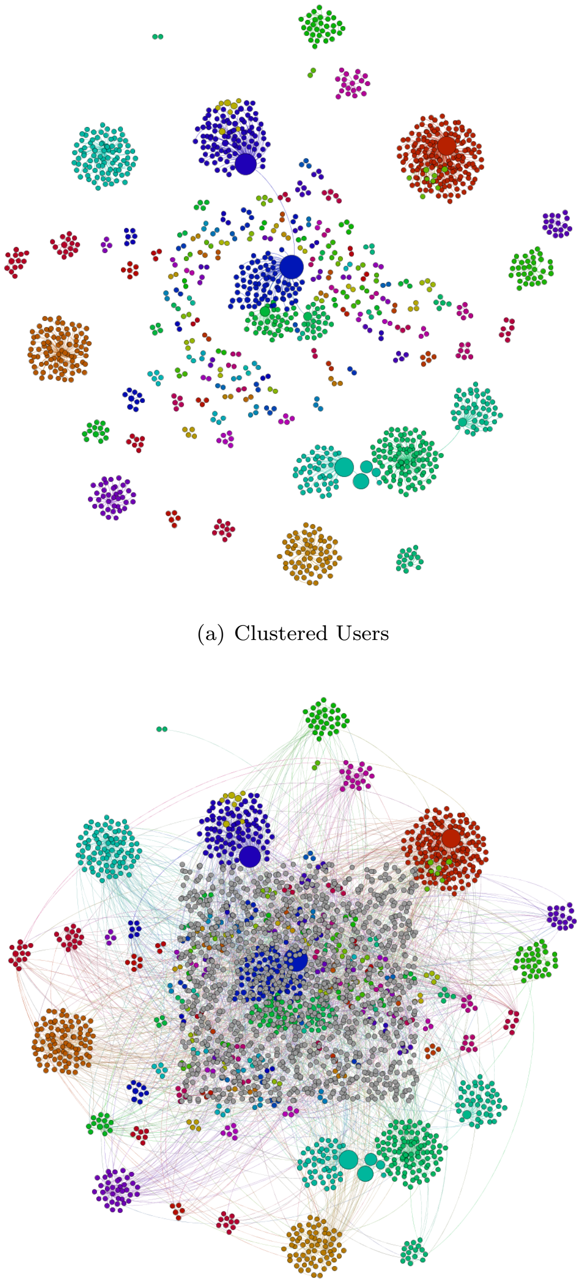
7 Social Network Analysis

Having captured information about the network, including who is posting, and which users were interacting with each other it was time to apply SNA techniques to see if the network structure on the social network correlates to the real-life situation. In addition to this we could augment the information with the semantic data extracted in the previous section. This allows a unique view on the network, which cannot be obtained in traditional manners.

The below table shows the total number of users, posts and comments extracted from the social network. Likes were omitted as this feature was under utilized with only around 30 likes registered within the system.

Metric	Value
Total Users	5,791
Users filled their biography	777
No of Posts	12,537
Post Edges Created	1,287
Users involved in Post	1,440
No of Comments	16,034
Node Edges Created	1,765
Users involved in Comments	2,059
Users involved in Posts or Comments	2,595

Due to a large number of inactive users within the network, highlighted above, the following sections will only focus on users who had any sort of interaction. Thus, this may not necessarily represent the actual network structure within the Impact Hub organisation. One can also notice that the large number of posts and comments within the network were mainly done through the same source and target, noticed easily when comparing the number of posts, and edges created in the network. For simplification purposes in our analysis, only users who are connected either through Posts or Comments are considered.



(a) Clustered Users

(b) Adding Connections through Comments

Figure 1: Users Connected through Posts and Comments

The first relationship which was studied was that created through Posts. Posts are usually sent from one entity to another, most of the time being users on the network, however could also possibly be directed to a group page. After converting the data into a graph entity and using a ForceAtlas algorithm for clustering the network structure found in Figure 1(a) was obtained. This visualisation shows that the users form clusters, with individuals who are the centerpoint of particular clusters. By visualising posts one can also notice that the network is interconnected and various nodes are only connected to one or two other nodes.

A delve deeper into this data, running a page rank algorithm in order to weigh the importance of members within the network resulted in a very interesting outcome. This algorithm gives higher importance to those users who have multiple posts directed to them, and as such their posts to other members also carry more weight. Out of the top 5 ranked individuals within the network, two were the Administrator of the NationalField platform, and the CEO of this platform. Whilst the Administrator is not really an active member, a significant number of posts from within the Impact Hub community were directed to him. The other three, included three important personnel within the Impact Hub network, the Global Managing director, a co-founder of Hub Copenhagen as well as the host of Hub Vienna.

This can be better understood when considering that these three individuals would have been very active in their respective communities, with the Global Managing Director, seemingly interconnecting the central cluster with other clusters. Each of the clusters also revolve around a location, with hosts of each cluster acting as an intermediary bringing people together. In the hub's actual structure, there is a host, whose job is to introduce new members and helping them connect. This role probably led to them welcoming other members in the network, and the new members in the network to go back to the host thanking them for being welcomed in the network one way or another.

Whilst the above network structure portrays very loosely interconnected clusters, using simply posts, introducing comments to the picture adds a totally different dimension as can be seen in Figure 1(b). In Figure 1(b), no ForceAtlas algorithm was used following the one run on Figure 1(a), and the new nodes just overlayed on top of the existing ones. One can observe the edges and nodes which were introduced in the network, representing comments, and users who only commented on other posts. One can notice connections from the outside clusters, also visible in figure 1(a), going from one end to the other. This shows that the relative distance and boundaries that existed in between the clusters no longer hold once users are allowed to comment on each other's posts.

Whilst the reasons for the increased interconnectivity when adding comments to our graph are not explored in detail in this paper there are underlying factors within the network which encourage this activity. First and foremost, unlike other public social networks, every member in the network is able to see all posts made by any user on the network. This allows users to respond to a query, or post to which a user would be interested in. Further research would be required to show whether the new links to outside clusters are linked to the semantic content of the posts, rather than a relationship which might exist outside of the social network.

8 Semantic Social Network Analysis

Having analysed the network structure, we take a look at the extracted semantic entities, and what machines could learn from the underlying network infrastructure. As biographies would be the most user representative these are taken into consideration, looking at the most commonly recurring entities the following are the 10 most popular:

- Sustainability
- Social Entrepreneurship
- Project
- Organisation
- Business
- Social
- Innovation
- Entrepreneur
- Design
- Idea

This after a small discussion with any of the members of the organisation, or even reading a bit on their website makes complete sense. The community is about social entrepreneurship, and sustainability achieved in a social manner, with various entrepreneurs sharing ideas to create new innovations, eventually leading to local businesses. That statement alone, covers over half of the keywords, in the top 10.

Each cluster is represented by a different semantic set of keywords, taking one of the clusters from Figure 1(a), the topics extracted vary slightly. Thus the composition of each cluster, is not only necessarily linked to location, but also to the skills available in the cluster. The top 10 entities from this cluster are:

- Communication
- Social Entrepreneurship
- Organisation
- Education
- Empowerment
- Culture
- Business
- Website
- Millennium
- Experience

One can notice that whilst there is an overlap in between the cluster and the main network, the cluster has keywords which are more specific, representative of topics discussed within the group. Keywords such as Social Entrepreneurship, Business and Organisation are quite generic and thus would be expected to feature all over the network. Semantic Keyword analysis could thus be used to identify key strengths of each cluster, and possible weaknesses by identifying keywords which are missing from these clusters.

9 Evaluation

Social Network Visualisation, can play a key role in understanding the structure of the virtual social network. This can easily help identify weak links within the networks, allowing administrators and moderators to take action and strengthen such links. One such example could be, that you want increased communication between two remote locations, or between two separate working groups.

Semantic entities as shown in the previous section, provide further information on skill and interest distribution within the network. This can be used more effectively to identify skill differences across the network, and better balance the members of an organisation where appropriate. In conjunction with other information showing the success of businesses within each cluster, this would help the network administrators to understand what mixture of skills create a successful enterprise.

The social network visualisation, though deemed effective, as it successfully clustered members across the network was not deemed to be a 100 % success. As one would have expected this visualisation to match the network structure of Impact Hub, which has over 53 Hubs around the globe. However a participation rate of around 50 %, is understood to have had a significant impact on the resulting visualization and analysis. This in itself is an important finding, as it helps the organisation to understand the lack of user activity, and thus be in a position to tackle it.

10 Arising Ethical Issues

When analysing a social network in such manner, one is inherently gathering information about each individual member and making assumptions on items discussed within the network. In particular cases, the semantic algorithms may pick up semantic connotations which are unwanted, or do not make much sense in context. When grouping data, such as in this paper these do not tend to show up however such connotations may be given to individuals where they do not exist.

This possibility of incorrect semantic connotations might be an issue especially when such data could face public entities. In this case the research was for internal purposes, and prior to showing data to other members within the network permission was sought from the related individuals.

11 Future Work

This work is certainly not the end of the line for private online social networks, following the proof of concept presented in this paper several items can be explored in more detail.

- Temporal changes in the private online social networks (do influencers change)
- Semantic composition of successful groups / projects within the network
- Overlaying SNA of an online private network with an organisation's real structure, how these overlap
- Explore weather users are more likely to comment on a post if this matches one of their interests.
- Using Semantic User Profiles to invite members to join particular working groups / clusters.

12 Conclusion

Overall the presented work outlines that in a limited private online social network, one can still conduct both Social Network Analysis and Semantic Social Network Analysis with a degree of success. In an organisation where this research is targeted to a particular purpose this sort of research can be quite useful in identifying both weak and strong points within the organisation. Then conducting a plan to improve the Social Network content delivery in order to increase engagement where required.

References

Abel, F., Gao, Q., Houben, G.-J. and Tao, K. (2011). Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web. *Lect. Notes Comput. Sci.* 375–389.

Abreu, A. and Camarinha-Matos, L. M. (2011). An Approach to Measure Social Capital in Collaborative Networks. *IFIP Adv. Inf. Commun. Technol.* 29–40.

Asur, S. and Huberman, B. A. (2010). Predicting the Future with Social Media. *2010 IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*

Fan, L. and Li, B. (2008). VisoLink: A User-Centric Social Relationship Mining. *Rough Sets Knowl. Technol.* 668–675.

Golbeck, J. and Hendler, J. (2004). Accuracy of Metrics for Inferring Trust and Reputation in Semantic Web-Based Social Networks. *Lect. Notes Comput. Sci.* 116–131.

Impact Hub. (2014). What is Impact Hub. [Online]. Available: http://www.impacthub.net/what-is-impact-hub.

Kazienko, P., Michalski, R. and Palus, S. (2011). Social Network Analysis as a Tool for Improving Enterprise Architecture. *Lect. Notes Comput. Sci.* 651–660.

Kumar, K. P. K. and Geethakumari, G. (2014). Analysis of Semantic Attacks in Online Social Networks. *Commun. Comput. Inf. Sci.* 45–56.

Meske, C. and Stieglitz, S. (2013). Adoption and Use of Social Media in Small and Medium-Sized Enterprises. *Lect. Notes Bus. Inf. Process.* 61–75.

Nakatsuji, M., Miyoshi, Y. and Otsuka, Y. (2006). Innovation Detection Based on User-Interest Ontology of Blog Community. *Lect. Notes Comput. Sci.* 515–528.

National Field. (2014). NationalField - Measure Change.

Sahnwaldt, C. (2014). wiki.dbpedia.org - About.

Stankovic, M., Rowe, M. and Laublet, P. (2012). Finding Co-solvers on Twitter, with a Little Help from Linked Data. *Lect. Notes Comput. Sci.* 39–55.

Tao, K., Abel, F., Gao, Q. and Houben, G.-J. (2012). TUMS: Twitter-Based User Modeling Service. *Lect. Notes Comput. Sci.* 269–283.

Teufl, P. and Kraxberger, S. (2011). Extracting Semantic Knowledge from Twitter. *Electron. Particip.* 48–59.

Wang, Q. and Jin, H. (2011). Social Analytics for Personalization in Work Environments. *Lect. Notes Comput. Sci.* 314–326.

Yammer. (2014). Yammer: The Enterprise Social Network. [Online]. Available: https://www.yammer.com/.

Z. C. P. Ltd. (2014). Zoho Connect — Your Social Workspace. [Online]. Available: http://www.zoho.com/connect/.

Zhu, W., Chen, C. and Allen, R. B. (2008). Analyzing the Propagation of Influence and Concept Evolution in Enterprise Social Networks through Centrality and Latent Semantic Analysis. *Adv. Knowl. Discov. Data Min.* 1090–1098.