A Study of the Online Profile of Enterprise Users in Professional Social Networks

Alessandro Bozzon, Hariton Efstathiades, Geert-Jan Houben
Delft University of Technology
Delft, The Netherlands
{a.bozzon, h.efstathiades, g.j.p.m.houben}@tudelft.nl

Robert-Jan Sips
IBM Benelux
Center for Advanced Studies
Amsterdam, The Netherlands
rjsips@nl.ibm.com

ABSTRACT
Understanding the impact of corporate information publicly distributed on the Web is becoming more and more crucial. In this paper we report the result of a study that involved 130 IBM employees: we explored the correctness and extent of organisational information that can be observed from the online profiles of a company's employees. Our work contributes new insights to the study of social networks by showing that, even by considering a small fraction of the available online data, it is possible to discover accurate information about an organisation, its structure, and the factors that characterise the social reach of their employees.

Categories and Subject Descriptors
H.4 [Information Systems Applications]: Miscellaneous

Keywords
Professional Social Network, User Analysis

1. INTRODUCTION
The Web is now a pervasive medium, where several technical aids have been created to stimulate content creation and publishing. Online social media allow users to create social profiles, where their personal and work information are available for consultation by a wide variety of actors, from friends to recruiters and enterprise analysts.

Companies mainly perceive social media as a way to support business processes. Surveys [2] demonstrate that corporations are mostly using social technologies to scan the external environment for new ideas, for marketing and for competitive intelligence purposes. The growing use of networking platforms within organisations (often in the workplace) may also pave the way for a new class of enterprise applications, where the online profile of employees can be a vehicle for a better understanding of the internal and external corporate dynamics.

In this paper we present a case study performed within IBM Netherlands, in which we aim to look at the organisation outside-in. The aim of this study is to gain insights about the extent and correctness of organisational information that can be observed from IBM-employee social media profiles. Our research aims at providing answers to the following questions.

1. Which kind of enterprise business knowledge can be discovered from professional online networks?
2. Which enterprise properties have an impact in the social reach of users in professional online network?

The main result of this paper is the empirical demonstration of the nature and extent of the corporate information that can be explicitly or implicitly observed from an online professional networking platform, even when considering only a small fraction of the published data. To the best of our knowledge, this is the first study of this kind.

We have focused on the most popular professional networking platform, namely LinkedIn. We engaged in a case study that involved 130 IBM employees working in different departments all over the world. We asked them to participate in an exploratory survey for demographical and online social activity characterisation and we created a realistic interaction scenario to obtain access to their professional online network profile on LinkedIn, through its public APIs, and adhering to the privacy settings of the involved users and their contacts. We then analysed the gathered data (including current occupation, social communities, etc.) to inspect the relationships that exist between their official allocation and responsibilities within IBM, and their professional environment in LinkedIn. Our approach can be adapted to various contexts, such as different online professional networks, and to other companies. Although the experiment specifically involved IBM employees, we believe that our results are of general interest.

The remainder of the paper is organised as follows: Section 2 describes related work; Section 3 describes our study; Section 4 reports the results of the exploratory survey; Section 5 provides a discussion on the outcome of the experiment and reports on gathered insights. Finally, Section 6 presents our conclusions.

2. RELATED WORK
Several studies show the increasing importance of Social media for organisations. [1] observes that new communication and collaboration technologies often encounter initial
organisational resistance, where about half of US companies reportedly block or restrict social network access. Despite that fact, [1] reports that surveys within Microsoft, conducted from 2008 to 2011 show a gradual rise in the workplace usage of Facebook, LinkedIn and Twitter. The same observation is shared by a recent survey [3] conducted by global consulting firm McKinsey, in which 65 % of companies report the use of social technologies within their organisation.

Social networks, referring either to physical or online communities, are a rich source of knowledge. During the last years, researchers studied a variety of social network communities in order to gain a more intimate understanding of the dynamics of online users. For instance, [2] focused on the analysis of LinkedIn profiles for the purpose of studying the career evolution of the alumni of a south-eastern US university after graduation. [4] investigates if the hierarchy of an organisation can be inferred from social networks.

The analysis of the social networking habits of employees could provide information very useful for the purposes of an organisation. A study conducted by the IBM Institute for Knowledge-Based Organizations [6] shows how the analysis of key figures in a company may lead to increased understanding of information flows within the organisation. Also, it concludes that knowledge about employees competencies and job roles are factors which affect the knowledge sharing within an organisation and, hence, could be used to drive the organisation’s activities. [5] studies enterprise users- social network activities to determine the type of interaction patterns that might reveal real-life relationships between colleagues. [4] analysed how the structural properties of professional networks (including organigrams) in a large international organisation affects the interaction patterns within the enterprise social network. They considered such factors as the user's geographic location and rank in company’s hierarchy, showing that users that hold a position higher in the hierarchy are more likely to receive replies from other users, as they receive more attention because of their influence.

Our study is inspired by previous results in the field, but departs from existing literature by being the first work that analyses the relationship existing between job roles of an enterprise employees, and his/her online profile and social reach in professional social networks.

3. EXPERIMENTAL SETTING

We organised our work into four research questions, hereby reported.

RQ1: Which enterprise organisation information is implicitly revealed by users in professional social networks?

An employee is intrinsically motivated to enrich the curriculum vitae with as many details as possible about his/her experience and skills, including past and current job responsibilities. As the user profile in a professional online network is directly related to his/her CV [5], we aim, by answering this research question, to analyse the extent to which an employee’s self-promotion in a professional online network can reveal information about the internal organisation of a company.

RQ2: Can an organisations’ key employees be identified from professional social network profiles?

In every organisation, key employees like managers play a central role in (strategic) decisions. This research question aims at investigating how the information published by employees in online profile would allow an external observer to identify such key employees.

RQ3: Which factors related to the employer influence the social reach of an employee?

Factors such as gender and culture are known to influence the social reach of users in online networks [9], especially concerning the size and composition of their connections. This research question aims at investigating which additional factors play a role in determining the social reach of an employee in professional online networks. Specifically, we investigate whether the employee’s nationality, industry of employment, and size of the employer organisation have an effect on the size of his/her community in a professional online network.

RQ4: Which factors related to the operational organisation within a company influence the social reach of an employee?

Considering the inherent organisational complexity and diversity of a corporation like IBM, the goal of this research question is to investigate how the operational organisation of a company plays a role in the composition of the network of their employees. Specifically, we investigate to what extent employees tend to establish connections with other member of the organisation, and how this aspect affects their social reach.

3.1 Methodology

To investigate such questions, we engaged IBM employees from all around the world. We first recruited personnel within IBM Netherlands, but then, thanks to internal promotion of the study, we were able to recruit a total of 134 IBM employees, working in several countries.

This study ran from January until March of 2013. Participants were invited to fill out an online introductory survey where they were asked demographic questions (e.g. country of origin, gender, etc.), but also questions intended to characterise their usage of online social platforms in the work environment (e.g. which platforms do they use, frequency of usage, etc.). All participants agreed to grant us readonly permissions to their LinkedIn profiles. Starting from the initial seed of 134 Core IBMers, we retrieved the profile information of their connections, totalling ~40000 user profiles, ~9000 of which employed at IBM. LinkedIn profiles obtained from IBM employees were then linked with corporate information retrieved from the IBM BluePages platform.

IBM BluePages is IBM’s internal tool which provides IBMers with yellow pages and expert finding capabilities. In our experiment, IBM BluePages is the data source from which we derive the correct and up-to-date employee information used for comparison with the information in LinkedIn profiles. Every IBMMer has a profile in BluePages, including details about the department office/building he/she works in, expertise, manager’s name, etc.

We distinguish three types of users, represented in Figure 1 and characterised by a different set of available data.

1. The LinkedIn User entity represents users of the LinkedIn platform. It specifies the information that is accessible for all the profiles we were granted access to through
Figure 1: A simplified meta-model for the information retrieved from IBM BluePages and LinkedIn.

the APIs, i.e. the Core IBMers. For LinkedIn User it is possible to access information such as: 1) Headline: description of the user’s job position. 2) Country: user’s country code. 3) Industry: the industry that the user has indicated that he/she belongs to, according to a taxonomy provided by LinkedIn at data insertion time. 4) Community Size: the number of connections a user has, with a maximum reported amount of 500. 5) Positions: information about current positions that a user reports in LinkedIn. For each position, the user can specify a Job Role, and information about the Company that user works for (e.g. the Industry in which the company operates, and its Size in terms of number of employees)

2. IBMers are LinkedIn users that work for IBM. In addition to LinkedIn profile information, an IBMer is described by data retrieved from IBM BluePages, including: 1) Employee Code: an unique identifier of the employee within IBM. 2) Job role: A description of the IBMer job role inside IBM. 3) Manager: The employee code of his/her manager. 4) Is_manager: a flag that indicates if the IBMer is responsible for the management of at least one other employee in IBM.

3. Finally, Core IBMers are the main participants of our study. Thanks to their access permission, we were able to access their full LinkedIn profile, including languages, skills and information about his/her education (e.g. the obtained degrees).

4. SURVEY RESULTS

This section details the demographic distribution of our pool of study subjects, and reports on the results of the survey.

<table>
<thead>
<tr>
<th>Work in country of origin</th>
<th>Percentage</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netherlands</td>
<td>51.5</td>
<td>69</td>
</tr>
<tr>
<td>Belgium</td>
<td>12.6</td>
<td>17</td>
</tr>
<tr>
<td>USA</td>
<td>11.9</td>
<td>16</td>
</tr>
<tr>
<td>Denmark</td>
<td>7.5</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country of origin</th>
<th>Percentage</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netherlands</td>
<td>73.4</td>
<td>101</td>
</tr>
<tr>
<td>Belgium</td>
<td>14.7</td>
<td>21</td>
</tr>
<tr>
<td>USA</td>
<td>11.9</td>
<td>16</td>
</tr>
<tr>
<td>Denmark</td>
<td>7.5</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Education</th>
<th>Percentage</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School Diploma</td>
<td>9.0</td>
<td>12</td>
</tr>
<tr>
<td>BSc Degree</td>
<td>29.1</td>
<td>39</td>
</tr>
<tr>
<td>MSc Degree</td>
<td>42.9</td>
<td>57</td>
</tr>
<tr>
<td>Ph.D Degree</td>
<td>6.7</td>
<td>9</td>
</tr>
<tr>
<td>Other Diplomas</td>
<td>12.7</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 1: Demographic characterisation of CoreIBMers

The survey allowed us to gather additional information about the social networking habits of our subjects, including 1) the monthly frequency and hourly duration of their visits; 2) the age of their network profile; and 3) their reasons for creating new connections.

Online social network adoption. Figure 2 depicts the distribution of online platforms of choice for our CoreIBMers. LinkedIn is the most used platform (90.3%) for professional networking, while Facebook is mainly used to connect with friends, family and acquaintances (70.9%). This result is aligned with findings in previous literature, and shows that our user sample exhibits a common behaviour.

Age of Online Network Profiles. Figure 3a shows that a significative amount of our subjects are also long-standing users. For social platforms, 63% of the participants created an account more than 3 years ago, while for professional platforms like LinkedIn the percentage is slightly lower, namely 56%.

Social Network Usage. A similar preference for social platforms can be also identified in the usage behaviour of our subjects, as depicted in Figure 3b. 36.57% of them visit their social online network more than once daily, while 32.84% access professional online networks more than once a week. In the same way, social networks are visited, on average, for a longer timespan (Figure 3c).

Motivation for Connection Creation. Figure 3d shows the distribution of preferences for new connection creation. The most common reasons to establish new connections on online social networks is friendship (89.55%), while being colleagues (36.57%), or business partners (e.g clients, potential clients) attracted roughly a third of the preferences. In professional online networks, work-related motivations prevail, where business partners (91.04%) and colleagues (77.61%) are the most important reasons for establishing a new connection.

Discussion. The results of the survey can lead us to the general observation that our employee sample is characterised by a more persistent interaction with social online networks, which might indicate how professional online networks are not perceived useful for social communication and interaction purposes. However, professional social networks are the platform of choice for maintaining digital relationship with colleagues, clients and business partners; the observed usage pattern suggests that our pool of IBMers has a consistent commitment toward visiting and maintaining their professional network profile.
5. FINDINGS

In this section we report on the analysis performed on the profile information retrieved from the LinkedIn professional networks, with respect to the research questions listed in Section 3.

5.1 RQ1: Information About Enterprise Organisation

LinkedIn provides users with several features to describe, in a well-structured way, current and past job positions. However such information can only be accessed by first-level connections, or by premium users. To cater for this limitation and to perform a study that relies on the most commonly available information, we decided to focus our attention on LinkedIn’s Headline, i.e. the field devoted to the description of the current job position of a given user.

The analysis included the 9100 IBMers and we compared their Headlines with the Job Role text field in the BluePages system, looking for similarities and differences. With this comparison we aim to investigate whether IBM employees reveal their internal job responsibilities as part of their job description in professional social networks, thus showing information about the internal organisation of the company. We recall that the Job Role field in BluePages contains manually curated (and therefore precise) information about the actual role of an employee within IBM. Moreover, at the time of the experiment, there was no policy from IBM according to the information that IBMers can publish about their job role.

After a manual inspection of the retrieved data, we noticed that the application of traditional textual (semantic) similarity techniques would not be effective to recognise textually different, but semantically similar job descriptions. To achieve maximal accuracy, we decided to manually compare the Headline and Job Role text fields when their content differed. We then categorised the result of the comparisons in three classes, respectively Equal, Similar, and Different.

- **Equal.** Means that the employee has the same LinkedIn Headline and BluePages Job Role text. This class contains 75% of the analysed employees.

- **Similar.** In this case, the textual description of the job role is different, but semantically related. This class contains 13% of the analysed employees, and it can be further specialised in: 1) Same role described with different terms (23.5%), i.e. headlines where the job responsibility is only rephrased (e.g. “CICS Product Line Manager at IBM” in LinkedIn and “CICS TS Product Manager” in BluePages). 2) Generalised role description (76.5%), e.g. “Service Sales at IBM” in LinkedIn and “Service Sales Public/ING” in BluePages. In these frequent cases, employees do not disclose specific details about the clients they are working for, although this information is required in the internal IBM system.

- **Different.** 12% of the employees used a completely different description in the professional social network. 80.2% of them published their general job role in LinkedIn, while BluePages contains their exact responsibilities. An example of this category is a user who has "Associate Project Manager IBM" in LinkedIn and "Mobility phones, data cards, blackberries Services" in BluePages. 19.8% of the employees with different job role description used a personalised, captivating description in LinkedIn, and their actual position and responsibilities in BluePages. An example is a user who describes his self in LinkedIn as "Passionate provider of Competitive Advantage for IBM" and as "European Sales Execution Leader" in BluePages.

**Discussion.** The analysis shows that the majority of IBMers reveal their internal organisation job role in LinkedIn. Only in few cases employees reported an inconsistent description of their job role, typically in favour of a more general or more appealing description of their current positions. This result clearly indicates that actual information about IBM internal organisation are also accessible to the outside world.

5.2 RQ2: Key Employee Identification

Considering the results obtain for RQ1, in RQ2 we focus on a specific property of employees, namely their importance within the organisation. In the context of this study, important employees are the ones which hold some management responsibility that involves the coordination of other people’s work. This information can be really important to understand the internal dynamics of an organisation, especially...
when it is also possible to infer the associated operational responsibilities.

To answer this question we try to identify patterns in the terminology that managers and non-managers use in their public profile job description. Thanks to BluePages we were able to identify IBM managers, i.e., employees that satisfy the above definition. The analysis resulted in the identification of 2233 managers, out of the 9100 considered employees.

We performed a linguistic analysis on the headlines specified by IBM managers in LinkedIn. Table 2 reports the frequency of the 10 most occurring terms and bigrams, highlighting the ones that co-occur in such lists. In both cases, a clear terminology separation is in place, although the word “manager” is surprisingly emerging as the most used in both categories. Out of the first 50 bigrams, only 16% are shared by both managers and non-managers. This result can be justified by the usage of those terms with a more generic meaning; for instance, the bigrams “program manager”, “marketing manager”, and “sales manager”. A punctual investigation performed with several IBM employees show that such terms are used to identify general responsibilities, such as the direction of a program, or of a partnership. Additionally, the term “senior” is popular between non-managers because it also indicates the level of knowledge and experience that they have in their specific field.

Discussion. The analysis shows that there is a clear linguistic distinction in the job role description of managers and non-managers, thus suggesting an easy identification of important coordination figures within the organisation. Overlapping terms generally refer to more operational responsibilities, which might not be related to management.

5.3 RQ3: External Factors Influencing Social Reach

This research question aims at investigating the factors, external to the organisation of the company, that contribute to the social reach of employees in terms of number of connections. To perform the study we retrieved the number and type of connections for all the Basic Users having a social reach (i.e., number of connections) lower than 500 from LinkedIn.

We investigated the following factors: 1) Country of the user; 2) user job role Industry as reported in LinkedIn; 3) the Industry of the company the employee is working for; 4) Size of the company the user is working for, as reported in LinkedIn.

Tables 3 and 4 report the results of a t-test statistical analysis performed on the comparison of the average number of connections held by users associated with a given factor (e.g., Dutch employees vs. Belgian). We only report factors that successfully passed a normality test on their distribution. A √ symbol indicates comparisons where the difference in the number of connections has been found statistically significant (p < 0.01).

Table 3 shows that the country of origin is a characterising factor for social reach, and it allows to tell employees from the Netherlands from employees from Belgium, Dutch employees from American employees, Belgian employees from English employees. No conclusions can be drawn for employees coming from Belgium and United States, and United

---

1 The same overlap can be seen also for other terms such as “director”, “partner”, “leader”, “executive”.
2 LinkedIn poses a limit to the retrieval of the connections for a given user to 500, with no possibilities (at the time of the study) to decided filtering or ordering retrieval conditions.

---
Table 5: Connection of in LinkedIn between employees at different level of the organisation’s hierarchy.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Manager</td>
<td>45.63</td>
<td>17.47</td>
</tr>
<tr>
<td>Manager</td>
<td>57.69</td>
<td>23.07</td>
</tr>
</tbody>
</table>

Table 6: Differences in community size of Core IBMers according to the presence of a LinkedIn connection with 1) the manager; and 2) the manager of his/her manager.

<table>
<thead>
<tr>
<th>Conn. With Man.?</th>
<th>Conn. With Man.’s Man.?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
</tr>
<tr>
<td># IBMers</td>
<td>21</td>
</tr>
<tr>
<td>Avg. Comm. Size</td>
<td>635.71</td>
</tr>
</tbody>
</table>

5.4 RQ4: Internal Factors Influencing Social Reach

We investigated the extent by which employees tend to connect with managers in professional social networks. To guarantee a result unbiased by the specific characteristics of our social crawl, we decided to perform the analysis only on the set of Core IBMers having less than 500 connections, a total of 129 employees. IBM BluePages provided information about the management relationship between employees (i.e. employee A manages employee B). Table 5 reports the fractions of managers and non-managers that, on LinkedIn, respectively connect with their managers, and with their manager’s manager. Table 3 shows the differences in the social reach between Core IBMers according to the presence of a connection with their manager and manager’s manager.

Discussion. In general, managers tend to connect more with employees higher in the organisation’s hierarchy; however, the overall percentage of employees that have a connection with their manager’s manager is significantly lower. Notably, employees who are connected with both their managers, and with their managers’ manager have a statistically significant larger community. An exception is represented by the employees that have a connection only with their manager’s manager, but the small number of employees in this category allows no conclusion. A more in-depth analysis of such communities in terms of similarities and differences is left for future work.

6. CONCLUSIONS

In today’s online society, understanding the impact of corporate information publicly distributed on the Web is becoming more and more crucial; this paper has explored the nature and extent of organisational information that can be observed on professional online network through the profiles of a company’s employees. From the results of our investigation we can conclude that, in the case of IBM Netherlands, small bits of information drawn from LinkedIn profiles were sufficient to discover accurate information about IBM’s organisational structure. We also observed significant differences in terms of social reach of employees that can be justified both by internal and external properties of the organisation such as the country of operation or the industry type. Although targeted at a specific population of users from a single company, we believe that the results of our research could be observed also in other organisations. Our results can be applied to several business intelligence scenarios and can be used to train or configure more automated methodologies.

As part of the future work, we plan to extend the analysis to more IBM employees, but also to other online networking platforms. The goal is to extend the reach of our study to other information that can be extracted from online profiles.

7. REFERENCES


