Recognizing Skill Networks and Their Specific Communication and Connection Practices

Sergiu Chelaru, Eelco Herder, Kaweh Djafari Naini, Patrick Siehndel
L3S Research Center
Leibniz University Hannover, Germany
{chelaru, herder, naini, siehndel}@L3S.de

ABSTRACT

Social networks are a popular medium for building and maintaining a professional network. Many studies exist on general communication and connection practices within these networks. However, studies on expertise search suggest the existence of subgroups centered around a particular profession. In this paper, we analyze commonalities and differences between these groups, based on a set of 94,155 public user profiles. The results confirm that such subgroups can be recognized. Further, the average number of connections differs between groups, as a result of differences in intention for using social media. Similarly, within the groups, specific topics and resources are discussed and shared, and there are interesting differences in the tone and wording the group members use. These insights are relevant for interpreting results from social media analyses and can be used for identifying group-specific resources and communication practices that new members may want to know about.

Categories and Subject Descriptors

H.5.4 [**Hypertext/Hypermedia**]: Navigation; H.3.5 [**Online Information Services**]: Web-based services

Keywords

skills, expertise, social networks, connections, topics, sentiment, content

1. INTRODUCTION

People who work in similar professions typically share particular skills. Further, if people are asked to indicate their skills, it is expected that the skills they mention vary in granularity. For example, someone working in public relations may indicate skills in social networking and marketing, but also specific skills such as DTP software, writing press releases and time management.

It is also known that people from different professions or cultural backgrounds have different practices in how they communicate with one another, the communication mechanisms that they choose and the topics that they discuss [12]. These differences can also be observed on a more private, personal level: programmers are usually more informal than bankers, people working in public relations are typically more active in social media than investors, and pastors will most likely talk about different topics than real-estate agents.

In this paper, we investigate differences in communities within self-reported skill networks. We are particularly interested in discovering differences in their communication practices: how well is a professional community connected, how often do people post updates via Twitter or Facebook, what are the topics that they talk about, and what is the overall tone or sentiment of these communications? Particularly for people who aim to identify and approach experts from a different profession, who wish to promote their services in other communities, or who consider a career switch, it is important to know the unwritten rules in a network. For example, what would programmers think of overly positive marketing language? How often can one repeat an announcement? Would it be a good idea to add a personal touch or will that be considered 'unprofessional'?

Being aware of differences between professional communities is also important for interpreting statistical data from social network analysis. For instance, in some communities the average number of followers is considerably higher than in other communities. As a consequence, a person from a well-connected community like online marketing with, say, 300 followers, may be considered isolated; for a programmer, this is actually a very good number. The same differences apply for interpreting centrality and other in- and out-degree measures.

The main contributions of our paper are: we provide an overview on how skills in professional networks are related and categorize these skills into professions. Further, we show to what extent different professions differ from one another in terms of connections, topics, sentiment and shared content. Finally, we discuss implications for social network analysis and the design of professional networking sites.

The remainder of this paper is structured as follows. In the next section we discuss related work, followed by a description of the dataset we used. In Section 4 we discuss the structure of the skill network derived from LinkedIn profiles and how this structure is reflected in the professions that we extracted using LDA. The results are presented in four subsections, covering: connections between people, topics that people discuss about, subjectivity and polarity of the word-

HT'14, September 1–4, 2014, Santiago, Chile.

http://dx.doi.org/10.1145/2631775.2631801 .

ing, and the resources that they share. We end the paper with a discussion and concluding remarks.

2. RELATED WORK

Our work draws upon two related strands of research. First, our focus on skill and expertise networks fits in the research area of automated expert finding, in which both explicit and implicit information is used for identifying experts in a particular area. Our interest in differences between expertise domains in how people connect and communicate online follows the tradition of social media analysis.

Yimam-Seid and Kobsa [20] argue that for the effective use of knowledge in organizations, it is essential to exploit tacit knowledge that is hidden in various forms, including in the people's heads. The authors also separate the need for 'information' from the need for 'expertise': the need for people who can provide advise, help or feedback - or who can perform a social or organizational role. Their expertise recommender made use of a hand-tailored expertise model.

MacDonald et al [15] indicate that, in order to identify experts, documentary evidence is needed. This evidence may be based on documents, emails, web pages visited, or explicitly created profiles with an abstract or a list of their skills. This evidence will than be ranked with respect to a given query or goal skill profile. Based on the TREC W3C and CERC test collections, they evaluated to what extent additional evidence could improve expert retrieval. They found out that the proximity of a candidate name to query terms and clustering of main expertise areas are the best indicators. Extracted text from homepages and the number of inlinks did not have much influence.

Balog and De Rijke's experiments [3] with data from the 2005 TREC Enterprise track show that user expertise can effectively be derived from email content; the persons being cc'd in an email were often authorities on the content of the message. Ghosh et al [10] leveraged social media (Twitter) content for seeking experts on a topic. Their results indicate that *endorsement* in other users' Twitter Lists (of which the topics need to be extracted) infers a user's expertise more accurately than systems that rely on someone's biography or tweet content.

Guy et al [11] examined indicators for expertise and interest as expressed by users of enterprise social media. The results are based on a large-scale user survey. They separate 'expertise' (being knowledgeable or skilled) from 'interest' (curiosity, basic knowledge, desire to learn more). As expected, interest and expertise ratings are correlated, with values for interest higher than for expertise. Results indicate that blogs and microblog provide different, more useful, information than communities and forums.

The above-mentioned studies suggest that people's skills and expertise can be derived from both explicitly provided lists and from their connections and communication patterns. This is consistent with Cingano et al's [8] observation that better-connected unemployed individuals, particularly those whose contacts were employed, are more easily reemployed. However, all of these studies were conducted in a single professional area or they generalized results between different areas. It is likely that considerable differences can be found between communities. For example, Hong et al [12] found that Twitter users of different languages adopted different conventions with respect to the inclusion of URLs, hashtags and mentions, as well as on replying and retweeting behavior. The main conclusion they drew is that the 'average' behavior of the English-speaking community does not necessarily translate to other communities.

In our study, we will look at differences in how people from different professions are connected, the topics that they discuss, the subjectivity and polarity in their wording, and the type of resources or websites that they share. These topics have been subject of research in various studies, a small selection of them is discussed in the remainder of this section.

Kumar et al [13] analyzed the structure and evolution of online social networks. They showed that networks typically have one well-connected core region, but most users are located in one of several more or less isolated communities around it. These communities are typically centered around one central person, and it is unlikely that two isolated communities will merge at some point. In the next section, we will see that the structure of our skill-based network matches these observations.

Abel et al [2] compared different approaches for extracting professional interests from social media profiles. Results indicate that dedicated tag-based profiles and self-created user profiles are most suitable for this task. Twitter profiles are more diverse but also more noisy; this effect can be reduced by extracting entities from running text. In a follow-up study, Abel et al [1] analyzed the completeness of user profiles in different social media. The outcomes suggest that user profiles in networking services, such as LinkedIn, are more complete than those in services like Twitter. Further, the topics that users talk about differs between channels, but the overlap in topics is higher between services that are used for similar purposes. It was also shown that combining information from different services was beneficial for tag and resource recommendations.

Siersdorfer et al [17] investigated the usefulness of comments, as perceived by YouTube users. They found out that positive comments were considered more useful than negative comments. Differences between categories were also found: for example, science videos receive predominantly objective comments, politics relatively many negatively rated comments, and music videos mainly attract positively rated comments. These findings suggest that different communities have different norms with respect to commenting - we expect that the same effect can be observed if one compares different professions.

3. DATASET

In order to create our dataset, we first collected a set of 94,115 public user profiles from About.me, using the crawling strategy employed by Liu et al. in [14]. About.me is a personal profile site where users can include all their social-web accounts. From each profile, we collected the users' LinkedIn, Twitter and Facebook accounts.

For LinkedIn, our crawler gathered the public profile data, including skills and expertise tags, industry, job and number of connections.

For each account from Twitter, we gathered the complete user profile with information like number of followers and friends or number of lists the user is in. Beside that, we crawled the latest 200 Tweets using the Twitter Rest API [1]. The average number of Tweets posted by the users is 5,833,

[1]https://dev.twitter.com/docs/api

with a median of 1812. This indicates that most of our users are quite active in Twitter. We also had 33 users with more than 100.000 followers, which is already pretty influential.

The Facebook subset was collected using the Facebook API [2], which provides access to the public profile information of the users. Here, our crawl was focused on the Facebook timeline of the user, which mainly contains the shared posts. On average, the number of posts per user is 210 (median 23), with 34 users having more than 5,000 posts. Further, we collected data on the most popular features in Facebook, including the number of likes, number of comments, and number of shares on the users' posts.

In total, we have 33,516 users with a LinkedIn profile, 46,799 users with a Twitter profile and 34,523 users with a Facebook profile. Since the LinkedIn account serves as a source for our topics describing the users, we use for our analysis only Twitter and Facebook profiles that have a corresponding LinkedIn profile, resulting in a final set of 7,740 users. Our datasets are inherently noisy, as they represent human behavior. For example, the skills from LinkedIn are self-reported. Similarly, tweet content and Facebook posts are a mix of - among others - work-related announcements, private updates, and responses to others. However, this noise is reduced by the fact that our analysis is based on a fairly large collection of users.

4. SKILL NETWORKS

LinkedIn users can list their skills in their profiles. It is a reasonable assumption that basic, more generic skills - such as 'management' - are more often mentioned than more specific skills - such as 'competitive analysis'. Further, one would expect that related skills - such as 'search engine optimization' and 'Web analytics' are often mentioned together, and that subskills are connected to one or two more generic skills - for example, 'Microsoft Word' would be often mentioned together with 'Microsoft Office' and 'Creative Writing'.

To verify whether these assumptions hold in LinkedIn, we visualized the network of skills using the graph visualization software Gephi [4] - see Figure 1, using a force-based layout, with the edge weights determined by how often skills are mentioned together. The four inlays that show parts of the network confirm the above-mentioned assumptions.

The largest node in the network is 'Social Media', which suggests that our sample is dominated by people who are professionally active in social media. Further, the areas surrounding the 'social media hub' have clearly defined sub topics. Top-right from social media are skills that are related to blogging and writing - with a subgroup of graphic design skills. The more technical professions, such as web design and programming are located bottom-right. 'Search engine optimization' forms the bridge to the more marketing-related skills in the left part of the visualization. Top-left is dominated by more traditional management skills, including team building and planning.

4.1 Subgroups in skill networks

The skill network, as displayed in Figure 1, suggests that the LinkedIn network can be divided into skill-based groups, or 'professions'. As explained in the introduction, different professions are expected to have differences in terms of communication behavior, the way people are connected, the topics they talk about, the resources they use, and the way they express themselves.

In order to study topics beyond individual tags and to obtain more context-related information, we additionally employed Latent Dirichlet Allocation (LDA) [5] and modeled each LinkedIn Skills and Expertise tag-based representation of a user as a mixture of latent topics. For this, we used the LDA implementation in the Mallet library[3]. Given a set of term sets (users u_i represented by their Skills and Expertise tags in our case) and the desired number of latent topics, k, LDA outputs the probabilities $P(z_j|u_i)$ that the Skills and Expertise topic z_j is contained (related) in the user profile u_i. In addition, LDA computes term probabilities $P(t_j|z_i)$ for tags t_j; the terms with the highest probabilities for a latent topic z_i can be used to represent that topic. We empirically chose the number of latent topics as 50 for our LinkedIn dataset.

Table 1 shows the top-10 most probable terms for the 50 latent topics (called $professions$ in the next sections), as assigned by the LDA method. In addition, the table contains short topic labels which were manually assigned and will be used throughout the rest of this paper.

5. CONNECTIONS AND ACTIVITIES

In this section, we discuss the results obtained from our analysis of differences in connections and activities between professions. We start with an overview of the differences in connections: which professions are better connected and more active. We continue with an analysis of the differences in topics that users post and tweet about: how generic or specific are these topics? Then we show that the differences in reasons why professions engage in social media have an impact on the sentiment and objectivity of the wording. Finally, we investigate which types of links and resources are shared in different professions.

5.1 Differences in connections

In this section, we look which professions are most and least connected with one another. Based on the insights obtained from the related work, we expect that professions that are in the core of the network are most connected and most active. In order to identify these differences, we took the following features into account:

- **LinkedIn:** We used the number of contacts as an indicator for the connections, no activity information was available.
- **Twitter:** Here, the user connections are based on the followers (incoming links), friends (outgoing links) and presence in lists (curated group of Twitter users). Activity is measured by the number of tweets.
- **Facebook:** As measure for connectivity, we used the number of likes, comments and shares (from friends) on the user's 'wall'. The number of posts of the users himself is an indicator of their activity.

For each of the social networks we created two lists of the top-5 highest and the top-5 lowest values on connections, presented in Figure 2. All professions displayed in this picture appear at least in one of these lists - all others are omitted.

[2]https://developers.facebook.com/docs/graph-api/

[3]http://mallet.cs.umass.edu/

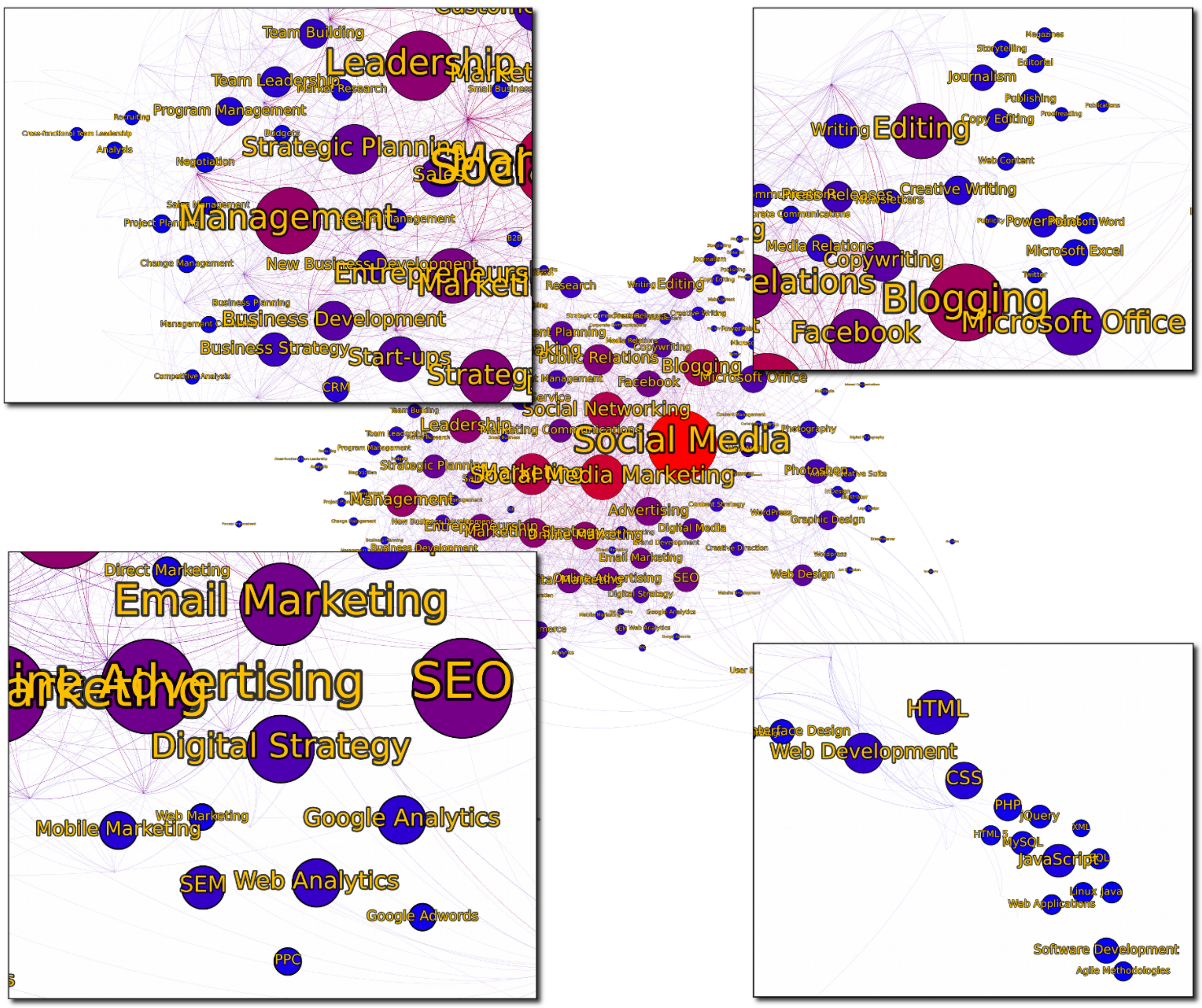

Figure 1: Skill network in LinkedIn. Larger nodes are more often mentioned. Skills that are often mentioned together are closer to one another. The four inlays are close-ups of parts of the network.

As can be seen, several professions have high connectivity scores in more than one network. These include Mobile Devices (actually startups in this field), Entrepreneurs, Marketing and Search Engine Optimization. Low connectivity scores in more than one network are found among Web Programmers, Software Engineers, Pastors, Team Managers and Health and Lifestyle Advisors. In general, the left side mainly contains marketing-oriented professions, the right side IT-oriented professions and 'offline' professions.

In Twitter, the number of followers is highly correlated with the number of friends ($r = .79$) and presence in lists ($r = .87$). The number of followers depends less ($r = .59$) on the number of status updates. Within Facebook, no significant correlations between the number of (public) posts and likes or comments can be found - apparently, Facebook is less 'quantity-driven'.

Interestingly, apart from the marketing-oriented professions, the top-5 professions in terms of status updates (tweets) also includes Content Creators, Journalists and Pastors. These people probably use Twitter for announcements and 'spreading the word', even though - on average - they do not score very high in terms of followers.

5.2 Differences in topics

In order to compare what users of different areas talk about in different networks, we indexed the tweets and Facebook posts of the users into a Solr[4] Index. All the messages were processed trough a standard text processing pipeline, in which we removed stop words and used a stemming algorithm. Beside this, we also removed links from the text as we are only interested in the 'real words' used by a user. For tweets, we also removed the mentions of other users as well as the hash-symbol from hash tags.

This indexing allows us to compute the cosine similarity between different users and different professions. The similarity is calculated using the Solr 'more like this' functionality, which finds documents similar to a given document or a set of documents, based on the terms within the given document. These terms are selected based on their TF/IDF

[4]https://lucene.apache.org/solr/

Table 1: The manually assigned topic labels and the most probable top-10 terms (assigned by the LDA method) for the 50 "Skills and Expertise" (SE) topics.

Topic Label	Top-10 Topic Terms
E-commerce-Strategy	marketing media social digital online strategy advertising analytics web management
Marketing-Strategy	research analysis strategy market product business development strategic competitive innovation
Social Media-Public Relations	media social creative public relations writing editing blogging press releases
Graphic Designer-Hands-On	design creative graphic direction art adobe suite illustration graphics identity
System/Network Administrator	windows server security network administration microsoft system vmware linux networking
Entrepreneur-Startup	business development strategy management start ups strategic entrepreneurship marketing planning
Search Engine Optimization-Tech.	marketing google web analytics online seo search advertising optimization sem
Web Designer-Graphical	web html design css wordpress photoshop adobe development graphic suite
Technical Support-Helpdesk	os mac office microsoft windows computer support technical hardware networking
Game Designer	design game games animation interior architecture video computer development planning
Social Media-'Spammer'/Analyst	social google media facebook twitter wordpress marketing analytics microsoft blogging
Manager or consultant	development community management program writing public leadership outreach planning education
Data Analysis-Programmer	data analysis science engineering research statistics computer design modeling matlab
Customer Management-People	customer management service sales retail team training satisfaction problem solving
Public Relations-International	policy public international research political relations english analysis writing government
Marketing-Events,Press	communications media marketing relations social management public strategic corporate event
Sustainability-Focused,Green	environmental energy management sustainability engineering sustainable construction project awareness water
Software Engineer-Commercial	software management cloud computing enterprise architecture data business integration saas
Financial Analyst	financial management analysis insurance finance planning business banking accounting risk
Marketing-Branding	marketing strategy media digital creative advertising brand social development online
Team Manager,Management	management team planning business project leadership development negotiation analysis strategy
Pastor-Church	pastoral church ministry youth leadership theology preaching studies development teaching
Professional Microsoft Product	microsoft office excel word powerpoint customer research service photoshop management
Marketing-Generic	marketing management media strategy social development advertising online brand business
Medical (Psychiatrists and co)	health healthcare medical clinical research psychology medicine counseling management mental
Beauty Industry	de fashion en styling trend beauty dise merchandising care comunicaci
Marketing-Networking	social marketing media public management event planning relations speaking networking
Marketing-Creator/Blogger	content media social marketing management web digital strategy online development
Web Programmer (#1)	sql net server asp development web microsoft software visual javascript
Manager-Project Planner	management business project process analysis improvement strategy leadership team planning
Real Estate	real estate homes home buyers sales property residential properties investment
Education-Teaching	learning education teaching technology development design curriculum educational training instructional
Creative Writer-Self-Employed	writing editing creative content publishing fiction copy blogging books articles
Web Programmer (#2)	development web html javascript css ruby java mysql php software
Journalist	journalism editing media writing news radio social style broadcast ap
Mobile Devices/Smart Phone	mobile product development devices applications strategy start web ups user
Film and Video Production	video production film editing final pro cut media television producing
Marketing-Generic Online	social media marketing networking online blogging digital web facebook design
IPR Person,Legal Analyst	law legal litigation property writing corporate intellectual research contract civil
Sales Manager	sales management business marketing development strategy selling product strategic account
Health and Lifestyle Advisor	coaching training sports wellness fitness nutrition health lifestyle weight personal
Music and Entertainment	music production audio sound theatre recording entertainment industry acting film
Photo Journalism-Art	photography art digital fine image painting portrait editing portraits photoshop
Hospitality and Tourism	food management hospitality event events travel wine tourism industry beverage
Software Engineer-Management	management software project testing agile analysis requirements quality assurance development
Training and Coaching	development management coaching leadership training business team organizational speaking change
Supply Chain Manager	security management military manufacturing supply chain operations engineering process improvement
Human Resources,Team Manager	skills problem solving communication team leadership thinking creative people building
Recruiter	recruiting management talent recruitment employee human search career resources sourcing
Usability Engineer	design user experience interface web information interaction usability mobile architecture

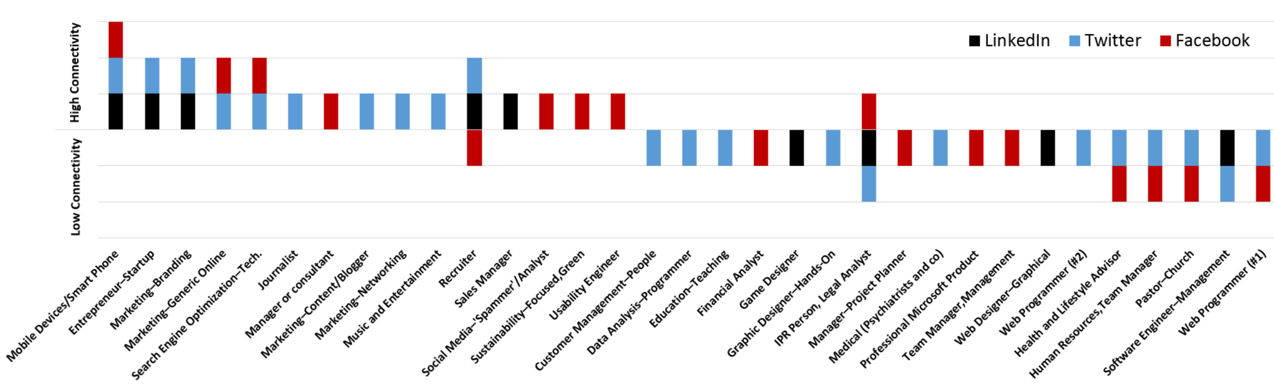

Figure 2: The professions with the highest and lowest connectivity for LinkedIn, Twitter, and Facebook.

values for the given document, which allows us to obtain a representative set of query terms for each user or profession. For all experiments, we selected the 500 most representative terms that occurred within at least five documents.

The different research questions we aim to answer with the experiments described in this section are the following:

- **Mentioning of skills.** Do people use Facebook and Twitter to talk about their professional skills as described in LinkedIn?
- **Similarity between networks.** How similar are users from the profession clusters in Twitter and in Facebook?
- **Specific & general topics.** Which profession clusters are very specific and which are very generic, based on the Facebook posts or Tweets?

To answer these questions, we built queries based on the terms used for the different skill and expertise groups (professions) from LinkedIn. Using these queries, we computed the score for the tweets or posts of every user. For each profession, we calculated the average score. The result of this computation is a matrix that shows how similar the users from the different professions are to the keywords of these professions. These matrices are shown in Figure 3. In order to make the differences better visible, we normalized the results for every query by dividing it by the maximum score. This ensures that the results are within a $[0, 1]$ interval and are comparable for every LinkedIn profession.

The diagonal lines in both diagrams show that most users use Facebook and Twitter to talk about their professional skills. In Twitter the diagonal is stronger than in Facebook, which indicates that Twitter is used for 'professional' communication to a larger extent than Facebook. Inside Twitter, we got an average self similarity (between the same profession cluster in LinkedIn and Twitter) of 0.884, while inside Facebook this values decreases to 0.741.

For answering the second question, how similar users behavior is in Facebook and Twitter, we indexed 50 users from each profession. We chose to use a similar amount of users per profession to remove the influence of the differences in cluster sizes. For each of the selected 2500 users, we computed the similarity to all other users based on the most representative terms used by the user in Facebook and in Twitter. The results are again two matrices, as shown in Figure 4. The matrix on the left uses the most common words in Twitter, the matrix on the right uses the most common words in Facebook. All values are normalized between 0 and 1.

Compared with the first two matrices, the first observation is that the diagonal is missing. This lack of within-cluster overlap indicates that users use Facebook and Twitter for different purposes. A remarkable difference between the two networks is the average similarity between random users: in Twitter, the average similarity is just 0.365 (the predominant green color in the left matrix); in Facebook, the average similarity is 0.818 (the predominant red color in the right matrix). The vertical lines in the left diagram indicate that some groups - particularly Creative Writing, Marketing and Social Media - write about very generic content within Twitter, while other groups use Twitter for more specific (professional) purposes. In summary, this indicates that Facebook is more general-purpose than Twitter, and that most profession clusters use Twitter for profession-specific purposes.

For analyzing the third question - which profession clusters discuss about more specific topics and which about more general topics - some first insights are already given by Figure 3, in which we ordered the LinkedIn profession clusters based on their average similarity to the users inside Twitter and Facebook. We see that professions related to Marketing and Social Media are listed on top in both diagrams, which indicates that the keywords used by these users are more generic and can be found in all professions. The bottom of both diagrams is dominated by technology-related professions as well as pastors, real estate and recruiters. Within these groups, the self-similarity is quite strong, which indicates that users within these professions exchange content-specific information.

We also indexed all messages from all networks and calculated the average similarity of one profession to all other professions, as shown in Figure 5. The blue bars show the similarity based on Facebook query terms and the red bars based on Twitter query terms. For some professions, like *E-commerce-Strategy* or *Usability Engineer*, we see large differences between the two networks. Other professions, like *Marketing*, *Journalist* or *Social Media*, are very general in both networks. The very general professions on the left seem all to be related to areas related to communication and marketing, the more specific professions on the right do not follow a clear scheme. Interesting to see is that many software-related topics are in the average area.

5.3 Differences in sentiment

In this section, we use the SentiWordNet [9] lexicon to study the connection between the users' professions and the sentiment features of tweets and Facebook posts written by these users. SentiWordNet is a lexical resource built on top of WordNet. It contains triples of sentivalues (pos, neg, obj) corresponding to positive, negative or objective sentiment of a word. The sentivalues are in the range of [0, 1] and sum up to 1. For instance (pos, neg, obj) = (0.875, 0.0, 0.125) for the word 'good' and (0.25, 0.375, 0.375) for the word 'ill'.

We assigned a sentivalue to each tweet and Facebook post, in a similar manner as [17, 7], where the authors analyse sentiment in short texts (YouTube comments and Web queries). Similar to the method used in these works, we restrict our analysis to adjectives, as we observed the highest accuracy in SentiWordNet. Finally, we computed the average positivity, negativity and objectivity over all tweets and Facebook posts that belong to a profession.

Table 2 shows the top-5 most positive, negative and objective professions with respect to user-expressed sentiments in Facebook posts and tweets. The users with skills in computer technical support and data analysis programmers tend to post the most negative messages in both Facebook and Twitter. Their posts or tweets often offer or request help for problems, i.e., *'@user Sounds like a hard drive issue. Either it's hitting bad sectors or the drive has literally slowed down and is having read/write issues'*. On the other side, users related to human resources, logistics and health, as well as lifestyle advisors post the most positive content in our collection. Some hand-picked examples from Twitter include *'Best food moments of 2013 #food http://t.co/JdYO36wVAY'*, *'Kids Eat and Stay Free at the Holiday Inn Washington DC. Bring the entire family for a holiday trip http://t.co/RhYa3zuHgu'*.

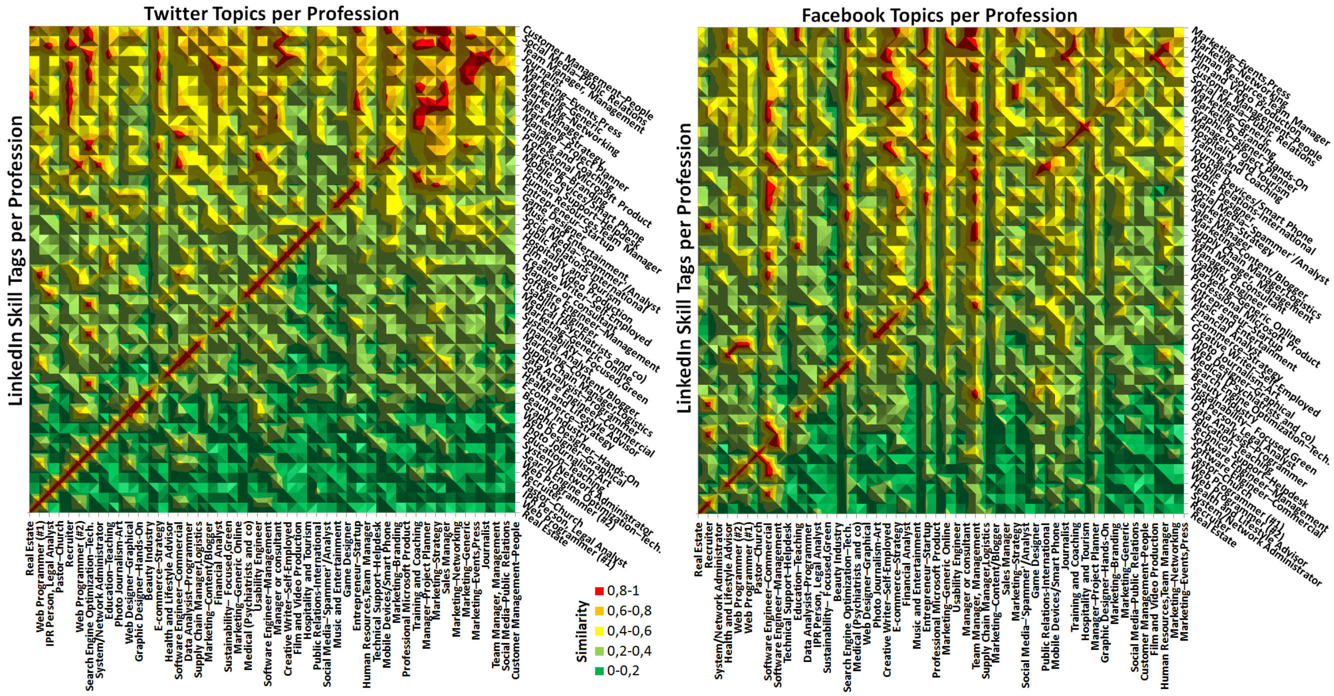

Figure 3: Similarity of skill tags from LinkedIn and terms used in Twitter (left) and Facebook (right). Similarities are summarized per profession.

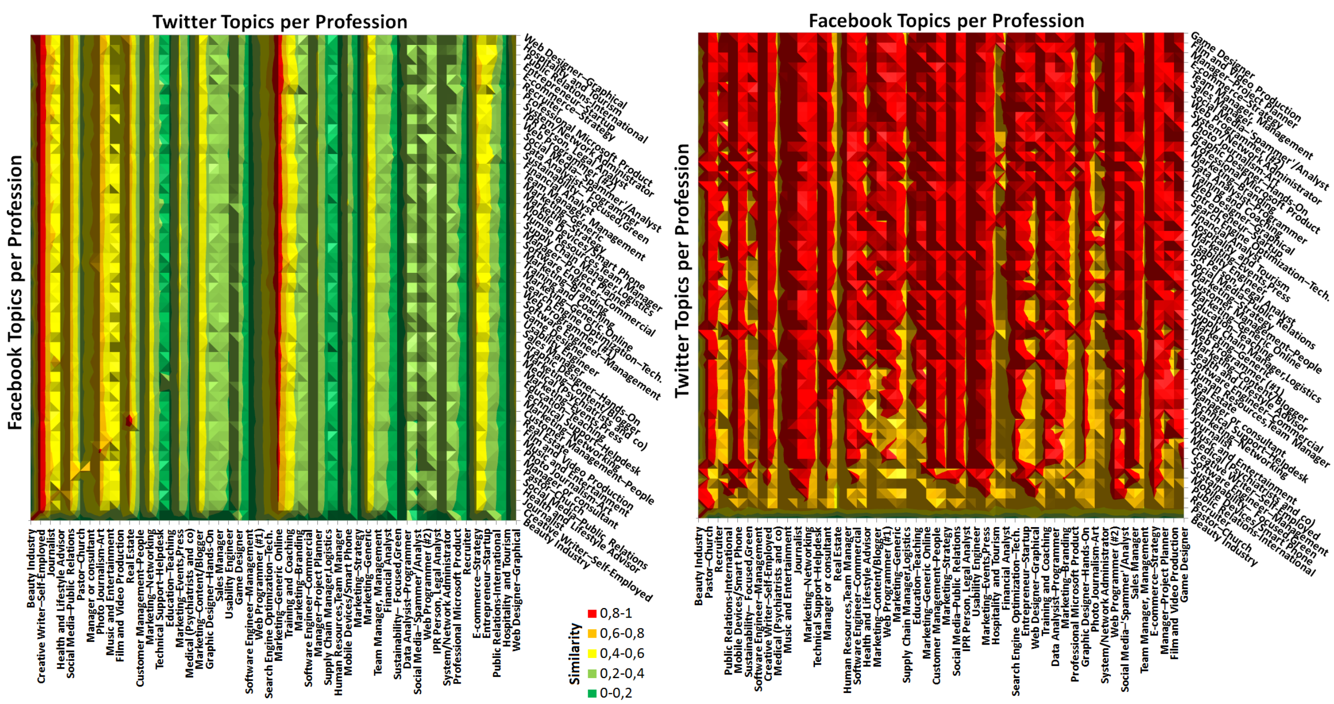

Figure 4: Similarity between topics that users talk about in Twitter (left) and Facebook (right), grouped by professions.

We also observed that users tend to be more objective in Twitter than Facebook, particularly for some of the professions. For instance, the average objectivity for *Pastors-Church* is up to 14% higher in Twitter than in Facebook. Many of the Facebook messages posted by users belonging to this profession express sympathy or commendation

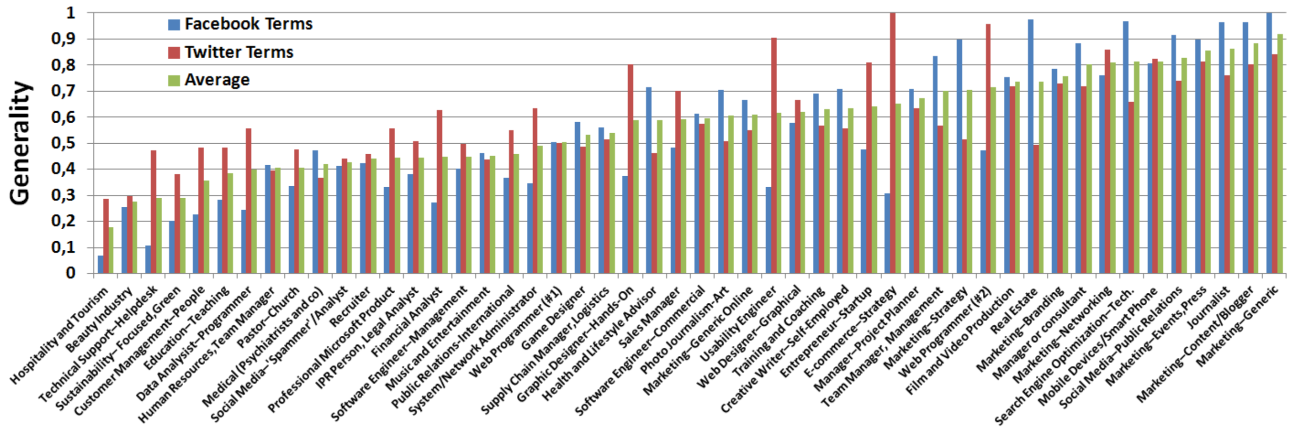

Figure 5: Comparison of generality of communication in different professions, based on terms from both Facebook and Twitter. Generality is the average similarity to all other professions.

Table 2: Top-5 most positive/negative/objective professions w.r.t. user-expressed sentiments in Facebook and Twitter.

Facebook	Twitter
Positive	
Supply Chain Manager,Logistics	Human Resources,Team Manager
Technical Support-Helpdesk	Hospitality and Tourism
Medical (Psychiatrists and co)	Health and Lifestyle Advisor
IPR Person,Legal Analyst	Customer Management
Pastor-Church	Marketing-Branding
Negative	
Pastor-Church	Technical Support-Helpdesk
Technical Support-Helpdesk	System/Network Administrator
Training and Coaching	Social Media-'Spammer'/Analyst
Film and Video Production	Human Resources,Team Manager
Data Analysis-Programmer	Journalist
Objective	
Manager-Project Planner	Recruiter
Recruiter	Public Relations-International
Team Manager,Management	Team Manager,Management
Beauty Industry	IPR Person,Legal Analyst
Professional Microsoft Product	Education-Teaching

towards a religious topic or event, such as: *'2013 EVANGELICAL HEALING CONVENTION "Arise, Go, Preach" (Jonah 3:2)'* or a religious greeting *'May God bless your day as you display responsible actions and superior performance'*.

The differences in sentiment between the different skills and expertise groups may reflect that people in some professions are more positive or negative in general, or that they tend to formulate their messages more positively or negatively. Our interpretation, however, is that the differences in sentiment are largely caused by differences in intentions of tweeting.

The most positive groups are professions that use social media for selling and promoting items and events; it seems natural that these promotional messages are positive and motivating. On the other hand, the most negative group consists of people who work individually on programming or writing tasks. We expect that these people mainly use social media for asking and providing help for problems and issues that they encounter. The least objective - or most subjective - topic groups mainly consist of people who provide advice and coaching in areas such as religion, health and lifestyle and entertainment. Most likely, these are people who aim to spread a particular message or opinion.

5.4 Differences in linked and shared content

Nowadays, a vast amount of content is shared by users through various social platforms. A recent study [18] shows that 71% of online users have shared some type of content on social media sites. The most popular shared items usually refer to a picture, an opinion/status update or a link to an article. Another user study [6] looks into the main motivation for sharing items, showing that most of the users (94%) carefully consider the usefulness of their shared content for the readers. While all of these recent studies imply the importance of users' shared content, there is no work that systematically investigates the link sharing patterns based on the users different expertise skills. We believe that our findings unleash the potential of analyzing users' shared links, which is a rather overlooked source of information up to now.

In this section, we first provide an overview on the amount of link-based content shared by different experts in their tweets and posts. Next, we investigate the type of content shared by different experts, by looking into the main web-domains extracted from the shared links.

Figure 6 shows the percentage of Facebook posts and tweets that contain links for each profession group. In Facebook, 60.97% of the posts share a link. Users belonging to the *Sustainability-Focused,Green*, *IPR Person,Legal Analyst* and *Public Relations-International* professions are most likely to post links. In contrast, software engineers, pastors or market strategists are less likely to include URLs in their posts. In Twitter, Web programmers and software professionals attach links less frequently. On the other side, real estate experts, photo-journalists and health care advisors contribute with a considerably higher amount of links across their tweets. This is in line with our observation in Section 5.1 that these professions make use of social media for posting announcements. Overall, 54.76% from the tweets in our collection contain a link.

As an illustrative example, we computed ranked lists of web-domains from a set of tweets and posts belonging to the top-3 and bottom-3 most active web-domain sharers in our dataset. For ranking the resulting web-domain terms, we used the Mutual Information measure [16, 19] from information theory, which can be interpreted as a measure of

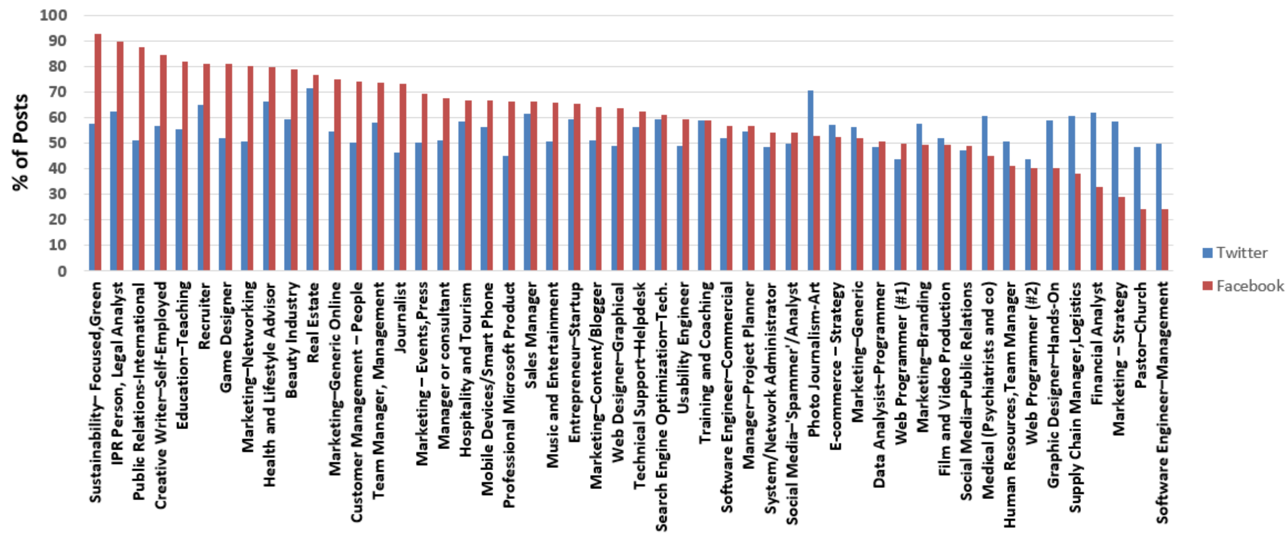

Figure 6: Percentage of Facebook posts and tweets sharing links, for each profession.

how much the joint distribution of features X_i (web-domain terms in our case) deviate from a hypothetical distribution in which features and categories (a specific profession versus all 'other' professions, in our case) are independent from each other. Table 3 shows the top-10 web-domains extracted from the links shared within: 1) the posts written by users belonging to top-3 and bottom-3 Facebook profession groups, based on the link-sharing frequency and 2) the tweets written by users belonging to top-3 and bottom-3 Twitter profession groups, based on the link-sharing frequency.

Different profession groups tend to prefer linking different type of content across their messages. In our collection, the most shared links refer to a Social Network. For instance, in Twitter, Web programmers show a preference for *foursquare.com* (a platform for discovering friends' best locations), while real estate users link a vast amount of Facebook content. Note that, while Table 3 indicates noticeable differences in the preference towards different Social Networks, analyzing the underlying reasons for such differences is beyond the scope of this study.

At the same time, people tend to include links related to their expertise domains, i.e., *activerain.com*, *houselogic.com* for Real-Estate users and *arstechnica.com*, techcrunch.com for Web Programmers. For Facebook, we noticed that most of the shared web-domains seem to be less connected to the user's profession.

6. DISCUSSION

In this paper, we investigated differences in communication and connection practices between professions, as represented by the skill and expertise groups that we extracted from a representative dataset.

In our analysis, we used a combination of exploratory analysis, visualization and interpretation. These methods are not suitable for drawing strong conclusions on the exact structure and growth of communities and the interactions between the members. Among others, Kumar et al [13] investigated these aspects as well. Our aim was to provide a complementary view on these structures and to give some insight in the people, professions and conversation topics that constitute these structures. Necessarily, these insights are partially given by means of representative examples. Keeping this limitation in mind, there are several key insights that can be drawn from the results.

In professional networks, connections between people based on shared skills follow the same structure as explicit connections, such as following, endorsing or befriending in social networks. The majority of mentioned skills are quite detailed and closely connected to a frequently mentioned more generic skill. By separating the skill network into clusters, skill and expertise groups - or professions - can be recognized.

The core of the skill network mainly consists of people who professionally use social media for specific purposes, such as marketing, promoting, branding and recruiting. These persons are typically well-connected, talk about common topics, share links from common resources and usually have a positive tone.

By contrast, several niche groups that are further away from the core are typically less connected and centered around a particular representative skill. Professions in which (individual) productivity is more important than communication - such as programming and writing - seem to use social networks predominantly for specific purposes, such as providing or asking for help or feedback. Due to this different intention of use, the activity level, the topics discussed and the resources shared differ highly from what happens in 'the core'.

These observations have clear implications for social network analysis, particularly for professional networks. Firstly, it is clear that averages for the whole population - and interpretation of these averages - are often only meaningful for the central core. The dynamics in subgroups are in many cases quite different - based on our qualitative evidence mainly caused due to differences in intention of use.

Zooming into the topics and links that are specific for a subgroup, and providing these to users who are new to the community or who aim to connect to it, seems to be a promising approach to get these users acquainted with the community and to get a feeling on the unwritten conventions

Table 3: Top-10 web-domains according to their Mutual Information values for tweets/posts written by users belonging to "One" profession vs. "Other" professions.

Top-10 distinctive web-domains for top-three professions, according to their % of links.					
Twitter			Facebook		
Real Estate	**Photo Journalism**	**Health and Lifestyle Advisor**	**Sustainability Focused,Green**	**IPR Person, Legal Analyst**	**Public Relations âĂŞ International**
facebook.com	facebook.com	facebook.com	facebook.com	facebook.com	apps.facebook.com
foursquare.com	instagram.com	networkedblogs.com	change.org	dangerousminds.net	nytimes.com
youtube.com	etsy.com	youtube.com	ulink.tv	politicususa.com	npr.org
paper.li	zazzle.com	graph.facebook.com	elpais.com	addictinginfo.org	nyti.ms
activerain.com	plus.google.com	articles.mercola.com	youtube.com	huffingtonpost.com	youtube.com
yelp.com	about.me	paper.li	avaaz.org	alternet.org	washingtonpost.com
trulia.com	vimeo.com	ebay.com	librarything.com	thinkprogress.org	salon.com
inman.com	blipfoto.com	amazon.com	europapress.es	forwardprogressives.com	behance.net
houselogic.com	post.ly	about.me	actuable.es	dailykos.com	change.org
scoop.it	fineartamerica.com	fitbit.com	zimbio.com	fab.com	i.imgur.com

Top-10 distinctive web-domains for bottom-three professions, according to their % of links.					
Twitter			Facebook		
Web Programmer #1	**Web Programmer #2**	**Profesional Microsoft Product**	Software Engineer **Management**	**Pastor-Church**	**Marketing Strategy**
foursquare.com	foursquare.com	instagram.com	youtube.com	apps.facebook.com	nike.com
youtube.com	twitter.com	foursquare.com	apps.facebook.com	ludia.com	buff.ly
fancy.com	youtube.com	twitter.com	facebook.com	barackobama.com	youtube.com
getglue.com	i.imgur.com	twittascope.com	nblo.gs	instagr.am	tripit.com
blogs.msdn.com	techcrunch.com	youtube.com	bbc.co.uk	gofundme.com	groupon.com
arstechnica.com	theverge.com	plurk.com	livingsocial.com	facebook.com	act.credoaction.com
engadget.com	twitpic.com	justunfollow.com	ludia.com	eventbrite.ca	secure.sierraclub.org
path.com	twitter.yfrog.com	gofundme.com	meetup.com	amzn.to	generalassemb.ly
fplus.me	plurk.com	runkeeper.com	mashable.com	amzn.com	gr.pn
techcrunch.com	meetup.com	infojobs.net	amazon.co.uk	itunes.apple.com	animoto.com

and rules within these communities. In addition, the group-specific resources - such as technology-oriented websites - often serve as a useful starting point for exploring a new expertise area. These insights can be used as starting points for new browsing and search functionality in professional networking sites.

7. CONCLUSION

Within a skill network, several subgroups - or professions - centered around a particular skill can be recognized. Our analysis shows that these subgroups have specific unwritten conventions and rules, mainly caused by differences in intention for using social media. These insights call for separate analysis or treatment of activities within these subgroups, and provide several starting points for new functionality in professional networking sites.

8. ACKNOWLEDGMENTS

This work was partially supported by EU FP7 projects ForgetIT (Contract No. 600826) and QualiMaster (Contract No. 619525).

9. REFERENCES

[1] F. Abel, E. Herder, G.-J. Houben, N. Henze, and D. Krause. Cross-system user modeling and personalization on the social web. *User Modeling and User-Adapted Interaction*, 23(2-3):169–209, 2013.

[2] F. Abel, E. Herder, and D. Krause. Extraction of professional interests from social web profiles. *Proc. Augmented User Modeling at UMAP*, 34, 2011.

[3] K. Balog and M. de Rijke. Finding experts and their details in e-mail corpora. In *Proceedings of the 15th international conference on World Wide Web*, pages 1035–1036. ACM, 2006.

[4] M. Bastian, S. Heymann, and M. Jacomy. Gephi: an open source software for exploring and manipulating networks. In *ICWSM*, pages 361–362, 2009.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, Mar. 2003.

[6] B. Brett. The psychology of sharing, 2012. Available at `http://nytmarketing.whsites.net/mediakit/pos/`.

[7] S. Chelaru, I. S. Altingovde, S. Siersdorfer, and W. Nejdl. Analyzing, detecting, and exploiting sentiment in web queries. *ACM Trans. Web*, 8(1):6:1–6:28, Dec. 2013.

[8] F. Cingano and A. Rosolia. People i know: job search and social networks. *Journal of Labor Economics*, 30(2):291–332, 2012.

[9] A. Esuli and F. Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC '06)*, pages 417–422, 2006.

[10] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi. Cognos: Crowdsourcing search for topic experts in microblogs. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 575–590, New York, NY, USA, 2012. ACM.

[11] I. Guy, U. Avraham, D. Carmel, S. Ur, M. Jacovi, and I. Ronen. Mining expertise and interests from social media. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, pages 515–526, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[12] L. Hong, G. Convertino, and E. H. Chi. Language matters in twitter: A large scale study. In *ICWSM*, 2011.

[13] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In P. S. Yu, J. Han, and C. Faloutsos, editors, *Link Mining: Models, Algorithms, and Applications*, pages 337–357. Springer New York, 2010.

[14] J. Liu, F. Zhang, X. Song, Y.-I. Song, C.-Y. Lin, and H.-W. Hon. What's in a name?: An unsupervised approach to link users across communities. In

Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, WSDM '13, pages 495–504, New York, NY, USA, 2013. ACM.

[15] C. Macdonald, D. Hannah, and I. Ounis. High quality expertise evidence for expert search. In *Proceedings of the IR Research, 30th European Conference on Advances in Information Retrieval*, ECIR'08, pages 283–295, Berlin, Heidelberg, 2008. Springer-Verlag.

[16] C. Manning and H. Schuetze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[17] S. Siersdorfer, S. Chelaru, W. Nejdl, and J. San Pedro. How useful are your comments?: Analyzing and predicting youtube comments and comment ratings. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 891–900, New York, NY, USA, 2010. ACM.

[18] J. Wiltfong. Majority (71%) of global internet users "share" on social media sites, 2013. Available at `http://ipsos-na.com/news-polls/pressrelease.aspx?id=6254`.

[19] Y. Yang and J. O. Pedersen. A comparative study on feature selection in text categorization. pages 412–420. Morgan Kaufmann Publishers, 1997.

[20] D. Yimam-Seid and A. Kobsa. Expert-finding systems for organizations: Problem and domain analysis and the demoir approach. *Journal of Organizational Computing and Electronic Commerce*, 13(1):1–24, 2003.