Discovering Topical Experts in Twitter Social Network

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Technology

in

Computer Science and Engineering

by

Naveen Kumar Sharma
07CS3027

supervised by

Dr. Niloy Ganguly

Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur
April, 2012

This Thesis is Dedicated to

My Parents
without whom nothing would have been possible...

and to
All my Teachers
for bringing me where I am today...

HATD

Handling and Archiving of Theses and Dissertations

submitted to the

Indian Institute of Technology, Kharagpur 721302

Declaration by the Author of the Thesis or Dissertation

I, **Naveen Kumar Sharma**, Roll No. **07CS3027** registered as a student of program **Integrated B.Tech and M.Tech (Dual Degree)** in the **Department of Computer Science and Engineering** Indian Institute of Technology, Kharagpur, India (hereinafter referred to as the Institute) do hereby submit my thesis, title: **"Discovering Topical Experts in Twitter Social Network"** (hereinafter referred to as my thesis) in a printed as well as in an electronic version for holding in the library of record of the Institute.

I hereby declare that:

1. The electronic version of my thesis submitted herewith on CD-ROM is in **PDF** format.
2. My thesis is my original work of which the copyright vests in me and my thesis does not infringe or violate the rights of anyone else.
3. The contents of the electronic version of my thesis submitted herewith are the same as that submitted as final hard copy of my thesis after my viva voce and adjudication of my thesis on **30th April 2012**.
4. I agree to abide by the terms and conditions of the Institute Policy on Intellectual Property (hereinafter Policy) currently in effect, as approved by the competent authority of the Institute.
5. I agree to allow the Institute to make available the abstract of my thesis in both hard copy (printed) and electronic form.
6. For the Institutes own, non-commercial, academic use I grant to the Institute the non-exclusive license to make limited copies of my thesis in whole or in part and to loan such copies at the Institutes discretion to academic persons and bodies approved of from time to time by the Institute for non-commercial academic use. All usage under this clause will be governed by the relevant fair use provisions in the Policy and by the Indian Copyright Act in force at the time of submission of the thesis.
7. Furthermore,
   - I **agree** to allow the Institute to place such copies of the electronic version of my thesis on the private Intra-net maintained by the Institute for its own academic community.

- I **agree** to allow the Institute to publish such copies of the electronic version of my thesis on a public access website of the Internet should it so desire.

8. That in keeping with the said Policy of the Institute I agree to assign to the Institute (or its Designee/s) according to the following categories all rights in inventions, discoveries or rights of patent and/or similar property rights derived from my thesis where my thesis has been completed:

   - with use of Institute-supported resources as defined by the Policy and revisions thereof,
   - with support, in part or whole, from a sponsored project or program, vide clause 6(m) of the Policy.

   I further recognize that:

   - All rights in intellectual property described in my thesis where my work does not qualify under sub-clauses 8(a) and/or 8(b) remain with me.

9. The Institute will evaluate my thesis under clause 6(b1) of the Policy. If intellectual property described in my thesis qualifies under clause 6(b1) (ii) as Institute-owned intellectual property, the Institute will proceed for commercialization of the property under clause 6(b4) of the Policy. I agree to maintain confidentiality as per clause 6(b4) of the Policy.

10. If the Institute does not wish to file a patent based on my thesis, and it is my opinion that my thesis describes patentable intellectual property to which I wish to restrict access, I agree to notify the Institute to that effect. In such a case no part of my thesis may be disclosed by the Institute to any person(s) without my written authorization for one year after the date of submission of the thesis or the period necessary for sealing the patent, whichever is earlier.

Naveen Kumar Sharma — **Dr. Niloy Ganguly**
(Name of Student) — (Name of Supervisor)

(Signature of Student)
Department of Computer Science and Engineering

Signature of the Head of the Department

Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur
April, 2012

Advisor's Certificate

This is to certify that the thesis entitled **"Discovering Topical Experts in Twitter Social Network"** submitted by **Naveen Kumar Sharma** (Roll No. **07CS3027**) to the Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur, in partial fulfillment of the requirements for the award of **Master of Technology** is a bonafide record of research work performed by him under my supervision and guidance.

In my opinion, this work fulfills the requirements as per the regulations of this Institute and has reached or exceeded the standard needed for submission. This thesis has not been submitted to any other university or institution for any degree or diploma.

Dr. Niloy Ganguly
Associate Professor
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur
April 2012

Acknowledgements

During the course of the past one year, there have been several people who have helped me finish this thesis, without whom it would have been incomplete. First, I would like thank my advisor Dr. Niloy Ganguly for providing me the opportunity to work under his supervision. He has been a wonderful guide throughout and has been very encouraging and supportive. His courses on "Complex Network Theory" and "Information Retrieval" have played a very important part in my understanding of this field. His critical comments and long discussions with him were very vital for successful completion of this project. He ensured that the work was carried out at the highest possible standard for which I am very grateful to him.

I would also like to immensely thank Dr. Krishna Gummadi and Saptarshi Ghosh for working along with me and acting as great mentors. They have always helped me out at crucial times, whenever I got stuck. I'll especially remember the night-outs we put together in the lab to meet conference deadlines. It was an enriching experience working along-side them and having those brain-storming sessions.

Next, I'd like to express my gratitude towards my parents whose tireless efforts have brought me up-to where I am today. They have taught me the value of hard work and perseverance, both of which have been invaluable during the past year. Many thanks go to my elder sister Sunita who has always been encouraging and pushes me that *one* step further. Special thanks to my brother Praveen who has been more like a godfather to me. He first sets benchmarks for me and then teaches me how to break them. I'm also grateful to all my friends and teachers at IIT Kharagpur without whom life would not have been the same.

Abstract

Finding topic experts on microblogging sites with millions of users, such as Twitter, is a hard and challenging problem. In this thesis, we propose and investigate a new methodology for inferring topical experts in the popular Twitter social network. Our methodology relies on the wisdom of the Twitter crowds – it leverages Twitter Lists, which are often carefully created by individual users to include experts on topics that interest them and whose meta-data (List names and descriptions) provide valuable semantic cues to experts' domain of expertise. We mined List information to build Cognos, an expert search system for Twitter. Detailed experimental evaluation based on a real-world deployment shows that: (a) Cognos infers a user's expertise more accurately and comprehensively than state-of-the-art systems that rely on the user's bio or tweet content, (b) Cognos scales well due to built-in mechanisms to efficiently update its experts' database with new users, and (c) Despite relying only on a single feature, namely crowd-sourced Lists, Cognos yields comparable, if not better, results in user tests, as compared to the official Twitter experts search engine for a wide range of queries. Our study highlights Lists as a potentially valuable source of information for future content or expert search systems in Twitter.

Contents

Acknowledgements **vi**

Abstract **vii**

Table of Contents **viii**

List of Figures **ix**

List of Tables **x**

1 Introduction **1**

1.1 Contributions of this Thesis 3

1.2 Thesis Organisation 4

2 Background **5**

2.1 Twitter Online Social Network 5

2.2 Related Work 6

3 Inferring Expertise **9**

3.1 Methodology 9

3.2 Quality of Inferred Expertise 11

3.3 Comparison with Existing Services 12

3.3.1 State-of-the-art research 12

3.3.2 Twitter's official WTF service 14

3.4 Coverage of Experts 16

3.5 Summary 18

4 Cognos Search System **19**
4.1 Ranking Experts 19
4.2 Building Experts Collection 20
4.2.1 Crawling Lists 20
4.3 Evaluating Cognos 21
4.4 Comparison with Existing Services 25
4.4.1 State-of-the-art research 26
4.4.2 Twitter's official WTF service 26
4.5 Summary 30

5 Finding New Experts **31**
5.1 Scalability problem with crawling Lists 31
5.2 Crawling experts efficiently 32
5.3 Evaluating coverage of our crawls 33
5.3.1 Coverage of most listed users 33
5.3.2 Coverage of newly joined experts 34
5.3.3 Coverage of experts identified by other systems 35

6 Conclusions **37**
6.1 Brief Summary 37
6.2 Future Work 38

References **40**

List of Figures

3.1 Fraction of users who are listed at least L times vs follower-rank 17

4.1 Distribution of the number of times a query was asked (out of 55 sample queries) 22

4.2 Screenshot of evaluation of relevance of results 23

4.3 Fraction of evaluations that judged a result relevant (across all queries) – shown for each individual rank out of top 10 . . 23

4.4 Fraction of evaluations / individual results that were judged relevant 24

4.5 Fraction of evaluations / individual results that were judged relevant – considering only those results for a query, which were evaluated at least twice 25

4.6 Screenshot of comparison of Cognos vs Twitter results 28

5.1 Fraction of estimated number of experts who are included in at least K Lists, that is discovered in the hub-based crawl, for $K = 10$, 100, 1000. 34

5.2 Distribution of the number of days elapsed since creation of accounts in random sample and those discovered in hub-based crawl (considering users listed at least 100 times) 35

5.3 Distribution of the fraction of the Twitter WTF top 20 results that is covered in our hub-based crawl 36

List of Tables

3.1 Examples of Lists, their description, and some members . . . 9

3.2 The top tags of expertise of some popular Twitter users, as identified using Lists 12

3.3 Some of the top results reported by Pal *et. al.* [16], for whom the topics inferred using Lists include the query-topic. 13

3.4 Top 10 results reported by Pal *et. al.* [16], for whom the topics inferred using Lists does *not* include the query-topic . 14

3.5 Examples of (i) users for whom topics inferred contain similar words but not the exact query-word (ii) wrong results within Twitter WTF top 20 results. 15

3.6 Examples of users who are related to niche topics, having intermediary follower-ranks (between 1 million and 10 million) 17

4.1 The sample queries used for evaluation of Cognos. 22

4.2 Top 5 results by Cognos and by Pal *et. al.* [16] for the three queries evaluated by Pal *et. al.*, along with their bio and number of followers 27

4.3 Evaluator-chosen queries for comparison of Cognos and Twitter WTF, where the verdict is given by majority voting. . . . 28

4.4 Top 5 results by Cognos and by Twitter WTF for the queries "music" and "windows phone". 29

5.1 Examples of very recently created expert accounts discovered by our Hub-based crawl (which ended on Feb 8, 2012) 34

Chapter 1

Introduction

Microblogging sites, out of which Twitter is the most popular, have emerged as an important platform for exchanging real-time information on the Web. Recent estimates suggest that 200 million active Twitter users post 150 million tweets (messages) daily [1, 24]. These messages contain a wide variety of information, varying from conversational tweets to highly relevant information on niche topics. The users posting these messages range from globally popular news organizations and celebrities to locally popular community organizers or activists and from domain experts in fields like computer science and astrophysics to spammers that fake the identities of well-known users.

As a result, the quality of information posted in Twitter is highly variable and finding the users that are recognized sources of relevant and trustworthy information on specific topics (i.e. topical experts) is a key challenge. Identifying topic experts is also the first step towards finding authoritative information on the topic. Recognizing this, Twitter itself has created a topical expert search system (known as the Twitter Who To Follow (WTF) service [22]). However, as we show later in this Thesis, the results from this service leave a lot of scope for improvement.

In this Thesis, we present **Cognos**, a system for finding topic experts in Twitter. Cognos is based on a new methodology for inferring users' expertise. Traditional approaches to identify topical experts in Twitter rely either on the information provided by the user herself (e.g., user bio) [23] or on analyzing the network characteristics and tweeting activity of users [26, 16]. Cognos takes a different approach to identify topical experts in Twitter utilizing *crowd-sourced* topical annotation of experts. Specifically, Cognos exploits the *Lists* feature in Twitter, using which any user can group Twitter accounts that tweet on a topic that is of interest to her, and follow their collective tweets. We observe that many users carefully create Lists to include other Twitter users who they consider as experts on a given topic. Furthermore, they generate meta-data, such as List names and descriptions, that provide valuable semantic cues to the topical expertise of the users included in the List. Our key idea is to analyze the meta-data of the Lists containing a user to infer the user's topics of expertise, which in turn enabled us to identify topical experts.

To build Cognos, we address three key challenges:

1. How to accurately and comprehensively infer individual user's topics of expertise from Lists?

2. How to rank the relative expertise of different users identified as experts on a given topic? and

3. How to crawl the Lists meta-data for hundreds of millions of Twitter users efficiently and scalably?

The main contributions of this work lie in the methodologies we propose to tackle the above challenges.

We present an extensive evaluation of Cognos based on user feedback obtained using a real-world deployment, which can be accessed at `http://twitter-app.mpi-sws.org/whom-to-follow/`. To summarize a few highlights from our evaluation: We find that Cognos performs as good as or better than the official Twitter WTF service in more than 52% of the queries. Cognos yields particularly better search results in cases in which experts do not have an account bio, or whose bio does not contain information about the user's topic of expertise. Moreover, Cognos rarely produces entirely irrelevant results, unlike the Twitter WTF service whose top results at times include a few users who are not related to the given query, but whose name or bio contains the terms in the query. Furthermore, as Cognos is based on the use of a single and simple feature (Twitter Lists) it is far more scalable as compared to prior approaches, which use computationally intensive machine learning algorithms over graph and content-based metrics [26, 16].

1.1 Contributions of this Thesis

This work has three major contributions towards discovering topical authorities in Twitter Social Network.

First we describe a novel method for inferring attributes that characterize individual Twitter users. As opposed to existing methods which attempt to infer topics for a user from the contents of the users tweets or profile, we infer topics by leveraging the wisdom of the Twitter crowds, as reflected in the meta-data (names and descriptions) of Lists created by the crowds. We used the proposed topic inference methodology to construct a who-is-who service for Twitter and showed that our service can automatically infer an accurate and comprehensive set of attributes for over a million Twitter users, including most of the popular users.

Next, based upon this methodology, we build and deploy Cognos, a topical expert search system. Our evaluation of the Cognos search system shows that a vast majority of its results are relevant for a wide variety of topics. In fact, Cognos rarely produces irrelevant results for user queries. Comparing Cognos with state-of-the-art research system by Pal *et. al.* and official Twitter WTF service highlights the advantages of relying on crowd-sourced Lists to identify experts. Cognos yields particularly better search results in the cases when the bio or tweets posted by a user does not correspond to or contain information about the user's topic of expertise. In fact, Cognos performs as good as or better than the official Twitter WTF service for more than 52% of the queries, even though it is based on a single and simple feature (Lists).

Lastly, we come up with an efficient solution to address the practical challenge of keeping our Cognos system up-to-date, even as hundreds of thousands new Twitter accounts and new Lists are created every day. We show that a *hub-based strategy* – periodically discovering experts through the Lists created by top hubs – can be used to efficiently discover newly joined experts (even very recently joined ones), and thus keep an expert search system up-to-date in the face of rapid increase in the Twitter population.

1.2 Thesis Organisation

Chapter 1 gives an introduction to the problem chosen and the motivation behind it. It describes at a high level, the challenges we address and the methodology we apply to solve these. The outline of the thesis and major contributions are also mentioned.

Chapter 2 introduces the reader to the Twitter online social network and related background. We review the existing approaches that have been proposed to solve similar problems and challenges. We point out fundamental differences between our approach and the existing methods.

Chapter 3 describes our methodology for finding topic experts using a recently introduced Twitter feature called Lists. We explain our method of inferring the expertise of individual Twitter users and then evaluate the accuracy and expressiveness of the inferred expertise. We also do a head-on comparison with the state-of-the-art research method and the official Twitter service.

Chapter 4 presents *Cognos*, a search system for topical experts in Twitter which leverages our previously discussed methodology to infer users' expertise. Cognos uses crowd-sourced Lists as the *only* source of information and hence its performance illustrates the potential uses of Lists in finding experts. We describe how we rank experts in Cognos and then present an extensive evaluation of the Cognos system.

Chapter 5 addresses the practical challenge of keeping our Cognos system up-to-date, even as hundreds of thousands new Twitter accounts and new Lists are created every day. We describe an efficient method which crawls recently joined "popular" experts on Twitter without incurring a lot of overhead in terms of API requests to Twitter.

Chapter 6 is the concluding chapter which summarizes the work done and points out several future directions and challenges that need to be solved.

Chapter 2

Background

Recently, the Twitter microblogging site has emerged as an important source of real-time information on the Web. Millions of users with varying backgrounds and levels of expertise post about topics that interest them. As a result, the quality of information posted in Twitter is highly variable and finding the users that are recognized sources of relevant and trustworthy information on specific topics (i.e. topical experts) is a key challenge. Identifying topic experts is also the first step towards finding authoritative information on the topic.

2.1 Twitter Online Social Network

Twitter is an online social networking service and microblogging service that enables its users to send and read text-based posts of up to 140 characters, known as "tweets". It was created in March 2006 by Jack Dorsey and launched in July the same year. The service rapidly gained worldwide popularity, with over 140 million active users as of 2012, generating over 340 millions tweets daily and handling over 1.6 billion search queries per day [2].

Twitter is primarily an "interest" based social network. Users either join Twitter to *speak* about things that they are interested in, or to *listen* to others who are talking on topics which they are interested in. Tweets are publicly visible to everyone on the web by default; however, senders can restrict message delivery to just their followers by making their accounts *private*. Users can tweet via the Twitter website, compatible external applications (such as Tweetdeck or Echofon), or by Short Message Service (SMS). Users may subscribe to other users' tweets, which is known as "following" and the subscribers are known as followers. Following someone implies that all of his/her tweets will be visible on your personal Twitter homepage – also known as *home time-line*.

Renowned technology author Steven Johnson describes the basic mechanics of Twitter as "remarkably simple" in the following words.

> "As a social network, Twitter revolves around the principle of followers. When you choose to follow another Twitter user, that user's tweets appear in reverse chronological order on your main Twitter page. If you follow 20 people, you'll see a mix of tweets scrolling down the page: breakfast-cereal updates, interesting new links, music recommendations, even musings on the future of education."

Today, Twitter contains numerous celebrities, politicians, sports-person, news-media outlets, bloggers, organisations and experts on a wide array of topics. There are even more users who just use Twitter as a medium to follow these "popular" users. Since inception, Twitter has been used for a variety of purposes in many industries and scenarios. Most notable examples include 2010-11 Tunisian protests, 2011 Egyptian revolution and the 2011 Japanese Earthquake.

As a result, Twitter has become a fertile playground for various measurement and analysis studies. There are several interesting challenges and problems which have come up over the past few years, and this thesis tries to solve one of them.

2.2 Related Work

As the number of users and information shared in Twitter has increased exponentially, different information retrieval tools, such as search [20] and recommender systems [22], are becoming very popular ways to find trend topics, users, and valuable content. A critical component of such mechanisms consists of identifying users who are important sources of information on specific topics (topical experts).

To facilitate information search in such platforms, it is useful and important to characterize the sources of information, e.g., infer semantic topics for pages or sites in the Web or infer attributes of users in Twitter. A lot of prior research has focused on discovering semantic topics for web-pages. Machine learning techniques have been applied over the contents of web-pages to automatically annotate the pages with their semantic topics [8]. It has also been shown that topic discovery of web-pages could be improved by exploiting *social annotations*, that is, annotations of web-pages provided by human users in social tagging sites such as Delicious [28].

In microblogging sites like Twitter, inferring the credentials (attributes) of individual users is necessary to determine how much trust to place in the content generated by them. Most prior works attempted to discover Twitter users' attributes from the contents of the *tweets* posted by the users themselves. For instance, Ramage *et al.* [18] used Latent Dirichlet Allocation to map the contents of a tweet stream into topical dimensions. Kim *et al.* [12] used chi-square distribution measure on the tweets posted by users included in a common List to identify topics of interest to the users. However, prior research has shown that tweets often contain conversation on day-to-day activities of users [10], making it difficult to identify meaningful topics from tweets alone. Hence, several studies have attempted to enhance the topics identified from tweet streams by querying Wikipedia [15, 17] or search engines [5] using words identified from tweets.

Further, there have been several attempts to measure the influence of Twitter users and hence to identify influential users or experts [6, 4, 13, 19]. However, none of the above mentioned efforts attempts to identify experts in any *specific topic*. To the best of our knowledge, there has been only two efforts that have approached the problem of identifying experts in *specific topics* [26, 16]. Weng *et. al.* [26] proposed a Page-Rank like algorithm Twitter-Rank, that uses both the Twitter graph and processed information from tweets to identify experts in particular topics. On the other hand, Pal *et. al.* [16] used clustering and ranking on more than 15 features extracted from the Twitter graph and the tweets posted by users.

Apart from the above research studies, there also exist some *services* for identifying topical experts in Twitter. Recognizing the importance of searching for experts on specific topics, Twitter itself provides an official "who to follow" (WTF) service [22] where one can search for experts on a given topic (query). Though the exact details of implementation of the service are not publicly known, it is reported that Twitter WTF uses several factors such as the profile information (e.g. name and bio) of users, their social links, their level of engagement in Twitter, and so on [23] to identify topical experts.

It can be noted that all the above approaches primarily rely on the information provided by a user herself (e.g. her account name and bio, the tweets posted by her) and her social graph, to infer the topics in which she is an expert. In contrast, the present work uses an entirely different methodology to infer the topics of expertise of an individual Twitter user, which relies on the 'wisdom of the Twitter crowd' (i.e. how others describe this user), collected through crowd-sourced Lists. Further, all of the above mentioned research studies use fixed Twitter datasets collected at a certain point in time. To the best of our knowledge, this study is the first to address the challenge of keeping an OSN-based search / recommender system up-to-date, a challenge that has become essential given the phenomenal rate of increase of population in today's OSNs [3].

Finally, it is important to mention that a few prior studies have used Twitter Lists for different purposes, such as identifying seed nodes for sampling algorithms or topic-sensitive Pagerank-like algorithms [27, 25] or for contextualizing a user [17].

Chapter 3

Inferring Expertise

In this chapter, we propose our methodology for finding topic experts using a recently introduced Twitter feature called Lists. We will describe our method of inferring the expertise of individual Twitter users and then evaluate the accuracy and expressiveness of the inferred expertise.

3.1 Methodology

Our methodology is based on the Twitter Lists feature. In late 2009, Twitter introduced Lists to help users organize their followings (i.e. the people whom a user follows) and the information they post [11]. By creating a List, a user can group other Twitter users, and view the aggregated tweets posted by all the listed users in the List timeline. When creating a List, a user typically provides a List name (free text, limited to 25 characters) and optionally add a List description. For instance, a user can create a List namely "celebrities" and add celebrities to this List. Then, the user can view tweets posted by these celebrities in the List timeline.

Name	Description	Members
News	News media accounts	nytimes, BBCNews, WSJ, cnnbrk
Music	Musicians	Eminem, ladygaga, rihanna, BonJovi
Tennis	Tennis players & news	andyroddick, usopen, ATPWorldTour
Politics	Politicians & experts	BarackObama, whitehouse, billmaher

Table 3.1: Examples of Lists, their description, and some members

Table 3.1 presents illustrative examples of Lists, extracted from Twitter users. The key observation here is that the List names and descriptions provide valuable semantic cues to the topic of expertise of the members of the Lists. For example, using List meta-data, we can associate BarackObama with Politics and Politicians and Eminem with music and musicians. Thus, Lists provide a way to annotate Twitter users with their topics of expertise. Interestingly, these annotations are generated by arbitrary Twitter users and so they reflect the collective wisdom of the crowds.

Our strategy consists of extracting frequently occurring topics (words) from the List meta-data (names and description) and associating these topics with the listed users. The intuition behind our strategy is that a user listed by many other users under a certain topic is very likely to be an expert on that topic. Previous efforts that analyzed Twitter Lists showed that *nouns* and *adjectives* in list names and descriptions are particularly useful for this purpose [17]. So our strategy to extract topics from List meta-data consists of the following steps:

1. We first apply common language processing techniques, such as case-folding, stemming, and removal of stop words. In addition to the common stop words, a set of domain-specific words are also filtered out, such as Twitter, list, and formulist (a tool frequently used to automatically create Lists).

2. Since list names cannot exceed 25 characters, users often combine multiple words using *CamelCase* (e.g. TennisPlayers). Thus, we separate these words into individual words.

3. We identify nouns and adjectives using a part-of-speech tagger.

4. As a number of list names and descriptions are in languages other than English, we group together words that are very similar to each other (based on edit-distance among words), e.g. politics and *politica*, journalist and *jornalistas*, etc.

5. As list names and descriptions are typically short, we consider only unigrams and bigrams as topics.

The above strategy produces a set of topics for each user, as well as the frequency with which a topic appeared in the names and descriptions of the Lists containing the user.

3.2 Quality of Inferred Expertise

When evaluating the quality of inferred expertise, we check for two metrics: (i) accuracy: is the user really an expert in the inferred topics of expertise? (ii) expressiveness: do Lists comprehensively capture all the different topics in which a user has expertise?

For our evaluation, we need to gather ground truth information about Twitter users' expertise. Since such ground truth is difficult to obtain for a random set of Twitter users, we consider the following strategies: First, we evaluate for a select set of *popular* users whose true topics of expertise are generally well-known or easily verifiable. Second, for a given set of topics, we collect the top experts identified by the state-of-the-art research system for identifying topical authorities [16], and by the official Twitter WTF service [22]. We then check if our methodology identifies these users as experts in the given topics. The results not only demonstrate the high quality of our expertise inference, but they also uncover drawbacks of competing state-of-the-art methods.

Inferred expertise for selected popular users

Table 3.2 shows the top 10 topics (obtained using our List-based method) for Twitter users whose expertise is well-known. It is evident that the main topics accurately describe the topics of expertise of the users. The inference is accurate and comprehensive not only for users with millions of followers, but also for users with hundreds or thousands of followers.

For instance, for Mark Sanderson (a well known professor in the field of information retrieval), even though his Twitter account is included in only *12 Lists*, the inferred topics identify that he is a researcher in computer science ("cs"), specializing in information retrieval, machine learning ("ml"), search and so on. Again, for US senators (two examples shown in Table 3.2 – Chuck Grassley and Claire McCaskill), this methodology could accurately identify a variety of topics, for instance, their political party (Republicans /

User	# Followers	Most frequent topics
Barack Obama	12,481,245	politics, celebs, government, famous, president, news, leaders, noticias, current events
Ashton Kutcher	9,479,352	celebs, actors, famous, movies, stars, comedy, funny, music, hollywood, pop culture
Mark Sanderson	320	information retrieval, ir, cs, ml, semantic, analysis, search, research, nlproc, tech
Chuck Grassley	34,710	politics, senator, congress, government, iowa, republicans, officials, conservative
Claire McCaskill	63,687	politics, senator, government, congress, democrats, missouri, progressive, women
BBC News	574,035	media, news, noticias, journalists, politics, english, newspapers, current, london
Linux Foundation	46,718	linux, tech, open, software, libre, gnu, computer, developer, ubuntu, unix
Yoga Journal	71,689	yoga, health, fitness, wellness, magazines, media, mind, meditation, body, inspiration

Table 3.2: The top tags of expertise of some popular Twitter users, as identified using Lists

Democrats), their state, their gender ('women' in case of Claire McCaskill), their political ideology (conservative / progressive) and even a number of the senate committees of which each senator is a member (e.g. 'health' in case of Chuck Grassley). We verified the accuracy of our inference using the Wikipedia pages for these people, and found them to be almost always accurate. Thus, List meta-data is often sufficiently rich to yield very high quality expertise inference for users over a large range of popularity (number of followers).

3.3 Comparison with Existing Services

In this section, for a given set of topics, we collect the top experts identified by the state-of-the-art research system for identifying topical authorities [16], and by the official Twitter WTF service [22]. We then check if our methodology identifies these users as experts in the given topics.

3.3.1 State-of-the-art research

Next we compare the extent to which the experts identified by a state-of-the-art research system built by Pal *et. al.* [16] can be recalled by our methodology. Pal *et. al.* use more than 15 features extracted from the Twitter social graph and the content of the tweets posted by users to identify topical experts. Though an implementation of this system is not publicly available, their paper lists the top 10 experts identified for three specific topics – *iphone*, *oil spill* and *world cup*. We test whether the topics inferred by our methodology for these experts match with the topic reported by Pal *et. al.*

We find that for a majority of the top 10 experts in each of the three topics, the set of topics inferred by us includes the topic for which they are reported by Pal *et. al.* – for 8 out of 10 for "iphone", for 7 out of 10 for "world cup", and for 6 out of 10 for "oil spill". Table 3.3 shows some of these experts, along with their bio.

User	**Extracts from Bio**
Query: iphone	
macworld	Mac, iPod, iPhone experts
TUAW	Unofficial Apple Weblog
Query: oil spill	
kate_sheppard	Reporter covering energy, environment
LATenvironment	Environmental news from California
Query: world cup	
FIFAWorldCupTM	FIFA soccer world cup tweets
itvfootball	News from ITV football

Table 3.3: Some of the top results reported by Pal *et. al.* [16], for whom the topics inferred using Lists include the query-topic.

However, for the rest of the cases, the topics inferred using Lists do *not* contain the topic reported by Pal *et. al.*. Table 3.4 lists these users along with their bios. Examining their bio, it is evident that these users are, in fact, *not* specifically related to the topic of the corresponding query. For example, a social media entrepreneur and technology blogger *teedubya* was identified as an expert on "iPhone", even though he is not a specialist on Apple products. Similarly, *Reuters*, *CBSNews* and *channel4news* are general news media and authoritative sources of information on a variety of topics, but they are not related specifically to the topics 'oil spill' or 'world cup'. It is likely that the algorithm used by Pal *et. al.* identified these users as experts because a number of their tweets were related to the topic in question during the period when the evaluation was done.

It is worth noting that Pal *et. al.* explicitly set out to discover experts that are not just overtly general and highly followed authorities like popular news media accounts. They highlight the discovery of dedicated specialists that mostly post tweets related to their specialization. Interestingly, our methodology has successfully recalled all such experts (i.e., 100% recall), even though it is based on a single feature (Lists). In comparison, Pal *et. al.* rely on 15 features, which indicates the relative advantages of using crowd-sourced Lists to identify users' expertise.

User	**Extracts from Bio**
Query: iphone	
teedubya	Social Strategy Shaman, SEO
macTweeter	*Account no longer exists in Twitter*
Query: oil spill	
Reuters	latest news from around the world
CBSNews	official Twitter feed of CBS News
TIME	Breaking news and current events
huffingtonpost	The Internet Newspaper
Query: world cup	
nikegoal	marketing, music, education, sport
Flipbooks	News, Random Information
channel4news	exclusive stories & breaking news

Table 3.4: Top 10 results reported by Pal *et. al.* [16], for whom the topics inferred using Lists does *not* include the query-topic

3.3.2 Twitter's official WTF service

The official Twitter Who-To-Follow (WTF) service helps to search for topical experts for a given topic (query), and is reported to use several factors such as the profile information (e.g. name and bio) of users, their social links, their level of engagement in Twitter, and so on [23] to identify experts. As part of a user survey to evaluate our system [cite here], we obtained the top 20 experts returned by the Twitter WTF service for a few hundred queries generated by users. We investigated the extent to which our methodology would recall these experts.

We find that out of the 3495 users returned by Twitter (top 20 results for some given query), the topics inferred using Lists include the corresponding topic (word in the given query) for 83.4% (2916) of the users. However, the topics inferred by the List-based methodology for the other 16.6% (579) users did *not* contain the topic (word) in the query. To understand these

Query	User	Extracts from Bio/Topics inferred
Users for whom tags inferred contain similar words		
dining	dineLA	official Twitter account of dineLA *Tags: restaurant, food, los angeles, chefs*
hubble	HubbleHugger77	Space Explorer, Director of Saving Hubble *Tags: space, universe, cosmology, nasa*
Wrong results in Twitter WTF top 20 results		
astrophysics	jimmyfallon	astrophysicist *Tags: celebs, comedy, funny, actors, humor*
cooking	danecook	When I tweet, I tweet to kill *Tags: celebs, comedy, funny, famous, actors*
origami	ScreenOrigami	Web developer from Germany *Tags: webdesign, webkrauts, html, designers*

Table 3.5: Examples of (i) users for whom topics inferred contain similar words but not the exact query-word (ii) wrong results within Twitter WTF top 20 results.

missing experts better, we manually verified 50 randomly selected users out of the 579 users.

We found 9 out of these 50 users (i.e. 18%) to be relevant experts on the query topics. Our methodology infers topics very similar to the query, but none matching the exact query-word. Table 3.5 shows two such examples. For the official Twitter account of the 'dineLA' restaurant, the inferred topics include 'food' and 'restaurant' but not the query-word 'dining' (for which it was returned by Twitter WTF). Similarly, for the Twitter user 'Hubble-Hugger77' who is a space explorer and directed the film 'Saving Hubble', we identify 'space', 'cosmology' and 'nasa' but not the query-word 'hubble'. This would appear to suggest that a user's name and bio occasionally contain clues to the user's expertise.

However, in 29 out of the 50 cases (i.e. 58%), we found that the official Twitter WTF service returns *wrong* results, i.e., the returned user is not at all related to the topic of the query for which he is returned. Interestingly, this is most possibly because the query-word appears in the name or bio of the user. For instance, the well-known comedian Jimmy Fallon has (mockingly) described himself as an astrophysicist in his bio, as a result of which he shows up in the top 20 Twitter WTF results for the query 'astrophysicist'. Table 3.5 shows other examples of users who are wrongly included within the top 20 results returned by Twitter WTF. We were not able to infer the relevance of the expert to the query in the remaining 12 out of the 50 (24%) manually verified user accounts, as we found the query to be ambiguous.

Further, we observe that a large number of very well-known users in Twitter either do not have bios, or have bios which do not reveal any topical information about them. This may explain the fact that for a number of queries, the Twitter WTF top 20 results do *not* include very important users even though they are strongly related to the given topic (query). For instance, the top 20 Twitter WTF results for the query 'actor' mostly contain Asian actors whose name or bio contains the word 'actor' but not Ashton Kutcher (the well-known actor from USA, who has close to 10 million followers in Twitter). Table 3.5 also shows some more examples of very important and relevant users who do not show up in the top 20 Twitter WTF results, possibly because their name or bio do not contain any information about their topics of expertise.

Thus, not only does our methodology recall a vast majority (83.4%) of the experts identified by the official Twitter WTF, but also a majority of the missing experts were incorrectly identified by Twitter. Our List-based methodology fails to recall only a small fraction of experts who are actually related to the given query, and even in those cases, we identify topics that are quite similar to the query word.

3.4 Coverage of Experts

In this section, we focus on the coverage of the List-based approach for inferring attributes for Twitter users. Specifically, we investigate how our ability to infer a user's attributes varies with the user's popularity in Twitter. We measure a user's popularity using *follower-rank*, a simple metric that ranks users based on their number of followers.

We ranked the users in our dataset based on their number of followers (as of November 2011) and analyzed how many times users with different follower-ranks are listed. Figure 3.1 shows how the fraction of users who are listed at least $L = 1, 5, 10, 20$ times varies with follower-rank. The follower-ranks on x-axis are log-binned, and the y-axis gives the fraction of users in each bin who are listed at least L times.

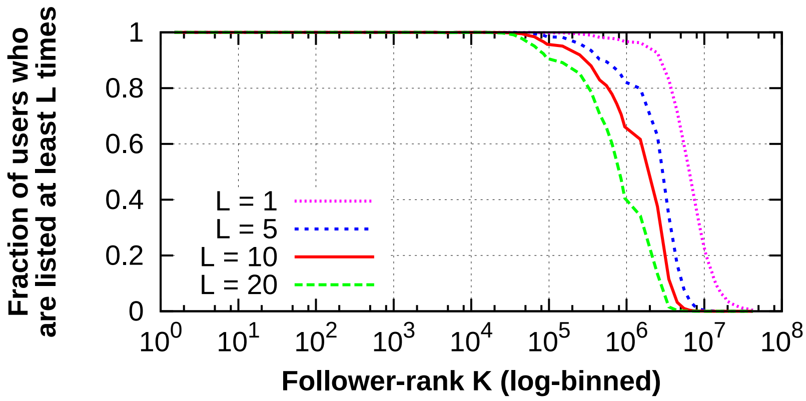

Figure 3.1: Fraction of users who are listed at least L times vs follower-rank

User & Extracts from Bio	Inferred attributes
spacespin: news on robotic space exploration	science, space exploration, nasa, astronomy, planets
laithm: Al-jazeera Network Battle Cameraman	journalists, photographer, al jazeera, media
HumphreysLab: Stem Cell, Regenerative Biology of Kidney	physicians, science, Harvard, stem cell, genetics, cancer, biotech, nephrologist

Table 3.6: Examples of users who are related to niche topics, having intermediary follower-ranks (between 1 million and 10 million)

Users with large numbers of followers: As shown in Figure 3.1, almost all the top follower-ranked users have been listed several times. 98,130 (98%) of the top 100,000 most followed users and 792,229 (79%) of the top 1,000,000 most followed users have been listed 10 or more times. Thus, the List-based methodology can be applied to discover topics related to a large fraction of the popular Twitter users.

Users with moderate numbers of followers: The fraction of listed users falls off dramatically with follower-rank. In fact, only 6% of users with moderate numbers of followers (i.e., users with follower-ranks between 1 and 10 million) are listed 10 or more times. To better understand these users, we manually examined a random sample of 100 users that are listed 10 or more times. Amongst these users, we found users who are experts on very niche topics, such as robotic space exploration, and stem cells. We show some examples of such users in Table 3.6. These users are known only within a small community of people interested in these niche topics, which explains their modest follower-ranks.

Users with few followers: Finally, we found only 1248 users listed more than 10 times amongst users with follower-rank beyond 10 million. Manually inspecting a random sample of these accounts, we found users attempting to abuse the Lists feature. For instance, 67% of these users have only 1 or 2 followers who have listed these users in multiple different Lists. Further, we found 64 users who listed themselves multiple times, which suggests an attempt to manipulate the Lists feature.

3.5 Summary

Our evaluation demonstrates that our proposed methodology of utilizing crowd-sourced List meta-data provides an accurate and comprehensive inference of topics of expertise of individual Twitter users. We also show that in many cases, the List-based methodology is more accurate, as compared to the existing techniques of inferring topics of a user from his profile data or his tweets. Moreover, we found that the List-based methodology to discover user attributes can be successfully applied for a large majority of the popular Twitter users. Only a small fraction of users with moderate popularity are listed multiple times, but they tend to be experts on niche topics.

Chapter 4

Cognos Search System

In this chapter, we leverage our previously discussed methodology to infer users' expertise to build *Cognos* [1], a search system for topical experts in Twitter. Cognos using crowd-sourced Lists as the *only* source of information and so its performance illustrates the potential uses of Lists in finding experts. We first describe how we rank experts in Cognos and then present an extensive evaluation of the Cognos system.

4.1 Ranking Experts

Ranking of users related to a given topic is a well-studied problem, and over the years, several ranking algorithms have been proposed for the Web [9], online topical communities [29], and even for topical experts in Twitter [16, 26]. The expert ranking schemes in Twitter take into account several metrics extracted from the social graph and the content of the tweets posted by users. In contrast, we decided to evaluate a ranking scheme that is based solely on the Lists feature, since one of our objectives is to evaluate crowd-sourced Lists as the *only* source of information for topical experts – we have already shown that Lists can be used to accurately infer topics of expertise, now we investigate whether Lists are also an effective metric to rank topical experts.

Using the method described in the previous section, we obtain for each individual user, a set of topics as well as the frequency of occurrence of

[1] The name is derived from the word *cognoscenti*, i.e. people who are considered to be especially well informed about a particular topic.

each topic in the names and descriptions of the Lists containing the user. Thus, for each user we obtain a vector of topics and we store this in a database. Given a query, we compute a topical similarity score between the topic vector for a user and the given query vector, using the algorithm in [7] which computes the *cover density ranking* between the vectors. We chose this similarity score (which is suited to queries containing one to three terms) since queries to expert search systems are almost always short, hence using cosine similarity on tf-idf based representations may not be very effective [14, 16]. Finally, we multiply the topical similarity score for a user with the logarithm of the number of Lists containing the user – the intuition behind this is that a user who is included in more number of Lists (by other users) is likely to be more popular in Twitter.

Thus, given a query (topic), Cognos identifies the set of experts related to the topic using the List-based methodology discussed in previous chapter, and then ranks them using the algorithm described above. In the remainder of this section, we extensively evaluate this List-based methodology of identifying and ranking topical experts in Twitter.

4.2 Building Experts Collection

The dataset used in this work uses extensive data from a previous measurement study [6] that included a complete snapshot of the Twitter social network and the complete history of tweets posted by all users as of August 2009. More specifically, the dataset contains 54 million who had 1.9 billion follow links among themselves and posted 1.7 billon tweets (as of August 2009). Out of all users, nearly 8% of the accounts were set as private, which implies that only their friends could view their links and tweets. We ignore these users in our analysis. For a detailed description of this dataset we refer the reader to [6].

4.2.1 Crawling Lists

Lists were introduced in Twitter *after* the above Twitter dataset was collected. To populate the Cognos expertise database, we started crawling all the Lists containing all Twitter users. We quickly realized that a brute-force crawl of all Lists for all users would be prohibitively expensive and would not scale. So we crawl the Lists containing only the 54 million Twitter users included in a complete snapshot of the Twitter social network taken in the above dataset. This is only a small fraction of the estimated 465 million Twitter users as of January 2012 [3]. We address the challenge of crawling Lists efficiently and scalably to include experts that joined after 2009, in the next Chapter.

Hence in November 2011, we re-crawled the profiles of all 54 million users in our dataset, which contains information about the number of Lists each user appears in. We found that **6,843,466** users have been listed at least once. In order to reliably infer topics of a user from Lists, it is important that a user has been listed at least a few times. We found that 20% of the listed users (**1,333,126** users) were listed at least 10 times. Since we intended to study users with a minimum level of expertise (i.e. who appear in at least a certain number of Lists), we considered these users and crawled their List information.

Using the Twitter API, we crawled the name and description of the Lists in which the top-listed users appear. Due to rate-limitations in accessing the Twitter API, we collected the information of at most 2000 Lists for any given user. However, as only 0.08% of the listed users are included in more than 2000 Lists, this has a negligible effect on the study. Overall for the 1.3 million top-listed users, we gathered a total of 88,471,234 Lists. Out of these, 30,660,140 (34.6 %) Lists had a description, while the others had only the List name.

4.3 Evaluating Cognos

Judgements on the quality of the results returned by a search system are to an extent subjective. So we chose to evaluate Cognos through an extensive user study where a set of human evaluators judged the relevance of the results returned by Cognos, using a web-based feedback service (available at `http://twitter-app.mpi-sws.org/whom-to-follow/`). We also gathered another set of human evaluations where the results returned by Cognos were directly compared with those returned by the official Twitter WTF service [22]. We also compared the top experts returned by Cognos with

Category	Sample queries
News	politics, sports, entertainment, science, technology, business
Journalists	politics, sports, entertainment, science, technology, business
Politics	conservative news, liberal politicians, USA/Indian politicians
Sports	F1, baseball, soccer, poker, tennis, NFL, NBA, Bundesliga
Entertainment	celebrities, movie reviews, theater, music
Hobbies	hiking, cooking, chefs, traveling, photography
Lifestyle	wine, dining, book club, health, fashion
Science	biology, astronomy, computer science, complex networks
Technology	iPhone, mac, linux, cloud computing
Business	markets, finance, energy

Table 4.1: The sample queries used for evaluation of Cognos.

those returned by the state-of-the-art research system [16].

The above URL was publicly advertised to all people in three academic institutes located across three different continents, inviting a few hundred people at each of the institutes to evaluate the system. It is to be noted that we preferred such an *in-the-wild* evaluation (instead of a controlled evaluation, e.g. with a fixed set of evaluators and few selected queries, as used by [16]) since this actually resembles a real-world deployment of the search system.

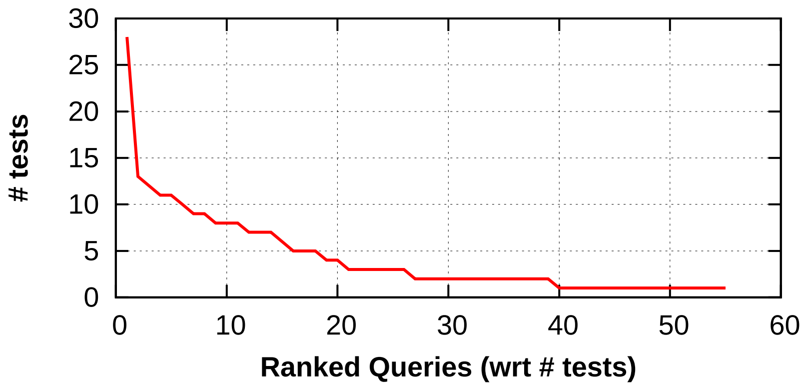

Figure 4.1: Distribution of the number of times a query was asked (out of 55 sample queries)

In this evaluation, an evaluator issues a query, for which she is shown the top 10 results returned by Cognos. Then the evaluator gives a binary judgement on each of the top 10 results as to whether it is relevant to the given query. The queries used for the evaluations could be selected from a given set of 55 sample queries spread over the 10 categories shown in Table 4.1. Fig. 4.1 shows the distribution of the number of times each query was asked, the 5 most frequently asked queries being “computer science”, “cloud computing”, “movie reviews”, “technology news”, and “travelling”. In the rest of this section, we use the term ‘evaluation’ to indicate a relevant or non-relevant judgement for an individual result given by Cognos for a particular query.

Figure 4.2: Screenshot of evaluation of relevance of results

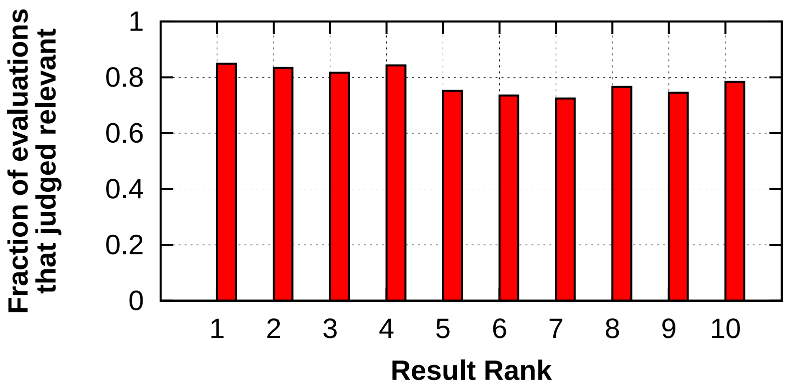

Figure 4.3: Fraction of evaluations that judged a result relevant (across all queries) – shown for each individual rank out of top 10

Overall, we obtained 2136 relevance judgements[2] over the top 10 results for the 55 sample queries, out of which 1680 (78.7%) judged the result (topical expert shown by Cognos) to be relevant to the query. Fig. 4.3

[2]Despite our request, some of the evaluators did not evaluate all 10 results for a particular query.

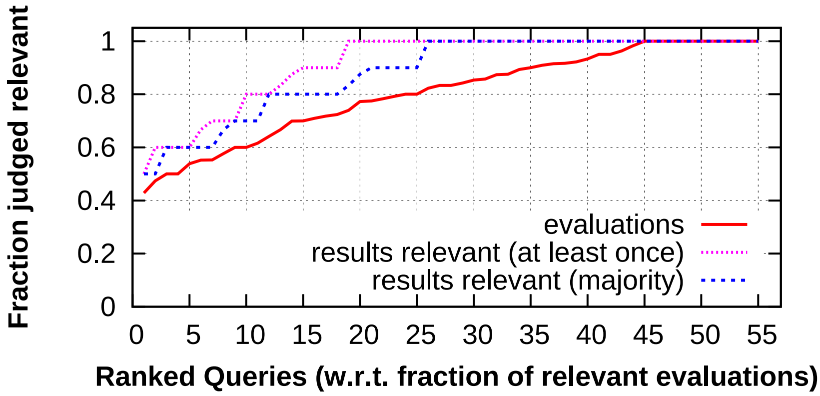

Figure 4.4: Fraction of evaluations / individual results that were judged relevant

shows the fraction of evaluations that judged a result as relevant, for each individual rank out of the top 10 (i.e. considering the results shown at a certain rank for any of the 55 queries) to be largely invariant – the top 4 results were judged to be relevant in more than 80% of the evaluations, while the results ranked 5–10 were judged relevant in more than 75% of the evaluations.

Next we examined the Cognos results that received the 456 (21.3%) 'non-relevant' judgements. We found that a large amount of subjectivity in these judgements driven by whether a particular user recognizes another user as a top expert on a given topic. We found a number of cases where the same result for the same query was judged relevant by some evaluator and non-relevant by others. For example, for the query 'cloud computing', Werner Vogels, who is one of the principal architects of Amazon's approach to cloud computing, was rated as relevant in 4 evaluations, and as non-relevant in 6 evaluations, possibly because the name was unknown to these evaluators.

To understand the subjectivity in our judgements, we consider for each particular query, (i) what fraction of evaluations judged a result for this query as relevant, (ii) what fraction of the top 10 results were judged relevant at least once, and (iii) what fraction of the top 10 results were judged relevant in the *majority* of evaluations. Fig. 4.4 shows the distribution of these fractions for all queries (where queries are ranked by the fraction of evaluations that judged a result as relevant). It can be seen that for 37 out of the 55 queries, every result was judged relevant by at least one evalua-

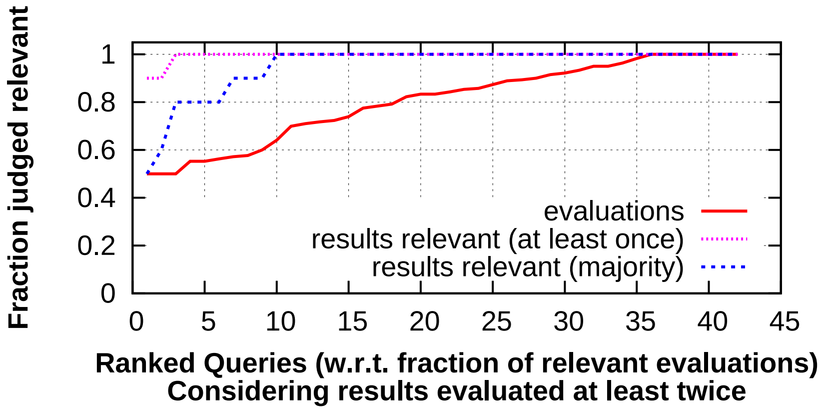

Figure 4.5: Fraction of evaluations / individual results that were judged relevant – considering only those results for a query, which were evaluated at least twice

tion, and for 30 out of the 55 queries, every result was judged relevant by the majority of the evaluations for that particular result.

The effects of subjectivity is seen even more clearly in Fig. 4.5 where we plot the above three fractions for each query, considering only those results that were evaluated at least twice. Note that there are 13 queries (out of the 55) for which no individual result was evaluated twice, hence Fig. 4.5 shows the other 42 queries. For as many as 40 out of these 42 queries, every result (that was evaluated at least twice) was judged relevant by at least one evaluation, and for 33 out of these 42 queries, every result (that was evaluated at least twice) was judged relevant by the majority of the evaluations for that result.

The above statistics show that a vast majority of the results returned by Cognos were judged topically relevant to the given query (topic) by at least some evaluators. Thus, Cognos can successfully identify relevant experts over a wide variety of topics.

4.4 Comparison with Existing Services

In this section, for a given set of topics, we collect the top experts identified by the state-of-the-art research system for identifying topical authorities [16], and by the official Twitter WTF service [22]. We then check if where our methodology ranks these users as experts in the given topics.

4.4.1 State-of-the-art research

As discussed in previous Chapter, Pal *et. al.* [16] list the top 10 experts identified by their algorithm for three specific queries: *oil spill*, *iphone*, and *world cup*. For these queries, Table 4.2 compares the top 5 results from Cognos and the top 5 results reported by Pal *et. al.*, along with the bio and number of followers of each user.

Note that while the top results reported by Pal *et. al.* contain some general news media sites, the top Cognos results are much more topic-specific, even if they are not as popularly followed as the news media sites. Interestingly, in their paper, Pal *et.al.* explicitly set out to discover such specialized topic-specific experts, even if they are highly visible. Cognos achieves this goal better than the state-of-the-art system.

Pal *et. al.* also mentions that for the query "toy story" (a popular movie), their algorithm returns *leeunkrich* (the director of the movie) as the top result; Cognos also returns *leeunkrich* as the top user for this query, followed by popular fan sites on the movie.

Given that Cognos uses only a single feature as compared to more than 15 network and content-based features used by Pal *et. al.* [16], these results further demonstrate the potential of crowd-sourced Lists in identifying topical experts in Twitter.

4.4.2 Twitter's official WTF service

In this evaluation, when an evaluator issues a query, she is simultaneously shown the top 10 results returned by Cognos as well as the top 10 results returned by the official Twitter WTF service for the same query. The results are anonymized, i.e. the evaluator is not told which result-set is from which service, in order to prevent bias in judgement. Then the evaluator indicates which set of results is better for the given query, or whether both result-sets are equally good or equally bad [3]. It is to be noted that since Cognos uses a Twitter dataset crawled in 2009, for this comparison to be fair, we filtered out from the Twitter WTF results those user-accounts which were created

[3] The search engines corresponding to the result-sets are revealed to the evaluators *after* the evaluation is done.

Cognos results		Results by Pal *et. al.*	
User/Bio/# Followers		**User/Bio/# Followers**	
Query: oil spill			
BP_America *BP America*	35,505	NWF *National Wildlife Federation*	76,796
TheOilDrum *energy, peak oil*	26,257	TIME *Breaking, news, events*	3,231,359
GOHSEP *Emergency Preparedness*	5,295	huffingtonpost *The Internet Newspaper*	1,574,848
usoceangov *National Ocean Service*	37,866	NOLAnews *Latest news and updates*	29,433
USCG *US Coast Guard*	20,513	Reuters *Latest news*	1,491,852
Query: iphone			
p0sixninja *iPhone Hacker*	127,631	macworld *Mac, iPod, iPhone experts*	182,248
iH8sn0w *made f0recast, iREB, iFaith*	105,015	Gizmodo *Technologies that change*	347,667
chronicdevteam *Hax*	107,541	macrumorslive *Updates from Apple events*	170,813
MuscleNerd *iPhone hacker*	330,625	macTweeter *Account Suspended*	–
iPhone_News *iPhone news and notes*	153,024	engadget *Twitter account of Engadget*	419,583
Query: world cup			
worldcupscores *Live 2010 World Cup Scores*	10,866	TheWorldGame *Australia's football website*	11,541
EdsonBuddle *Soccer playerFC Ingolstadt*	30,808	GrantWahl *Sports Illustrated writer*	180,290
thefadotcom *Website for England Football*	102,536	owen_g *Guardian's Olympics editor*	14,930
nytimesgoal *New York Times Soccer Blog*	11,699	guardian_sport *Sport news from Guardian*	121,095
herculezg *US National Team Forward*	31,454	itvfootball *News from ITV football*	54,395

Table 4.2: Top 5 results by Cognos and by Pal *et. al.* [16] for the three queries evaluated by Pal *et. al.*, along with their bio and number of followers

after 2009 [4]. In order to test the performance of Cognos 'in-the-wild', we allowed the evaluators to issue any query of their choice.

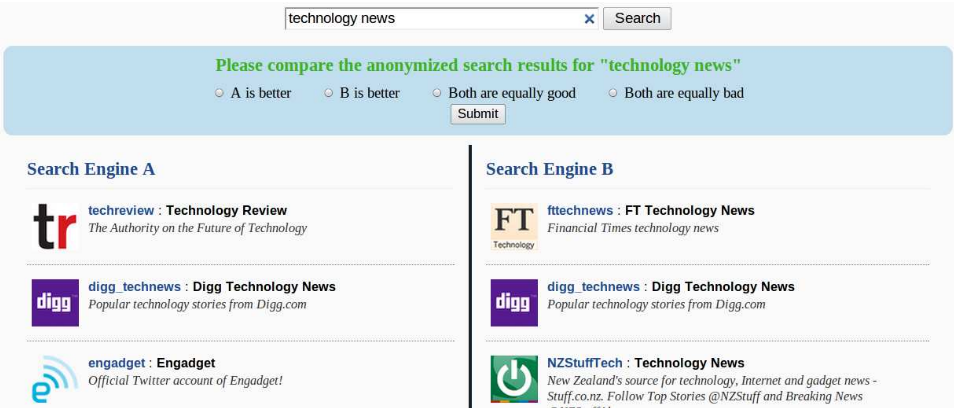

Figure 4.6: Screenshot of comparison of Cognos vs Twitter results

We obtained relevance judgements for 325 total queries of which 259 are distinct. These queries are evaluator-chosen and they cover a wide variety of topics. Given the high subjectivity observed in user relevance judgements in the previous section, we choose to focus our evaluation on the 27 distinct queries that were asked at least two times. In total, these 27 queries were asked 93 times.

Verdict	Queries
Cognos Better	Linux, computer science, mac, India, Apple, Facebook, internet, ipad, markets, windows phone, photography, politic journalist
Twitter WTF Better	politic news, music, Sachin Tendulkar, Twitter, Alka Yagnik, Anjelina Jolie, cloud computing, Delhi, Harry Potter, metallica
Tie	Microsoft, Dell, Kolkata, Sanskrit as an official language

Table 4.3: Evaluator-chosen queries for comparison of Cognos and Twitter WTF, where the verdict is given by majority voting.

Table 4.3 shows the 27 queries that were asked at least twice. For each query, we consider the verdict – Cognos better / Twitter WTF better / tie – based on majority voting. The queries for which there was a unanimous verdict (i.e. all evaluations for this query agreed that one was better) are italicized in Table 4.3. Cognos was judged to be better for 12 out of the

[4]The date on which an account was created is available from the profile information.

27 queries, while Twitter WTF was judged better for 11, and there was a tie for 4 queries. The fact that Cognos was judged to be better than the official Twitter WTF service for 44% of the queries, clearly indicates the potential of crowd-sourced Lists (the only feature used in Cognos) in identifying topical experts in Twitter. It can be noted that a significant fraction of the cases where Twitter was unanimously judged better are names of individuals (celebrities) or organizations. Since such names appear very rarely in the List names / descriptions, Cognos does not handle these queries well.

Cognos results	**Twitter WTF results**
User/Bio	**User/Bio**
Query: music	
Katy Perry *i kissed a girl...*	iTunes Music *Music updates for U.S.*
Lady Gaga *mother monster*	YouTube *YouTube news, trends, videos*
taylorswift13 *Bio not written*	SonyMusicGlobal *home of Sony Music*
jtimberlake *Official Justin Timberlake*	50cent *It's the kid 50 Cent*
Pink *It's all happening*	guardianmusic *Squashing music*
Query: windows phone	
BrandonWatson *Developers on Windows Phone*	Windows Phone *Official Windows Phone*
wmpoweruser *Windows Phone Power Users*	pocketnow.com *Windows Phone news*
Charlie Kindel *Founder, CTO, Mentor*	WP Dev Team *Windows Phone Dev Team*
joebelfiore *Runs team doing W. Phone 7*	WindowsPhoneNL *Windows Phone in Nederland*
pocketnow.com *Windows Phone news*	WPCentral *All thing Windows Phone 7*

Table 4.4: Top 5 results by Cognos and by Twitter WTF for the queries "music" and "windows phone".

It can also be noted from Table 4.3 that the top 10 Cognos results show very low overlap with top 10 Twitter WTF results across all queries. This is in spite of the fact that 83.4% out of the Twitter WTF top 20 results for some query (topic), were inferred by our List-based methodology to be related to the same topic. This implies that the low overlap between the top Cognos results and Twitter WTF results is primarily due to the List-based *ranking* used in Cognos. We observe that in general, the top Twitter WTF results mostly include organizations / business accounts while the Cognos top results mostly include personal accounts. We present some examples in Table 4.4 for the queries “music” (for which the majority voted Twitter WTF better), and “windows phone” (for which the majority voted Cognos better). This is possibly because the Twitter WTF considers the name and bio of users [23], and organizational / business accounts are more likely (compared to personal accounts) to have names or bios which contain terms related to their topics of expertise. As such, these examples again bring out the subjective nature of human judgement, where some evaluators preferred the personal accounts while others preferred the organizational accounts.

4.5 Summary

Our evaluation of the Cognos search system shows that a vast majority of its results are relevant for a wide variety of topics. In fact, Cognos rarely produces irrelevant results for user queries. Comparing Cognos with state-of-the-art research system by Pal *et. al.* and official Twitter WTF service highlights the advantages of relying on crowd-sourced Lists to identify experts. Cognos yields particularly better search results in the cases when the bio or tweets posted by a user does not correspond to or contain information about the user's topic of expertise. In fact, Cognos performs as good as or better than the official Twitter WTF service for more than 52% of the queries, even though it is based on a single and simple feature (Lists).

Chapter 5

Finding New Experts

In the section, we address the practical challenge of keeping our Cognos system up-to-date, even as hundreds of thousands new Twitter accounts and new Lists are created every day.

5.1 Scalability problem with crawling Lists

We begin by analyzing the scalability of a simple updation strategy that relies on periodically crawling all the Twitter users and the Lists that contain them. Recent reports indicate that 200 million new users joined Twitter in the last 9 months [3], which roughly amounts to 740,000 new users joining per day. Twitter rate-limits the number of profile crawls from a single machine (IP address) to 150 API requests per hour [21], i.e., to 3600 user profile crawls per day. For each user, we would need to make at least one extra request to crawl her Lists. In fact, Twitter returns only 20 Lists per request. For instance, for a user with more than 2000 lists, it would be necessary to make 100 requests to Twitter API. Thus, just to keep the system up-to-date, a lower-bound rate limit would be of at least 1,480,000 requests per day. Fortunately, three of our machines were *white-listed* by Twitter, which allows each of them to crawl at a significantly higher rate of 20,000 user profiles per hour. Thus, we can fetch at most 1,440,000 (20,000 $\times$ 3 $\times$ 24) user profiles per day from all three of our white-listed machines. Note that our maximum crawl rate is still lower than the lower-bound rate we would need to gather the Lists of all new users joining Twitter. Given that we would need to periodically crawl the new Lists for the already existing 465 million Twitter users [3], it becomes quite evident that our simple strategy of crawling all users' Lists would not scale.

Next, we estimated the number of highly listed users amongst the 465 million Twitter accounts as of January 2012. Since Twitter assigns user-ids in an integer sequence starting from 1, we took a random sample of 300,000 integers in the range 1 to 465 million, and attempted to crawl the profiles of Twitter user-ids in the sample. The distribution of experts within this large random sample can be expected to be similar to the distribution of experts among all Twitter users. For instance, only 363 out of the 300,000 sampled users (i.e. 0.12%) were Listed 100 or more times; hence we expect the total number of Twitter users who are Listed 100 or more times to be 0.12% of the entire Twitter population. Thus, only a small fraction of all Twitter users are highly listed experts and once they are identified, it would be possible to crawl the Lists containing these experts periodically. The key challenge, however, lies in efficiently identifying these experts from the large Twitter user population.

5.2 Crawling experts efficiently

Our discussion above showed that we cannot exhaustively crawl Lists for all Twitter users. However, we can crawl Lists for the small fraction of *expert* users, if we somehow identified them from the Twitter user population. We now propose and evaluate a strategy to efficiently identify expert users.

We tried several approaches for discovering experts who did not appear in our dataset. For instance, one approach was to identify those Lists which include several of the *known* experts (i.e. highly-listed users who already appeared in our dataset), and then using those Lists to find new experts. However, we describe in detail that approach which turned out to be the most successful in discovering new experts efficiently.

We begin by observing that the Twitter social network consists of a number of *hubs*, users who follow a large number of popular experts and include them in Lists. Our strategy is to first identify popular hubs in an older snapshot of the network (when the network was considerably smaller) and then leverage the Lists created by the top hubs in order to find new authorities. It can be noted that this strategy also relies on crowd-sourcing – we expect the Twitter crowd (in particular, the top hubs) to discover experts who newly join Twitter, and we can utilize their discovery by periodically crawling the Lists created by the top hubs.

We used the well-known HITS algorithm to identify the top hubs in the snapshot of the Twitter network gathered in 2009 [6], when the network had only 54 million users. We then crawled the Lists created by the top *1 million* hubs in the network to efficiently discover experts. In all, the top 1 million hubs had created 479,129 Lists, which taken together contained 4,100,367 unique users. Out of these, 2,064,373 (i.e. 50.3%) have been included in 10 or more Lists. In comparison, only 1.13% of all the users in our large random sample of Twitter users are listed 10 or more times. The difference clearly indicates that our strategy is effective in focusing our crawls on experts in Twitter. Also, the crawl for the top 1 million hubs took about 3 weeks (January 20 – February 8, 2012) using the machines white-listed by Twitter, and hence can be repeated every month to discover new experts.

5.3 Evaluating coverage of our crawls

In this section, we estimate the fraction of experts covered by our strategy to crawl Lists created by top hubs.

5.3.1 Coverage of most listed users

We measure the fraction of most Listed users in Twitter, that is covered by our methodology as follows. First, we estimate the number of Twitter users listed at least K times by computing the number of such users in our 300,000 random sample of users, and then scaling it to the total Twitter user population of 465 million users. Next we calculate the fraction of the estimated number of users Listed at least K times, that are actually discovered by crawling the Lists created by the top hubs.

Figure 5.1 plots the fraction of experts discovered, against the number of top hubs crawled. We find that by crawling the Lists created by the top 1 million hubs, we discovered 25,887 experts who are Listed 1000 or more times, which is 70.6% of our estimated total number of experts Listed

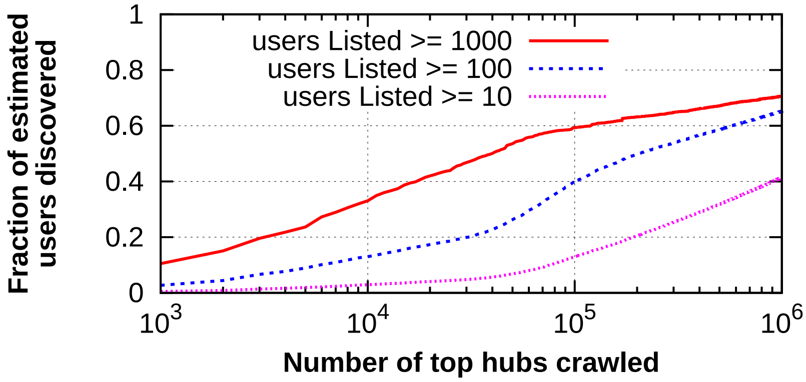

Figure 5.1: Fraction of estimated number of experts who are included in at least K Lists, that is discovered in the hub-based crawl, for K = 10, 100, 1000.

at least 1000 times in Twitter. Further, we find that crawling the Lists created by only the top 100,000 hubs is sufficient to discover 53.3% of the estimated number of experts Listed 1000 or more times in Twitter. Thus, the hub-based updation methodology can be used to efficiently discover a large fraction of new experts in Twitter.

5.3.2 Coverage of newly joined experts

Here we evaluate how our discovery of experts varies with the age of their Twitter accounts. Fig. 5.2 plots the distribution of the age of accounts (in days) of experts who are listed 100 or more times, in the random sample described above, and in the set of new users discovered by our hub-based crawl. It is evident that the crawl discovers old users as well as very new users too.

Account	Bio / Description	Listed	Created
MartiRiverola	F.C.Barcelona	67	Feb 6
annekirkbride	English Actress	23	Feb 4
AaronAStanford	Canadian Actor	32	Feb 1
Shay Given	Ireland goalkeeper	107	Jan 27
CourteneyCox	American actress	294	Jan 24
PMOIndia	Prime Minister India	309	Jan 23

Table 5.1: Examples of very recently created expert accounts discovered by our Hub-based crawl (which ended on Feb 8, 2012)

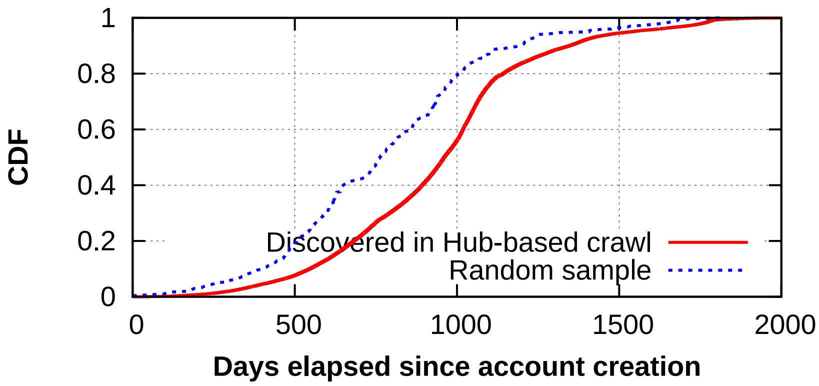

Figure 5.2: Distribution of the number of days elapsed since creation of accounts in random sample and those discovered in hub-based crawl (considering users listed at least 100 times)

Our expert discovery strategy is effective in discovering newly joined experts. For example, even though our top hubs were selected using a 2009 snapshot of the Twitter network, more than 42.3% of the 4,100,367 users in the Lists created by these hubs have joined Twitter after 2009. Further, we show some examples of very recently created Twitter accounts that our hub-based crawl could discover, in Table 5.1. Our crawl of Lists created by the top 1 million hubs, which ended on February 8, 2012, discovered some experts who joined Twitter as recently as Feb 6 or Feb 4 (i.e. while the crawl was going on). This validates our hypothesis that the top hubs quickly discover newly joined experts and add them to Lists, and hence shows the effectiveness of the hub-based updation strategy.

5.3.3 Coverage of experts identified by other systems

We evaluate whether our updation methodology can discover topical experts returned by the Pal *et. al.* research system and Twitter WTF service. Out of the 30 topical experts stated by Pal *et. al.* (for the three topics "oil spill", "world cup" and "iPhone"), 29 are included in the crawls of Lists created by the top 1 million hubs (the remaining account no longer exists in Twitter). Next, we consider the top 20 experts returned by Twitter WTF service for all the 259 queries obtained by our user-survey and calculate what fraction of these experts are covered by our hub-based crawls. Figure 5.3 plots the distribution of the fraction of Twitter WTF top 20 results included in our

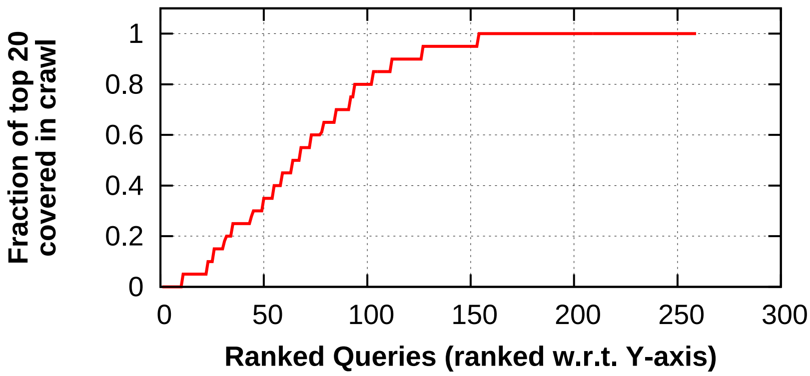

Figure 5.3: Distribution of the fraction of the Twitter WTF top 20 results that is covered in our hub-based crawl

hub-based crawls, across all queries. It is seen that our crawls include all Twitter WTF top 20 results for more than 50% of the queries and at least 15 out of the Twitter WTF top 20 results for close to 80% of the queries.

The above results indicate that the hub-based strategy – periodically discovering experts through the Lists created by top hubs – can be used to efficiently discover newly joined experts (even very recently joined ones), and thus keep an expert search system up-to-date in the face of rapid increase in the Twitter population.

Chapter 6

Conclusions

As Twitter emerges as a popular platform for users to search for interesting topical content, an important research challenge lies in identifying experts in specific topics. In this thesis, we show that an effective solution to this hard problem lies in exploiting wisdom of the Twitter crowds via a recent feature introduced by Twitter called *lists*.

6.1 Brief Summary

First we propose a methodology for finding topic experts using a recently introduced Twitter feature called Lists. We observe that individual Twitter users, for their own convenience, annotate and classify experts in various topics using the Lists feature. We show that by aggregating the *List* information for a Twitter user, we can discover an extremely rich and varied characterization of the topical expertise of the user as perceived by the Twitter crowds. We describe a novel method for inferring attributes that characterize individual Twitter users. As opposed to existing methods which attempt to infer topics for a user from the contents of the users tweets or profile, we infer topics by leveraging the wisdom of the Twitter crowds, as reflected in the meta-data (names and descriptions) of Lists created by the crowds. We used the proposed topic inference methodology to construct a who-is-who service for Twitter and showed that our service can automatically infer an accurate and comprehensive set of attributes for over a million Twitter users, including most of the popular users.

Our evaluation demonstrates that our proposed methodology of utilizing crowd-sourced List meta-data provides an accurate and comprehensive inference of topics of expertise of individual Twitter users. We also show that in many cases, the List-based methodology is more accurate, as compared to the existing techniques of inferring topics of a user from his profile data or his tweets. Moreover, we found that the List-based methodology to discover user attributes can be successfully applied for a large majority of the popular Twitter users. Only a small fraction of users with moderate popularity are listed multiple times, but they tend to be experts on niche topics.

Based upon this methodology, we build and deploy Cognos, a topical expert search system. Our evaluation of the Cognos search system shows that a vast majority of its results are relevant for a wide variety of topics. In fact, Cognos rarely produces irrelevant results for user queries. Comparing Cognos with state-of-the-art research system by Pal *et. al.* and official Twitter WTF service highlights the advantages of relying on crowd-sourced Lists to identify experts. Cognos yields particularly better search results in the cases when the bio or tweets posted by a user does not correspond to or contain information about the user's topic of expertise. In fact, Cognos performs as good as or better than the official Twitter WTF service for more than 52% of the queries, even though it is based on a single and simple feature (Lists).

Finally, we come up with an efficient solution to address the practical challenge of keeping our Cognos system up-to-date, even as hundreds of thousands new Twitter accounts and new Lists are created every day. We show that a *hub-based strategy* – periodically discovering experts through the Lists created by top hubs – can be used to efficiently discover newly joined experts (even very recently joined ones), and thus keep an expert search system up-to-date in the face of rapid increase in the Twitter population.

6.2 Future Work

A crucial future challenge lies in making our expertise inference methodology robust against attackers, who create fake Lists including a target user to manipulate the inferred expertise of the user. While we did not find much evidence of such attacks to date, such attacks could be launched in a straightforward manner. Defending against such attacks would require the inference methodology to consider the reputation (or influence) of the users creating Lists, a subject for future research.

The main contributions of the present study a methodology and a service to accurately infer topics related to Twitter users have a number of potential applications in building search and recommendation services on Twitter. For instance, the inferred user attributes can be utilized to search for topical experts in Twitter, who can provide interesting news on a given topic. We plan to explore these possibilities in future.

We demonstrate that even though Cognos is built utilizing *only* the Lists feature, it can compete with the commercial who-to-follow system deployed by Twitter itself. We believe that crowd-sourced Lists provide a valuable foundation for building future content search / recommendation / discovery services in Twitter.

Bibliography

[1] There Are Now 155m Tweets Posted Per Day, Triple the Number a Year Ago. `http://rww.to/gv4VqA`, April 2011.

[2] Twitter - Wikipedia. `http://en.wikipedia.org/wiki/Twitter`, April 2012.

[3] Twitter to hit 500 million accounts by February. `http://bit.ly/twitpopulation`, Jan 2012.

[4] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. Everyone's an influencer: quantifying influence on Twitter. In *ACM WSDM*, pages 65–74, 2011.

[5] Michael S. Bernstein, Bongwon Suh, Lichan Hong, Jilin Chen, Sanjay Kairam, and Ed H. Chi. Eddi: interactive topic-based browsing of social status streams. In *ACM symposium on User interface software and technology (UIST)*, 2010.

[6] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *AAAI ICWSM*, May 2010.

[7] Charles Clarke, Gordon Cormack, and Elizabeth Tudhope. Relevance ranking for one to three term queries. *Information Processing and Management*, 36:291–311, 2000.

[8] Stephen Dill et al. Semtag and seeker: bootstrapping the semantic web via automated semantic annotation. In *ACM WWW*, 2003.

[9] Taher H. Haveliwala. Topic-sensitive pagerank. In *ACM WWW*, pages 517–526, 2002.

[10] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: understanding microblogging usage and communities. In *WebKDD/SNA-KDD*, 2007.

[11] Nick Kallen. Twitter blog: Soon to Launch: Lists. `http://blog.twitter.com/2009/09/soon-to-launch-lists.html`, Sep 2009.

[12] Dongwoo Kim, Yohan Jo, Il-Chul Moon, and Alice Oh. Analysis of Twitter Lists as a Potential Source for Discovering Latent Characteristics of Users. In *Workshop on Microblogging: What and How Can We Learn From It? (with ACM CHI)*, Apr 2010.

[13] Changhyun Lee, Haewoon Kwak, Hosung Park, and Sue Moon. Finding influentials based on the temporal order of information adoption in Twitter. In *ACM WWW*, 2010.

[14] Donald Metzler, Susan Dumais, and Christopher Meek. Similarity measures for short segments of text. In *ECIR*, 2007.

[15] Matthew Michelson and Sofus A. Macskassy. Discovering users' topics of interest on Twitter: a first look. In *Workshop on Analytics for Noisy unstructured text Data (AND)*, 2010.

[16] Aditya Pal and Scott Counts. Identifying topical authorities in microblogs. In *ACM WSDM*, pages 45–54, 2011.

[17] Ravali Pochampally and Vasudeva Varma. User context as a source of topic retrieval in Twitter. In *Workshop on Enriching Information Retrieval (with ACM SIGIR)*, Jul 2011.

[18] Daniel Ramage, Susan Dumais, and Dan Liebling. Characterizing Microblogs with Topic Models. In *AAAI ICWSM*, 2010.

[19] Daniel M. Romero, Wojciech Galuba, Sitaram Asur, and Bernardo A. Huberman. Influence and passivity in social media. In *ACM WWW*, pages 113–114, 2011.

[20] Jaime Teevan, Daniel Ramage, and Merredith Ringel Morris. #twittersearch: a comparison of microblog search and web search. In *ACM WSDM*, pages 35–44, 2011.

[21] Rate Limiting — Twitter Developers. `https://dev.twitter.com/docs/rate-limiting`.

[22] Twitter: Who to Follow. `http://twitter.com/#!/who_to_follow`.

[23] Twitter Improves "Who To Follow" Results & Gains Advanced Search Page. `http://selnd.com/wtfdesc`.

[24] L. Rao, Twitter Seeing 90 Million Tweets Per Day, 25 Percent Contain Links, *TechCrunch*, 2010. `http://tinyurl.com/27x5cay`.

[25] Michael J. Welch, Uri Schonfeld, Dan He, and Junghoo Cho. Topical semantics of Twitter links. In *ACM WSDM*, 2011.

[26] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twitterrank: finding topic-sensitive influential twitterers. In *ACM WSDM*, 2010.

[27] Shaomei Wu, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. Who says what to whom on Twitter. In *ACM WWW*, pages 705–714, 2011.

[28] Xian Wu, Lei Zhang, and Yong Yu. Exploring social annotations for the semantic web. In *ACM WWW*, 2006.

[29] Jun Zhang, Mark S. Ackerman, and Lada Adamic. Expertise networks in online communities: Structure and algorithms. In *ACM WWW*, 2007.