Inferring Who-is-Who in the Twitter Social Network

Naveen Sharma
IIT Kharagpur, MPI-SWS

Saptarshi Ghosh
IIT Kharagpur, MPI-SWS

Fabricio Benevenuto
UFOP

Niloy Ganguly
IIT Kharagpur

Krishna P. Gummadi
MPI-SWS

ABSTRACT

In this paper, we design and evaluate a novel *who-is-who* service for inferring attributes that characterize individual Twitter users. Our methodology exploits the `Lists` feature, which allows a user to group other users who tend to tweet on a topic that is of interest to her, and follow their collective tweets. Our key insight is that the List meta-data (names and descriptions) provides valuable semantic cues about who the users included in the Lists are, including their topics of expertise and how they are perceived by the public. Thus, we can infer a user's expertise by analyzing the meta-data of crowdsourced Lists that contain the user. We show that our methodology can accurately and comprehensively infer attributes of millions of Twitter users, including a vast majority of Twitter's influential users (based on ranking metrics like number of followers). Our work provides a foundation for building better search and recommendation services on Twitter.

Categories and Subject Descriptors

H.3.5 [**On-line Information Services**]: Web-based services; J.4 [**Computer Applications**]: Social and behavioral sciences

General Terms

Design, Human Factors, Measurement

Keywords

Twitter, who is who, Lists, topic inference, crowdsourcing

1. INTRODUCTION

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. The democratization of content authoring has contributed tremendously to the success of Twitter, but it also poses a big challenge: *how can microbloggers tell who is who in Twitter?* Knowing the credentials of a Twitter user can crucially help others determine how much trust or importance they should place in the content posted by the user.

In this paper, we present the design and evaluation of a novel *who-is-who* inference system for users on the popular Twitter microblogging site. Figure 1 shows an illustrative tag cloud of *attributes* inferred by our service for Lada Adamic, who is an active Twitter user and a well-known researcher in the area of social networks [9]. Note that these attributes not only contain her biographical information (she is a *professor* at *umsi, umich* – University of Michigan's School of Information), but also capture her expertise (she is an expert on *social media*, *network-analysis*, *social networks*, *csresearch*, *hci*, *statphysics*) as well as popular perceptions about her (she is a *bigname*, a *thinker*, and a *goodblogger(s)*.)

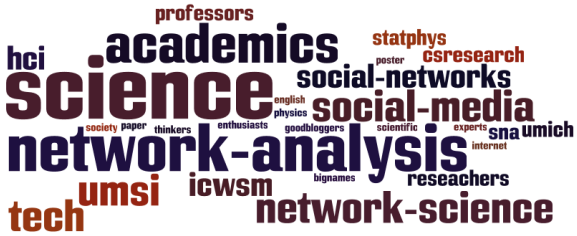

Figure 1: Attributes inferred for Lada Adamic, a noted social network researcher, by our who-is-who system

Existing approaches to infer topics related to a user rely either on the profile information provided by the user herself (e.g., name and bio) or on analyzing the tweeting activity of the user [13, 14]. The problem with relying on bios is that many users do not provide sufficiently informative bios, and worse, the information provided by the users is mostly unvetted. The problem with analyzing tweets to infer users' attributes is that tweets often contain conversation on day-to-day activities of users [6], which makes it difficult to extract accurate topical information about the users (as we show in Section 2). Compared to existing approaches, the attributes inferred by our *who-is-who* service provide a more accurate and comprehensive characterization of Twitter users.

In this paper, we take a different approach to construct the *who-is-who* service for Twitter users. We exploit the *Lists* feature on Twitter, which allows users to group together Twitter accounts posting on a topic of interest to them, and

WOSN'12, August 17, 2012, Helsinki, Finland.

follow their collective tweets. We observe that many users carefully curate their Lists, generating meta-data, such as List names and descriptions, that provide valuable semantic cues about who the users included in the Lists are. Our key idea is to analyze the meta-data of the Lists containing a user to infer the user's attributes. By crowdsourcing our inference, we discover a more accurate and comprehensive set of attributes related to a user, which is often richer than the bio or tweets posted by the user herself.

To evaluate our methodology, we gathered List data for all 54 million Twitter users who joined the network before August 2009. We find that there is sufficient List meta-data to automatically infer attributes for over 1 million users in the Twitter network. These users include nearly 80% of the top 1 million most widely followed Twitter users. We used the List data to build our *who-is-who* system and we deployed it for public testing at `http://twitter-app.mpi-sws.org/who-is-who`. We encourage interested readers to test the system for themselves. User feedback from our public deployment indicates that our attribute inference for Twitter users is both accurate and comprehensive. We argue that our system provides a fundamental building block for future designers of content or people search and recommendation services.

2. RELATED WORK AND MOTIVATION

Like the Web, Twitter has become a popular platform for finding news and information. 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 [5]. It has also been shown that topic discovery of webpages could be improved by exploiting *social annotations*, that is, annotations of webpages provided by human users in social tagging sites such as Delicious [19].

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.* [13] used Latent Dirichlet Allocation to map the contents of a tweet stream into topical dimensions. Kim *et al.* [8] used chi-square distribution measure on the tweets posted by users included in a common List to identify topics of interest of the users. However, prior research has shown that tweets often contain conversation on day-to-day activities of users [6], 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 [10, 12] or search engines [3] using words extracted from tweets.

Additionally, efforts to identify experts on specific topics in Twitter attempt to judge whether a user is related to a given topic. Weng *et al.* [17] and Pal *et al.* [11] use features extracted from the Twitter graph and the tweets posted by users to identify whether a user is related to a given topic. Similarly it has been reported [15] that Twitter's own "Who To Follow" service [14] uses the profile information (e.g., name and bio) provided by the users to identify experts on a given topic.

Thus, existing attempts to discover attributes for Twitter users [8, 10, 13, 14] rely on analyzing tweets or bios posted by users themselves, analogous to examining the contents of web-pages. In this work, we explore an alternative approach which relies on leveraging crowdsourced social annotations, which are gathered from Twitter Lists feature. Though the main purpose of Lists is to help users organize the tweets they receive, we show that the feature can be effectively exploited to derive social annotations, that is, how the Twitter crowd views and describes other users.

Table 1 illustrates the advantages with our approach. It compares the quality of information that can be extracted from users' bio, tweets, and Lists for some well-known Twitter users. To infer a user's attributes from her tweets and the Lists containing them, we extracted the most frequently repeated nouns and adjectives in the tweets and List meta-data and removed common stop-words. Note that many popular users either do not provide a bio, or have a bio which does not provide any topical information. Sometimes the bios may be misleading – for instance, the well-known comedian Jimmy Fallon has mockingly described himself as an astrophysicist in his bio. Further, for many users, tweets primarily contain day-to-day conversation [6], making it difficult to identify meaningful topics. For example, for the popular actor, Ashton Kutcher, none of the top words from tweets describe who he is. However, in all cases, words extracted from crowdsourced Lists identify the user's attributes accurately and comprehensively.

Thus, social annotations provide a rich source of information to characterize a user. Recently, their utility has been explored by Bernstein *et al.* [2] who proposed a game on Facebook that encouraged friends to annotate one another. A few studies [16, 18] have used Twitter Lists to identify users related to a small number of selected topics, such as, celebrities and media sources. However, to the best of our knowledge, ours is the first large-scale attempt to discover attributes for users in a social network using existing social annotations.

3. INFERRING *WHO-IS-WHO* FROM LISTS

In order to help users organize their followings and the information they post, Twitter introduced a new feature called *Lists* [7] at the end of 2009. By creating a List, a user can group together some Twitter users, so that all tweets posted by the grouped users can be viewed in the List timeline. To create a List, a user needs to provide the List name (free text, limited to 25 characters) and optionally add a List description. For instance, a user can create a List called 'celebrities' and add celebrities to this List. Then, the user can view tweets posted by these celebrities in the List timeline. In this section, we first describe how we gathered List data and then discuss how we extract user attributes from the data.

3.1 Twitter dataset

The dataset used in this work includes extensive data from a previous measurement study [4] 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 billion

User	Extracts from Bio	Top words from tweets	Top words from Lists
Barack Obama	Account run by #Obama2012 campaign staff. Tweets from the President are signed -bo	health, visit, American, vote, event, Iowa, debate, reform, president	politics, celebs, government, famous, president, media, news, barack obama, leaders
Ashton Kutcher	I make up stuff, stories mostly, collaborations of thoughts, dreams, and actions. Thats me.	love, daily, people, time, great, gui, movie, video, happy, life	celebs, actors, famous, movies, stars, comedy, funny, music, hollywood, pop culture
Jimmy Fallon	astrophysicist	#fallonmono, happy, love, fun, #fjoln, roots, funny, video, song, game, hope	celebs, funny, famous, humor, music, movies, laugh, hollywood, comics, television, entertainers

Table 1: Comparing three possible approaches for identifying attributes of a Twitter user – (i) from the account bio (ii) from tweets posted by the user (iii) from Lists containing the user

List Name	Description	Members
News	News media accounts	nytimes, BBCNews, WSJ, cnnbrk, CBSNews
Music	Musicians	Eminem, britneyspears, ladygaga, rihanna, BonJovi
Tennis	Tennis players and Tennis news	andyroddick, usopen, Bryanbros, ATPWorldTour
Politics	Politicians and people who talk about them	BarackObama, nprpolitics, whitehouse, billmaher

Table 2: Examples of Lists, their description, and some members

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 [4].

3.2 Crawling Lists

Lists were introduced in Twitter *after* our Twitter dataset was collected. 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. We refer to this top-20% most listed users as the *top-listed* set of users, and we focus our study on these users in the next sections.

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 limited 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 (i.e., 34.6 %) Lists had a description, while the others had only the List name. Table 2 presents illustrative examples of Lists, extracted from our dataset. We can immediately note that the List names and descriptions provide valuable semantic cues about who the members of the Lists are.

3.3 Using Lists to infer user attributes

Our strategy to discover attributes that characterize a given Twitter user consists of extracting frequently repeated words from the names and description of the Lists that include the user. More specifically, we apply the following five processing steps. (1) Since List names cannot exceed 25 characters, multiple words are frequently combined using *CamelCase*, e.g., TennisPlayers. We separate them into individual words. (2) We apply common language processing approaches like case-folding, stemming, and removing stop words. In addition to the common stop words, we also filter out a set of domain-specific words, such as Twitter, list, and formulist–a tool frequently used to automatically create Lists. (3) Prior research on social annotations showed that nouns and adjectives are especially useful for characterizing users [12]. So we used a standard part-of-speech tagger to identify nouns and adjectives. (4) A number of List names and descriptions are in languages other than English. We observed several cases where the same topic is represented using different words in different languages, and these words are *not* unified even by stemming, e.g., political and *politicos*. Hence we group together words that are very similar to each other based on edit-distance among words. (5) Finally, as List names and descriptions are typically short, we consider only unigrams and bigrams (2-word phrases) as candidates for attributes.

The above strategy produces a set of attributes for each user as well as the relative importance of the attributes, based on how frequently they appeared in different Lists containing the user. In the next section, we evaluate how well the inferred attributes characterize individual Twitter users.

4. INFERENCE QUALITY

In this section, we evaluate the quality of the attributes inferred by the List-based methodology. There are two aspects to consider when evaluating quality – (i) whether the attributes inferred are *accurate*, and (ii) whether the attributes inferred are *informative*. A set of attributes inferred for a user may have high accuracy, but low information content and vice-versa. For example, if the set of attributes inferred for Barack Obama contained the single attribute 'politician', the inference would indeed be highly accurate, but it is not informative and is of limited practical usage.

In order to evaluate whether the inferred attributes are accurate and informative, we need to compare our results with ground truth attributes for some Twitter users. Since such ground truth is difficult to obtain for a random set of Twitter users, we adopt two strategies for our evaluation. First, we evaluate the attributes inferred for a set of *popular* Twitter accounts, for whom the relevant attributes are generally well-known or easily verifiable. Second, we col-

User	Top attributes inferred from Lists		
	Biographical information	**Topics of expertise**	**Popular perception**
Well-known users			
Barack Obama	government, president, usa, democrat	politics	celebs, famous, leaders, current events
Lance Armstrong	sports, cyclist, athlete	tdf, triathlon, cancer	celebs, famous, influential, inspiration
Linux Foundation	computer, linux, tech, open source, software	libre, gnu, ubuntu, suse	geek
News media			
The Nation	media, journalists, magazines, blogs	politics, government	progressive, liberal
townhall.com	media, bloggers, commentary, journalists	politics	conservative, republican
SportsBusiness Daily	media, journalists, bloggers	sports, football, athletes, nba, baseball, hockey, nhl	experts, influencers
Guardian Film	guardian, media press, journalists, reviews	movies, cinema, theatre, actors, directors, hollywood	film critics
US Senators			
Chuck Grassley	politics, senator, congress, government, republicans, iowa, gop	health, food, agriculture	conservative
Claire McCaskill	politics, senate, government, congress, democrats, missouri, women	tech, cyber-crime, security, power, health, commerce, military policy	progressive, liberal
Jim Inhofe	politics, senators, congress, republican, government, gop, oklahoma	army, energy, climate, foreign	conservative

Table 3: Examples of top attributes (biographical information, topics of expertise and popular perception) for some example popular users, as inferred by the List-based methodology. All attributes are case-folded to lower case.

lect feedback from human volunteers on the quality of the inferred attributes for a number of different users.

4.1 Evaluation of popular accounts

As popular accounts, we consider (i) a set of well-known Twitter users / business accounts for whom relevant attributes are generally well-known, (ii) news media sites, and (iii) US senators, for whom detailed and authentic information is available in the Web, e.g., in the corresponding Wikipedia pages. Table 3 shows the top attributes inferred by our methodology for some of these users.

Well-known users: For the well-known accounts in Table 3, it is evident that the inferred attributes accurately describe the users. Moreover, the set of inferred attributes include not only biographical information of the users, but also their topics of expertise, and even the popular perception of these users. For instance, the inferred attributes for Lance Armstrong not only contain his biographical information that he is a *sports* person and a *cyclist*, but they also indicate more specific topics of expertise such as Tour de France (*tdf*) and *cancer*.[1] Further, the attributes also capture the popular perception that Lance Armstrong is a *famous celeb* and *inspirational*. Similarly, for the business account 'Linux Foundation', the inferred attributes not only say that this account is related to *computers*, *tech*, *software* and *Linux* but also indicates more specialized attributes related to Linux (*libre*, *gnu*, *ubuntu*). This illustrates both the accuracy and the rich information content of the set of attributes inferred using crowdsourced Lists.

News media sources: For media accounts in Twitter, the inferred attributes not only indicate that they are news media (biographical information), but also indicate the specific topics the media focuses on, such as, *politics* for The Nation and townhall.com, *sports* for SportsBusinessDaily and *movies* for Guardian Film. Furthermore, for political news media sources, the attributes also indicate the perceived *political bias* of the media sources (if any), e.g. *progressive* for The Nation and *conservative* for townhall.com. In total, we observed the attributes for 36 political media sources, out of which 6 were inferred to be conservative, 11 were inferred to be progressive/liberal and the rest did not have any inferred bias. These inferences were verified using ADA scores [1] and the Wikipedia pages for the corresponding media sites, and the inferences were found to be correct for all 6 conservative media sources and for 7 of the 11 progressive / liberal media sources.

US Senators: Since a large amount of authoritative information on the US senators is readily available, we chose to demonstrate the quality of the inference over the Twitter accounts of US senators. Out of the current 100 US senators, 84 have Twitter accounts. We obtained the main attributes for each of these senators and analyzed their accuracy and information content (some examples are shown in Table 3.) (i) *Biographical information*: Apart from identifying that they are *politicians* and *senators*, our *who-is-who* system accurately identified the *political party* to which each of the 84 senators belonged to. In total, 41 were inferred to be democrats and 41 to be republicans, which is in agreement with their Wikipedia pages. The other two US senators (Joe Lieberman and Bernie Sanders) had both the attributes 'democratic' and 'independent' very prominently, which is accurate as they are pro-democratic independents. The attributes also correctly identified the *states* represented by each of the 84 senators (see Table 3) as well as their genders– the attribute 'women' or 'female' was found for each of these 15 female senators. (ii) *Topics of expertise*: Senators tend to be members of senate committees on the topics that they have expertise or interest in. Our List-based method correctly identified a number of *senate committees* of which each senator is a member, as shown in Table 4. We verified the committee memberships from the

[1]Note that the topics of expertise can be thought of as part of an extended biographic information as well.

Senate committee	# senators on Twitter	# correctly identified
Super-committee	5	5
Appropriations	22	21
Banking, Housing & Urban Affairs	17	12
Budget	20	16
Commerce, Science & Transportation	24	21

Table 4: Senate committee memberships of US senators, identified by our *who-is-who* system.

Wikipedia pages on the respective committees. (iii) *Popular perception*: Apart from accurately inferring several interesting factual information about the US senators, the set of inferred attributes also reveal the public perception about the *political ideologies* of the senators. We were able to identify several attributes that indicate a certain political ideology, such as progressive, liberal, libertarian, conservative, and tea-party. For instance, most democrats were inferred to be progressive or liberal; the only democrat who was inferred to be conservative was Joe Lieberman, who is described as neo-conservative in Wikipedia.

4.2 Evaluation using human feedback

Since the attributes associated with a person or a Twitter user are inherently subjective by nature, one of the best ways of evaluating the quality of inferred attributes is through human feedback. In order to evaluate our approach, we deployed our *who-is-who* service on the public Web at `http://twitter-app.mpi-sws.org/who-is-who/`. Anyone can access the service by entering the name of a Twitter user and see a *word cloud* of the top 30 attributes inferred for the chosen user. We advertised the URL to other researchers in the home institutes of the authors, inviting them to evaluate the system. Evaluators of the service are shown the word cloud (of the top 30 attributes) for Twitter users and are asked to judge whether the inferred attributes are (i) accurate and (ii) informative. For both questions, the evaluator chooses from Yes / No / Can't Tell feedback options. An evaluator is allowed to choose from the three sets of Twitter users described above (well-known users, news media sites, and US senators). Alternately, an evaluator can choose to provide feedback for any user that she is interested in.

In total, we obtained 345 evaluations for *accuracy* of tag clouds, out of which 53 chose the 'Can't Tell' option. These are evaluations, where the evaluator does not know the Twitter user sufficiently well to rate the accuracy of the user's inferred attributes. Ignoring these 53 evaluations, 274 (94%) of the 292 remaining evaluations were rated as accurate. Next, we investigated the small number of Twitter users for whom our inference received one or more negative evaluations. Interestingly, every single one of these Twitter users received more positive evaluations than negative evaluations, highlighting the subjectiveness in accuracy judgements. Thus, not only is our inference highly accurate for most Twitter accounts, but also the occasional negative evaluations are subjective in nature and are always outvoted by positive evaluations.

The evaluations for *informativeness* of inferred attributes are very similar. In total, we obtained 342 evaluations for informativeness of tag clouds, out of which 45 chose the 'Can't Tell' option. Ignoring these 45 evaluations, 277 (93%) of the 297 remaining evaluations were rated as informative. Once again, analysis of the Twitter users with one or more negative evaluations shows that every single one of them received more positive evaluations than negative ones. Thus, feedback from human evaluators indicates that our inference is not only highly accurate but also quite informative for most Twitter accounts.

5. INFERENCE COVERAGE

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 2 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.

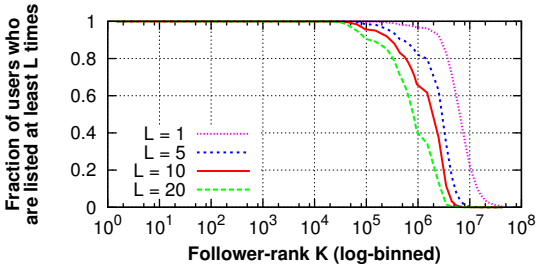

Figure 2: Fraction of users who are listed at least L times vs follower-rank

Users with large numbers of followers: As shown in Figure 2, 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 million 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 million 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 5. 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

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

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.

In summary, 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. Finally, our analysis of the top-listed users with very few followers suggests potential misuse of the Lists feature.

6. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel methodology 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 user's tweets or profile, we infer topics by leveraging the wisdom of the Twitter crowds, as reflected in the meta-data 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.

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.

Finally, our current methodology is vulnerable to List spamming attacks, where malicious users can create fake Lists to manipulate the inferred attributes for a target user. While we did not find much evidence of List spamming to date, such attacks could be launched in a straightforward manner. To make our methodology robust against List spam, we plan to consider the reputation of the users who create the Lists in future research.

7. ACKNOWLEDGEMENT

The authors thank Dr. Parag Singla, IIT Delhi, for his constructive suggestions, and the anonymous reviewers whose comments helped to improve the paper. This research was supported in part by a grant from the Indo-German Max Planck Centre for Computer Science (IMPECS). Fabricio Benevenuto is supported by Fapemig.

8. REFERENCES

[1] Americans for Democratic Action. `www.adaction.org`.

[2] M. Bernstein, D. Tan, G. Smith, M. Czerwinski, and E. Horvitz. Collabio: a game for annotating people within social networks. In *ACM symposium on User interface software and technology (UIST)*, 2009.

[3] M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi. Eddi: interactive topic-based browsing of social status streams. In *ACM symposium on User interface software and technology (UIST)*, 2010.

[4] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *AAAI Conference on Weblogs and Social Media (ICWSM)*, May 2010.

[5] S. Dill et al. Semtag and seeker: bootstrapping the semantic web via automated semantic annotation. In *ACM Conference on World Wide Web (WWW)*, 2003.

[6] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *WebKDD and SNA-KDD workshop on Web mining and Social Network Analysis*, 2007.

[7] N. Kallen. Twitter blog: Soon to Launch: Lists. `http://tinyurl.com/lists-launch`, Sep 2009.

[8] D. Kim, Y. Jo, I.-C. Moon, and A. Oh. Analysis of Twitter Lists as a Potential Source for Discovering Latent Characteristics of Users. In *ACM CHI Workshop on Microblogging*, 2010.

[9] Lada Adamic – University of Michigan. `www.ladamic.com`.

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

[11] A. Pal and S. Counts. Identifying topical authorities in microblogs. In *ACM Conference on Web Search and Data Mining (WSDM)*, 2011.

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

[13] D. Ramage, S. Dumais, and D. Liebling. Characterizing Microblogs with Topic Models. In *AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010.

[14] Twitter: Who to Follow. `twitter.com/who_to_follow`.

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

[16] M. J. Welch, U. Schonfeld, D. He, and J. Cho. Topical semantics of Twitter links. In *ACM Conference on Web Search and Data Mining (WSDM)*, 2011.

[17] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In *ACM Conference on Web Search and Data Mining (WSDM)*, 2010.

[18] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts. Who says what to whom on Twitter. In *ACM Conference on World Wide Web (WWW)*, 2011.

[19] X. Wu, L. Zhang, and Y. Yu. Exploring social annotations for the semantic web. In *ACM Conference on World Wide Web (WWW)*, 2006.