

How People Describe Themselves on Twitter

Konstantinos Semertzidis
Dept. of Computer Science
University of York, UK
ks989@york.ac.uk

Evaggelia Pitoura
Dept. of Computer Science
University of Ioannina, Greece
pitoura@cs.uoi.gr

Panayiotis Tsaparas
Dept. of Computer Science
University of Ioannina, Greece
tsap@cs.uoi.gr

ABSTRACT

Twitter, being both a micro-blogging service and a social network, has become one of the primary means of communicating and disseminating information online. As such, significant amount of research has been devoted to analyzing the Twitter graph, the tweets, and the behavior of its users. In this work, we undertake a study of the user profile bios on Twitter. The goal of our study is two-fold: first, to understand what Twitter users choose to expose about themselves in their profile bio, and second, to investigate if it is possible to exploit the information in the user bio for tasks such as predicting connections between Twitter users.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous;
H.2.8 [Database Applications]: Data Mining

General Terms

Experimentation, Measurement

Keywords

Twitter bio, Link prediction

1. INTRODUCTION

The evolution of the Internet in the past few years has been marked by the emergence of online social networks and media which have altered significantly the habits and behavior of online users. Today information between users is exchanged mostly via social networks like Facebook and Twitter. Twitter is an especially interesting case, since it combines the features of both a social network and a microblogging service and has become the primary medium for communicating breaking news and making real-time commentary, often supplanting news agencies and portals.

Due to the high importance of Twitter as both a social network and news medium, it has been the focus of several research studies. The structure of the graph, the contents of

tweets and the propagation of content via Twitter have been studied extensively (e.g., see [6, 2, 10]). There has been also work on predicting future connections between Twitter users, as part of the general problem of link prediction on social networks [7, 12, 1, 3].

Twitter provides a bio box, where users can give some information about themselves in fewer than 160 characters. When filled in by the user, this box is used in many different ways: to provide demographic information such as age, marital status, location, or occupation; to state a preference, an interest or some personal information; to make a distinctive comment and remark; as another tweet that can be about anything. Similar to the tweets, it can reveal information about the user interests and identity, and it is interesting to understand it further.

In this work, we undertake a study of the user profile bios on Twitter. To the best of our knowledge, this is the first detailed study of the Twitter bios. Our goal is first to study what Twitter users expose on their bio, and then investigate if it is possible to use the bio information for tasks such as link prediction. Predicting follower relationship is a task of great importance for online social networks like Twitter. It is a means for growing the network and increasing the engagement of the users, but it is also a way for advertisers to expand their user base. Sponsored follower recommendation is one of the monetization strategies for Twitter

The contributions of this work are the following.

- We provide a detailed analysis of the content of the bios of Twitter users. Our goal is to understand what Twitter users choose to say about themselves in the bio box. We study the keywords that are common in Twitter bios, and categorize them according to the role they play in the user bio.
- We study the correlation between the information in the bio of connected users. More specifically, we address the following question: Is there any noticeable similarity in the way connected users describe themselves? We demonstrate that this is actually the case, connected users are more likely to have similar bios than random users. This similarity becomes stronger when the relationship is reciprocal.
- In life, the friends you choose often describe you better than you describe yourself. We study if this is true in Twitter, by creating an extended bio description which aggregates the bio profiles of the followings of a user. We perform a similar study as for bios and observe that this extended bio yields much higher similarity

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DBSocial '13 New York, NY USA

Copyright 2013 ACM 978-1-2191-4 ...\$15.00.

Table 1: Basic Dataset Statistics

Number of users	553,690
Users without bio	105,041 (18.97%)
Users with unreadable bio	3,027 (0.55%)
Users with readable bio	445,622 (80.48%)
Average bio length	87

for connected users than the bios that users expose for themselves.

- Given the similarity between bios of connected users, we study whether bios can be used for link prediction. In our experiments, we compare against methods that use only the neighborhood information. We observe that the bio (especially, the extended bio) provides a useful signal for predicting connections between users, although not as strong as that of the neighborhood.

The rest of the paper is structured as follows. In Section 2, we present related work. In Section 3, we provide an analysis of the bio characteristics. In Section 4, we study bio similarity between connected users and the use of bio information for link prediction. Section 5 concludes the paper.

2. RELATED WORK

Twitter has been the focus of much analysis from its early stage, e.g., [5]. There is a large body of work on both the underlying followers-following graph and the content of information propagated through it. The repertoire of topics studied is wide including amongst others: understanding and characterizing influence and influencers, e.g., [2, 9], monitoring real time interactions for event identification, e.g., [13, 4] and trend detection, e.g., [10]. The authors of [6] provide a study of the entire Twitter-sphere including information diffusion, the Twitter topology and influential users.

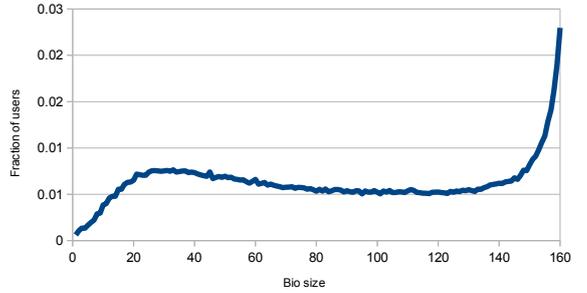
In this paper, we look into one of the less studied aspects, that of user bios. Very recently, bios were used along with other features in the context of topical expertise of Twitter users, namely, to form computational expertise models of Twitter users and assist expertise judgments by humans [14], and were shown to enhance the credibility of a tweet [11].

Link prediction in social networks has also attracted a lot of attention. A survey of various link prediction algorithms for a variety of complex networks is provided in [8]. For social networks, methods presented in [7] focus on node proximity in the underlying graph, such as graph distance or number of common friends, defined from both a local and a global perspective. The problem has also been considered for Twitter, in terms of understanding the link formation process [12, 1] and predicting reciprocal links [3]. Here, we investigate whether similarity of bios may play a role.

3. ANALYSIS OF BIO CONTENT

In this section, we study the characteristics of the Twitter bios. For our study, we used a sample of 10K users from the set of users collected in [2] and all the users that they follow. This amounted to a set of 553,690 users. Using the Twitter API¹, we crawled the profile information for these users, which also contains the bio information. The statistics for

¹<https://dev.twitter.com/docs/api/1.1/get/users/show>

**Figure 1: Distribution of bio length**

our findings are shown in Table 1. Out of the 553,690 users, 445,622 (80.48%) had a readable non-empty bio. Figure 1 depicts, for each bio length, the percentage of users with a non-empty, readable bio that have bios of this length. As we can see, the most frequent case is that users make use of the full 160 characters for describing themselves.

We now turn to the content of the bios. We did some basic preprocessing, where we removed all characters that are not letters or numbers. We tokenized into words, and threw out stop-words, since they do not have any information content. We also removed duplicates, so each bio becomes a collection of distinct words. Given this representation, we computed the frequency of each word, that is, the number of bios that the word appears in. The top-30 most frequent words are shown in Table 2.

It is surprising to observe that the word “love” is the most frequent word. However, to disappoint the romantics, after examining the bios, most of the times the word love does not refer to the sentiment of love, but it is rather used to express a strong interest of a user (e.g., “i love hokey”). It is common for Twitter bios to describe some interest of the user and there are several words in the top-30 around the user interests (e.g., “music”, “fan”, “food”). Several users use the bio to state their profession (e.g., “marketing”, “author”, “artist”, “designer”). There also are many words that are related to Twitter itself, to Web, or to social networks in general (e.g., “twitter”, “follow”, “tweets”, “internet”). Some examples of bio descriptions are shown in Table 3.

Although there is an impressive variety and uniqueness in the way people express themselves in the bio, there were also some obvious common patterns. To make our observations more concrete, we looked at a larger set of words. We took the top-200 most frequent words, and after laborious manual inspection of the bios, we categorized them into the six categories shown in Table 4 and explained below. In choosing the categories, we looked into the general themes of the user bios.

- *Occupation* refers to words that are used to describe the occupation of the user.
- *Interests/Preferences/Hobbies* refers to words that are used to describe some interest or past-time of the user.
- *Personal Info* refers to words that are used to describe personal information such as family status, religion, pets etc. It is very common for people to state their marital status, or the fact that they are proud parents in the bio.

Table 2: Top-30 most frequent words

Rank	Word	Frequency	Fraction
1	love	33,691	0.0067
2	life	17,631	0.0047
3	music	15,993	0.0043
4	twitter	15,915	0.0043
5	follow	15,349	0.0041
6	news	14,190	0.0038
7	media	12,702	0.0034
8	social	12,315	0.0033
9	marketing	12,195	0.0023
10	world	10,877	0.0029
11	business	10,567	0.0028
12	official	9,832	0.0026
13	people	9,418	0.0025
14	writer	9,308	0.0025
15	online	8,810	0.0024
16	lover	8,180	0.022
17	live	8,130	0.022
18	web	8,093	0.022
19	fan	7,706	0.0021
20	time	7,331	0.0020
21	author	6,851	0.0018
22	design	6,665	0.0018
23	free	6,599	0.0018
24	artist	6,424	0.0017
25	good	6,186	0.0017
26	living	5,927	0.0016
27	internet	5,881	0.0016
28	tweets	5,875	0.0016
29	designer	5,852	0.0016
30	food	5,847	0.0016

Table 3: Bio examples**Bio examples**

- i love all things sporting and sports related, like my cervelo bike!
- love good music, great movies, going on vacation, good food, reading inspiring books, spring time, earning
- self employed silicon jockey. married a good wife, have a good dog, what more can u ask 4? chk my blog when you get time for more info.
- the official twitter of the wikistupida facts site!
- new and successful day trader. no b.s. make real income in less than 10 minutes a day.
- original and unbiased news and opinion on events happening in the digital media and internet marketing world.
- librarian interested in digital curation, library technology & social networking. gadget lover
- linux journal editor. writer. geek.
- proud new england native. always interested in solid story ideas. football obsessed.
- hi, my name is brian and im a mac. im a technology and video game enthusiast. i create my own electronic music. i enjoy water skiing and riding wave runners!

Table 4: Top-200 frequent word categories

<i>Occupation</i>	
social, marketing, business, writer, photographer, cofounder, author, design, artist, internet, designer, founder, science, online, web blogger, director, student, art, work, singer, professional, digital, entrepreneur, tv, editor, radio, fitness, working, fashion, technology, producer, manager, teacher, university, consultant, company, coach, owner, writing, website, ceo, management, magazine, dj, book, expert, development, personal, services, video, speaker, agency, service, film, developer, college, journalist, host, model, software, wine, industry, products, marketer, write	
<i>Interests/Preferences/Hobbies</i>	
love, music, writer, lover, fan, enthusiast, video, book, coffee, food, blogger, sports, blog, games, photography, film, books, travel, tv, technology, loves, stuff, football, arts, software, tech, art, geek, entertainment, rock, play, wine, loving, dog, write, writing, science, fitness	
<i>Personal Info</i>	
wife, mom, family, girl, husband, father, god, creative, years, geek, guy, mother, young, beautiful, enjoy, proud, year, jesus, passionate, school, learning, expert, loving, happy, christian, view, dog, culture, friend, passion, player	
<i>Social Networks/Internet</i>	
twitter, follow, official, people, tweets, account, tweet, email, page, site, network, facebook, join, mobile	
<i>Community</i>	
news, media, community, events, information, updates, info, member, email, contact, join, club, visit	
<i>General</i>	
life, world. live, time, free, good, living, real, day, home, fun, team, great, fashion, find, health, creative, big, show, latest, man, friends, money, based, helping, check, personal, local, city, making, public, group, stuff, tips, 2012, daily, rock, kids, international, support, heart, small, full, leading, high, call, change, education, top, national, dream, london, dedicated, source, estate, share, independent, global, culture, place	

Table 5: Category popularity

Category	Word Fraction
1 <i>Occupation</i>	7.5%
2 <i>Interests/Preferences/Hobbies</i>	4.6%
3 <i>Personal Info</i>	2.6%
4 <i>Social Networks/Internet</i>	2.3%
5 <i>Community</i>	1.5%

- *Social Networks/Internet* refers to words that are Twitter or internet related. These can be thought of as Twitter stop-words, since they do not carry a lot of information about the user.
- *Community* refers to words that have to do with communities that people belong to, such as being members of different clubs, or participating in organized events.
- *General* includes all words that do not fall into any of the other categories. These are often words that are too specific to define a general category.

Table 5 orders the categories (except from *General*) according to the fraction of words belonging to each category. As we can see, when creating their bio, users are more likely to talk about their occupation and then their interests. This is not surprising, since many users use Twitter for professional purposes (e.g., journalists, politicians, athletes). It is also common for people to use their interests to define themselves online. When people use Twitter more as a social networks to keep in touch with their friends, then they are likely to include personal information, such as their marital status, religion, or the community they belong to.

4. BIO SIMILARITY AND LINK PREDICTION

In the previous section, we studied the characteristics of the profile descriptions. In this section, we study how the content of the bio relates to the connections of the user in the Twitter network. Is it the case that connected users have similar bios? And can similarity in the bio content be used for predicting a connection? We break our analysis into two steps. First, we show that connected users are indeed more likely to use common words in their profile than random users. Then, we show that similarity can be used as a signal for predicting a connection between two users.

In our analysis, apart from the bio information that a user exposes for herself, we also consider an inferred bio that is constructed by aggregating the profile descriptions of the users that the user in question follows. We will refer to this profile description as the *following bio description* of the user, or the *f-bio*. In many ways, a following bio is more informative than the bio description of the user. It is richer in content, and it expresses the interests of the user as they are reflected by her choice in the users that she follows. Note also, that it is possible to construct a following bio even for users that have not specified a bio for themselves.

4.1 Bio Similarity

We first address the following question: Are connected users more likely to have similar bios than random users? We consider two types of connections: (i) the simple follow relationship, where user u follows user v , but v does not follow back and (ii) the mutual follow relationship, where user u follows user v and user v follows back. The latter is more indicative of a true social connection between two users [3].

To ensure that we have a rich dataset for this task, where we have many users with descriptions, and many follow and mutual follow relationships, we crawled a new larger dataset, sampling 300,000 users that all have non-empty descriptions, and no more than 5,000 followers/followings (as a

Table 6: Average bio similarity

	Avg S_b	Conf-Int
Non-connected Pairs	0.00514	$\pm 1.61 \cdot 10^{-13}$
Simple Follow Pairs	0.01825	$\pm 3.30 \cdot 10^{-13}$
Mutual Follow Pairs	0.02278	$\pm 2.14 \cdot 10^{-12}$

Table 7: Average f-bio and bio similarity

	Avg S_f	Conf-Int	Avg S_b	Conf-Int
Non-connected Pairs	0.44871	$\pm 1.18 \cdot 10^{-12}$	0.00574	$\pm 1.41 \cdot 10^{-13}$
Simple Follow Pairs	0.78697	$\pm 6.99 \cdot 10^{-7}$	0.01898	$\pm 2.13 \cdot 10^{-7}$
Mutual Follow Pairs	0.83202	$\pm 2.04 \cdot 10^{-8}$	0.01912	$\pm 7.20 \cdot 10^{-9}$

rough threshold for removing special types of users, such as celebrities). We found all the follow relationships between these users, resulting in 595,573 simple follow relationships, and 245,828 mutual follow relationships.

To measure the similarity between two bio descriptions, we use the cosine similarity. For a user u , let $B(u)$ denote the vector that represents the bio of the user. $B(u)$ has length equal to the total number of words in all descriptions. For simple bios, we remove multiple occurrences of a word in the profile, thus the resulting $B(u)$ is a 0/1 vector, with value 1 if the corresponding word appears in the profile and 0, otherwise. The bio similarity $S_b(u, v)$ of users u and v is defined as follows:

$$S_b(u, v) = \frac{B(u) \cdot B(v)}{\|B(u)\| \|B(v)\|}$$

Bio similarity takes values between zero and one, and it is zero if the two users do not share any word, and one if the two vectors are identical.

Table 6 shows the average bio similarity for different types of pairs of users from our study set. Bios were pre-processed as described in Section 3 (removed special symbols and stop-words). The first row is the average similarity over 595,578 randomly sampled non-connected pairs. The second row is the average similarity of pairs of users where u follows v but v does not follow u . The third line shows the average similarity between two users that mutually follow each other. We also include the confidence intervals for confidence level 95%. As we can see connected users have more similar bios compared to non-connected users, and the difference is statistically significant. Similarity increases as the strength of the relationship between the pair of users increases. That is, if user u follows user v , then it is more likely that they share similar words in their description than two non-connected users. If user v reciprocates the follow link, then the two users are even more likely to share common words than two users that have a one-way connection. This increase in similarity is also statistically significant.

The fact that there is a statistically significant similarity between the bio descriptions of two connected users, even for one-way connections is particularly interesting. Our finding indicates that users are likely to follow other users who describe themselves in a similar way. This is consistent with the principle of homophily observed in social networks. Our analysis demonstrates that bio descriptions contain enough signal to capture this phenomenon.

We perform the same analysis for the bio description con-

constructed by the bios of the followings (the f-bio). We construct the following bio as follows. Let u be a user, and let $N(u)$ denote the followings of user u . We define the following bio $fB(u)$ of user u as $fB(u) = \sum_{v \in N(u)} B(v)$, i.e., the sum of all the bios of the followings of u . Note that contrary to $B(u)$, $fB(u)$ may contain a word multiple times. This is intentional, since multiple occurrences of a keyword in the f-bio means that the user follows many users that are described by this word, and thus she has a stronger affiliation with this word, and we want the f-bio to capture this. We measure the similarity between two f-bios using again the cosine similarity. We use $S_f(u, v)$ to denote the cosine similarity between the f-bios of two users u and v .

In order to compute the f-bio we need to crawl the bio description of all the followings of a user. Given that each user has on average around 100 followings this is an expensive process. We thus restricted ourselves to a set of 1,126 users, consisting of the users that have at least 10 reciprocal connections. This is the same dataset that we use next for link prediction. Table 7 shows the results. The trends are the same as before, the stronger the relationship the higher the similarity. To compare the f-bio with the bio description, we also computed the average bio similarity for the same set of users. We observe that the similarity of the f-bio is significantly higher than that of bio. It appears that the bios of the followings of a user provide a better description of the individual than the bio that the user picks for herself. As the saying goes, your friends say a lot about you, and this is true also in the case of Twitter.

4.2 Link Prediction Using Bio

Given the increased likelihood that two users share common words in their description if they are connected on Twitter, it is reasonable to ask if we can make use of this description similarity in order to predict links between users. In this section we address this question. We consider three different link prediction tasks:

- **Follower prediction:** In this case, given a pair of users u , and v , we want to predict if there is a follower relationship between u and v (in either, or both directions).
- **Follow-back prediction:** In this case given a pair of users where u follows v , we want to predict if user v will follow user u back. It is often important for users to follow someone who will follow back. It is also possible for corporate accounts to follow users, so that they will follow them back.
- **Mutual-following prediction:** In this case, given a pair of users u , and v , we want to predict if user u will follow v and user v will follow u . Mutual relationships in Twitter are often perceived as the true social relationships, or friendships. It is useful to be able to predict such relationships.

For each prediction task, we predict that the users will connect, if their similarity is above a threshold. We introduce three different similarity measures and corresponding prediction techniques.

- **Bio Similarity:** Given two users u and v , we compute the similarity $S_b(u, v)$ between the bio vectors of the two users.

- **Following Bio Similarity:** Given two users u and v , we compute the similarity $S_f(u, v)$ between the f-bio vectors of the two users.

- **Neighborhood Similarity:** Given two users u and v , we compute the similarity of their neighborhoods $N(u)$ and $N(v)$. This is a commonly used technique for link prediction [7]. The similarity measure used is the Jaccard similarity of the sets $N(u)$ and $N(v)$. Therefore, we define the neighborhood similarity $S_n(u, v)$ as follows

$$S_n(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$$

We evaluate our approach on the set of 1,126 users that have at least 10 reciprocal connections. This dataset gives us a manageable, yet large enough number of users to study, for whom we can compute the f-bio, and there is sufficient density of links between them to make the task interesting. There are 4,509 follow relationships between these users, and 85% of those are reciprocal.

We use the following methodology to evaluate the different approaches. Given the set of users U , we consider all possible pairs of users of U and compute the three similarity measures we described above. For a given similarity measure S , and a threshold value θ , let $T_S(\theta)$ denote the set of pairs that have similarity at least θ . If C is the set of connected pairs, we define the precision $P_S(\theta)$ and the recall $R_S(\theta)$ as follows.

$$P_S(\theta) = \frac{|T_S(\theta) \cap C|}{|T_S(\theta)|}, \quad R_S(\theta) = \frac{|T_S(\theta) \cap C|}{|C|}$$

We consider different values of recall and compute for each of the three similarity techniques the threshold that results in this recall. Using this threshold, we compute the precision of each technique and produce a precision-recall curve. Figure 2 shows the precision recall curve for the follow prediction task, Figure 3 for the follow-back prediction task, and Figure 4 for the mutual follow prediction task.

From the plots, the first observation is that the neighborhood similarity performs the best. This is to some extent expected, since it is established that connected users tend to follow similar accounts, and also the Twitter recommendation engine biases the follow links this way. Between the two bio-based methods, the f-bio similarity performs better than bio-similarity, confirming that the bios of one’s followings give a better description of the user than her own bio. The performance of the f-bio similarity also indicates that there is useful signal in the bio description of the users, and it can be used towards predicting connections (possibly in conjunction with the neighborhood information). Link prediction is a notoriously hard task, so any information that offers additional signal is important.

Of the three different tasks, predicting follow-back links is the easiest, all algorithms achieve high precision for high recall values. However, we should take into account that 85% of our links are reciprocal, so this task is significantly easier than the other two. Still we manage to get a non-negligible increase in the prediction accuracy for higher thresholds, using either the f-bio, or the neighborhood similarity. Predicting follow relationship is slightly easier than predicting mutual follow relationship; we achieve higher precision for all recall values. This is expected since it is a less demanding task.

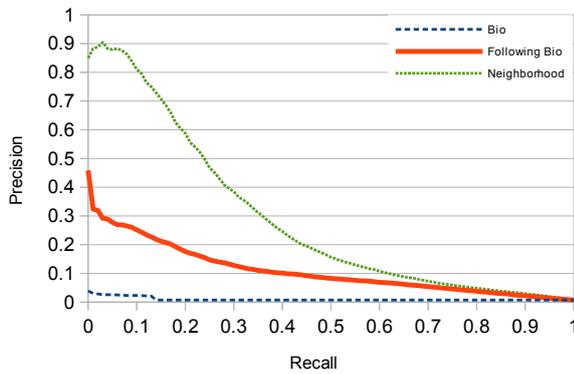


Figure 2: Follower prediction

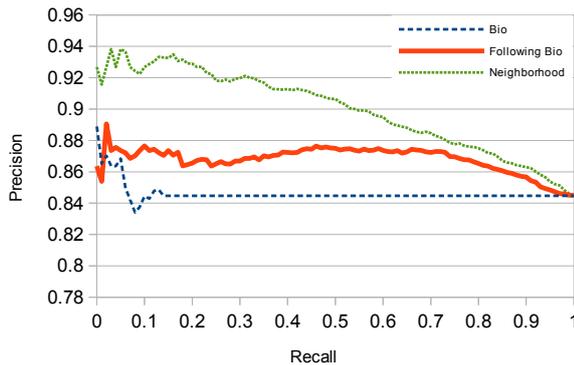


Figure 3: Follow-back prediction

5. CONCLUSION

In this work we studied the user bio descriptions on Twitter. To the best of our knowledge, this is the first such detailed analysis for bios. We analyzed the frequent keywords, and characterized them according to the purpose that they (usually) serve. We then studied whether connected users tend to share common keywords in their description and we observed that this is indeed the case, establishing homophily with respect to descriptions. This phenomenon is more pronounced when a description for a user is constructed from the bios of the users that she follows (the f-bio). Given these observations, we use similarity to predict connections between Twitter users. We observe again that the f-bio works better for this purpose. The followings of a user provide more information for her than her own description.

For future work, it would be interesting to study how the signal from the bio can be combined with other signals for different mining tasks. For example, it would be interesting to combine all the similarity measures into a single link prediction algorithm. Bios could also be useful for understanding or predicting other processes on Twitter, such as when a tweet will be retweeted, or whether a user will express interest in a new hashtag.

6. ACKNOWLEDGMENTS

Work supported by project Intersocial, through the European Territorial cooperation programme "Greece-Italy 2007-2013" co-funded by the European Union (ERDF) and National Funds of Greece and Italy.

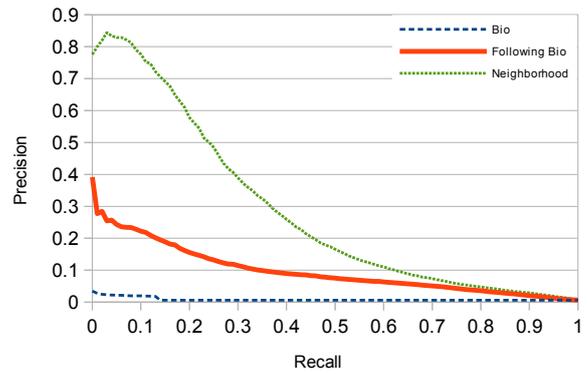


Figure 4: Mutual-following prediction

7. REFERENCES

- [1] M. J. Brzozowski and D. M. Romero. Who should i follow? recommending people in directed social networks. In *ICWSM*, 2011.
- [2] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *ICWSM*, 2010.
- [3] J. Cheng, D. M. Romero, B. Meeder, and J. M. Kleinberg. Predicting reciprocity in social networks. In *SocialCom/PASSAT*, pages 49–56, 2011.
- [4] G. De Francisci Morales, A. Gionis, and C. Lucchese. From chatter to headlines: harnessing the real-time web for personalized news recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining, WSDM '12*, pages 153–162, New York, NY, USA, 2012. ACM.
- [5] A. Java, X. Song, T. Finin, and B. L. Tseng. Why we twitter: An analysis of a microblogging community. In *WebKDD/SNA-KDD*, pages 118–138, 2007.
- [6] H. Kwak, C. Lee, H. Park, and S. B. Moon. What is twitter, a social network or a news media? In *WWW*, pages 591–600, 2010.
- [7] D. Liben-Nowell and J. M. Kleinberg. The link-prediction problem for social networks. *JASIST*, 58(7):1019–1031, 2007.
- [8] L. Lu and T. Zhou. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 396(6), 2011.
- [9] M. Mathioudakis and N. Koudas. Efficient identification of starters and followers in social media. In *EDBT*, pages 708–719, 2009.
- [10] M. Mathioudakis and N. Koudas. Twittermonitor: trend detection over the twitter stream. In *SIGMOD Conference*, pages 1155–1158, 2010.
- [11] M. R. Morris, S. Counts, A. Roseway, A. Hoff, and J. Schwarz. Tweeting is believing?: understanding microblog credibility perceptions. In *CSCW*, pages 441–450, 2012.
- [12] D. M. Romero and J. M. Kleinberg. The directed closure process in hybrid social-information networks, with an analysis of link formation on twitter. In *ICWSM*, 2010.
- [13] T. Sakaki, M. Okazaki, and Y. Matsuo. Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE Trans. Knowl. Data Eng.*, 25(4):919–931, 2013.
- [14] C. Wagner, V. Liao, P. Pirolli, L. Nelson, and M. Strohmaier. It's not in their tweets: Modeling topical expertise of twitter users. In *SocialCom/PASSAT*, pages 91–100, 2012.