Weighted *Content Based* Methods for Recommending Connections in Online Social Networks

Ruth Garcia
Universitat Pompeu de Fabra
Barcelona,Spain
rutholimpia.garcia01@campus.upf.edu

Xavier Amatriain
Telefonica R+D
Barcelona, Spain
xar@tid.es

ABSTRACT

Online Social Networks currently have an important role in the life of millions of active internet users. Cases like Twitter are of special attention since a lot of connections are made between people who never met before and with no need of reciprocation. For this reason it is important to find new ways to provide recommendations that may be of interest for users. Should these recommendations focus on the popularity, on the activity, location, common friends or content? Should recommendations be influenced by egocentric or global network metrics? This research is the first phase of an in-depth study of a large dataset based on Twitter which aims to answer the previous questions. Despite many studies based on global rankings, the authors believe that recommendations should mostly be based on the preferences made by users in their own networks. This stage of the study focuses on the popularity and activity of links as indicators to predict connections.For this end, the authors compute a weight for each of these features, which varies for each user. Each pair tested is accepted if it satisfies a minimum total weight. Results show a slight but important improvement in performance when using two features instead of one, the results gives an insight that if more features are considered more improvements in predictions will be found. The results of this paper can and should be accompanied with more research.

Keywords

followee, follower, features, Activity, Popularity, weight, O.S.N

1. INTRODUCTION

Nowadays, Online Social Networks have an important role in the life of millions of active internet users and an even stronger growth and penetration is expected in the following years. According to Mccann [4], 62.5% of Active Internet Users in the world belonged to at least one O.S.N by March of 2009 compared to the 57% in the previous year. From the first group, 56.4% try to "Find New Friends" in these networks.

In contrast to networks such as Facebook, Orkut and Myspace where privacy, reciprocity and knowing your connections is important, suggesting friends in environments like Twitter goes beyond the "you may know this person" approach since many connections with strangers do not need reciprocation. Furthremore, the majority of the times profiles are public. In fact, KwaK et al.[3], researches concluded that only 22% of all connections on Twitter are reciprocal. By April of 2010, Twitter has 105 million registered users and 37% of active users use Twitter on their phone[1].

For this reason, recommendations in this type of environment could be more broad and complex. This challenge has attracted major attention from several researchers interested in finding efficient and better ways to connect people. For example, there are methods to rank people in twitter such as TunkRank [9] or TwitterRank [10] while others have focused in avoiding spammers by measuring popularity such as the one presented by Avello [2]. Scellato et al. [6] analyzed how the location affects the social structure and Sudheendra et al.[7] proposes algorithms to find the most efficient path between two users that are not connected, among others. In contrast to recommendations of products where global rankings strongly influence people's choices, recommending people in O.S.N is more abstract. A high ranking does not necessarily mean a better recommendation since people get connected for reasons that may be hard to predict. The authors attempt to identify the reasons of connectivity by measuring how five features influence people when deciding to connect with someone. The features are a) popularity, b) activity, c)location, d)friends in common and e)content of tweets. In this research, which is the first stage of a deeper study, the authors study if the two first features are good indicators to predict connections and if they work better together than alone.

In this regard, the main objectives of this paper are: 1) Propose a method to weight popularity and activity for each user 2)Base predictions of future followees according to these measured weights 3)Observe if the two features together perform better in predictions than alone.

This study does not aim to propose a "good" or "better" way to recommend people in Twitter but it rather tries to explore new possibilities on personalized recommendations. To the best of our knowledge this paper is the only one

[1]http://www.blogherald.com/2010/06/28/twitters-meteoric-rise-compared-to-facebook-infographic/twitter-statistics-infographic-911/.

proposing personalized recommendations in twitter based on *content based* weighted features. In the following sections we will explain the approach, the thresholds we used, the weighting algorithms for the two features, the calculation of predictions and the evaluation method. At the end of the paper, we will discuss the results and propose new challenges for future research.

2. APPROACH

In environments similar to Twitter, users have three types of connections: only followers, only followees and reciprocal connections. Followees are people the user choose to follow and from whom they receive tweets in their so called *public timeline*, followers are those users who have chosen to follow the user in question but their tweets will not appear in the user's timeline whereas reciprocal connections are users that are both followers and followees and therefore information is shared in both directions. Basing recommendation on users who are only followers is not a good idea since in the majority of cases users do not have control over who follows them. Spammers or heavy advertisers looking for reciprocations can become their followers.

For this reason, we have carefully studied the role of followees because we believe that the user's preferences are better reflected in them but we do not substract reciprocal connections as Avello did in [2]. It is true that sometimes users follow back others for "politeness" but with time the tendency is to un-follow those heavy advertisers or people that do not provide information of interest. Active users aiming to find interesting information build a network of followees where the majority of members represent people who the users *care* to keep.

Studies of Kwak et al. [3] and Avello[2] have found homopholy in Twitter network showing that users "engaged in a social activity seem to be associated more closely with ones who are similar to them along a certain dimension such as location, age, political view or organization affiliation, compared to ones who are dissimilar." This shows that despite the lack of reciprocity in Twitter there is a tendency for people to get together with people similar in one or more aspects. In this part of the study, we analyze how similar followees are in a specific network considering popularity and activity. *Popularity* is the ratio of followers and followees and *activity* is the number of tweets a user has posted since entering the O.S.N.

In this phase of the study , weights are calculated for these two aspects. In order to do that, we first determine if that specific feature (popularity and/or activity) is actually relevant for that user. A feature is considered relevant if it is present in a certain percentage of followees. A recommendation is accepted if its calculated weight is bigger or equal than a threshold. In Section 3.2 we propose one metric for computing these weights.In Section 3.3 we explain how to make predictions according to this algorithm. We assume that these weights are indicators of how important that specific feature is for the user. In the following sections Popularity and Activity are referred as features, characteristics, indicators and aspects interchangeably.

We also focus only in the user's egocentric network of followees [2] and therefore in this study we call the recommendations as *content based people recommendations*. Due to our personalized recommendation approach, we omit calculation of global networks, global rankings and the use of collaborative filtering information since we only take information from the users' own networks. The results in Section 5 show the influence of each one of the features analyzed as well as the combination of both in the performance of finding predictions.

[2]egocentric network is used as "all the edges leading into and out of single user" in [5]

3. DATASET

The dataset was taken from the Web site of Munmon de Choudhury [3]. For this part of the study, we focused on two of the three files provided: the user file and the social graph file. The information given is that Tweets were collected between 2006 and 2009, the user file contains details of 456,107 profiles and the social graph file has a sample of 2,476 users with 815,554 followees links. The details of the profiles in the socialgraph are listed in the user file. Due to the evaluation method adopted and explained in Section 4, we considered those users with a total number of followees bigger or equal to 10. This reduces the sample to 2,381 users and 815,188 followees links. We leave for future research the evaluation of the content of tweets, location and friends in common.

3.1 Thresholds

For the calculation of weights we take into account the following thresholds:

- α : Popularity Threshold (followers/followees).
- $\beta1$-$\beta2$: Activity Range (Range of a total number of posts).
- δ : Feature Acceptance (%).
- γ : Weight Acceptance (≤ 1).

The popularity threshold (α) is calculated by the division of followers/followees. This ratio is not an accurate indicator of the prestige of a user. A ratio of 2 does not necessarily mean more popularity than a ratio of 1.5 but we have preferred not to use other more complex methods (see Section 1) because we are not calculating levels of prestige for recommendations but rather weather the user cares or not about popularity. We say that if a user's ratio of followers/followees is bigger than or equal to α then that user cares about popularity and therefore recommendations should consider popularity. We assume that any popular person will have a ratio bigger than one. For this reason, we start by setting α=1 but this value can be changed if needed. In Section 5, we analyze the impact of different α starting values. After determining if a user cares or not about popularity, we recalculate α based on the most likely α value the user will prefer. Section 3.3 specifies how this is done. For more information regarding the different methods of calculating prestige refer to [2].

The activity (β) is the total number of tweets a user has posted since the moment of joining the O.S.N. We assume that if a users has more followees with activities within the range $\beta1$-$\beta2$ then that user cares about the level of activity when choosing to follow someone. We assume that values

[3]For more information refer to http://www.public.asu.edu/ mdechoud/datasets.html

lower than $\beta 1$ mean inactivity or a new user and values above $\beta 2$ are candidates for spammers or heavy advertisers.We have analyzed the total dataset and observed that the majority of users are within 200-29000 number of posts. Just like the case of popularity, after determining if the activity is relevant for the user we determine the $\beta 1$ value that best suits the user. Section 3.3 also specifies how this is done. We chose not to recalculate $\beta 2$ because it is very unlikely that a user can post more than 29000 posts during 2 years but $\beta 1$ can vary and therefore other values for $\beta 1$ are evaluated in Section 5.

The feature acceptance threshold (δ) is the minimum percentage allowed in the network of followees for a feature to be considered relevant. For example, if δ=0.3 and α=1 , the popularity is relevant if 30% or more of the followees in that network have a popularity ratio bigger or equal than 1. Likewise, the weight acceptance threshold (γ) is the minimum weight a recommendation should have in order to be accepted. For example, if (γ)=0.7 and the total weight of a recommendation, $TW(r)$, is 0.6, then that recommendation is rejected. As it will be seen in Section 3.2, weight will not be bigger than 1. This approach resembles the System of Symeonidis et al. [8] where items with ratings bigger than 3 stars (varies according to user) are considered for calculations, the rest is not.

3.2 Weighting features

The weights are calculated after determining the relevant features for a particular user. The weight of a relevant feature for a user, u, is determined by the number of followees satisfying α, *PF*, and/or within the range $\beta 1$-$\beta 2$, *BF*, divided by the total number of followees. In his study we only consider this division for the calculation of weights because it is based only in the *egocentric network of followees*, leaving more complex metrics for future research. If Popularity is relevant for user u then the weight is calculated as :

$$W_p = \frac{\sum\limits_{\forall f \in PF} f}{\sum f} \tag{1}$$

If Activity is relevant for user u, the weight is calculated as :

$$W_a = \frac{\sum\limits_{\forall f \in BF} f}{\sum f} \tag{2}$$

The total Weight of relevant features for u is denominated as $TW(u)$. If only popularity is relevant then $TW(u) = W_p$, if it is only activity then $TW(u) = W_a$, if it is both $TW(u) = W_a + W_p$.

3.3 Filtering Recommendations

After obtaining the W_p and/or W_a for u, we should filter the future followees candidates, we follow the following steps:

1. Recalculate α and/or $\beta 1$
2. Filter Recommendations according to α and/or $\beta 1$
3. Weight Recommendatios
4. Accept or reject Recommendations according to weight

Step 1 is an step forward that aims to obtain more personalized values for α and/or $\beta 1$. For user u, we calculate this value by finding the median in group PF and/or BF. we have analyzed several other methods such as the average and metrics with thresholds but the Master thesis of the first author of this paper shows that the median is a good indicator [1] for this value.

After new values for α and/or $\beta 1$ are obtained, step 2 says that the recommendations, r, should be filtered according to these thresholds. For each one of the relevant features for u, a recommendation, r in the testing set is considered if $popularity_r \geq \alpha$ or in its turn $\beta 1 \leq activity_r \leq \beta 1$.

Step 3 says to weight the recommendation. We assumed that features that are not relevant for u should not be evaluated in r. The total Weight of each relevant feature r is denominated as $TW(r)$. If only popularity is relevant for r then $TW(r) = W_p(u)$, if it is only activity then $TW(r) = W_a(u)$, if it is both $TW(r) = W_a(u) + W_p(u)$. In other words, we weight r according to the parameters of u since we should judge the importance of the feature according to u.

Step 4 says that a recommendation, r, should be in the accepted group, AR, if the total $TW(r)$ of that recommendation is bigger or equal than γ. In this way, we have that:

$$r \in AR, \text{ if } TW(r) \geq \gamma, \tag{3}$$

4. EVALUATION METHOD

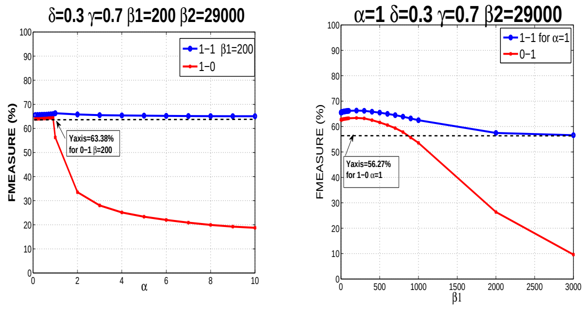
The evaluation method chosen was 10-fold cross-validation. The data has been divided randomly into 10 parts and one part is held out for testing 10 times. The training set is based on the remaining nine parts every time thus executing 10 times the learning procedure. According to Witten et al.[11], extensive tests on numerous datasets have shown that 10 is about the right number of folds to get the best estimates of errors. Nevertheless, we do not repeat the sampling 10 times as it is recommended in [11].

Since every testing set fold has actual connections <u,f> only, we have decided to randomly add fake connections to the testing set. Fake connections were carefully built from the 2,381 users mentioned in Section 3 that are not actually connected. We know that this could cause some unreal assumptions since some of these pairs could be actually connected by now but we consider this to be the most rigorous way of evaluating our algorithms. A recommendation is positive when we predict an existing followee and a recommendation is negative when the followee recommended is a fake connection.

The results show the level of precision, recall and the f-measure of predicting future followees for cases when popularity , activity and both the popularity and activity were considered. If a user u did not receive any recommendations, the recall was 0 and precision is not calculated. We also considered the recommendation percentage, which is the percentage of users in the testing set who received recommendations.

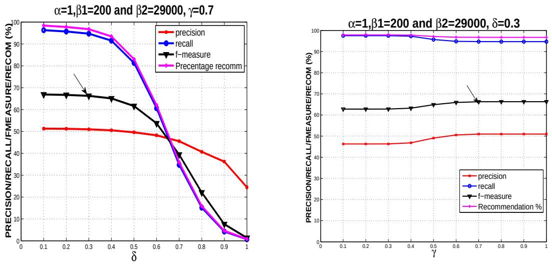
5. RESULTS

Labels with a) 1-0 represent cases where only feature popularity was evaluated, b) 0-1 only activity and c) 1-1 when both features were used. Each point in the x and y axis represent the average of the 10 learning procedures executed for each testing fold. The fake connections and the 10 different testing sets were the same every time.



(a) 1-0 and 1-1 (b) 0-1 and 1-1

Figure 1: FMeasure for a) 1-0, 1-1 with dashed line Yaxis of 0-1 at β1=200 and b) 0-1, 1-1 with dashed line Yaxis of 1-0 at α=1



(a) δ 1-1 (b) γ 1-1

Figure 2: Performance,recall,precision, recommendation percentage for 1-1 cases when a)δ and b) γ change

Improvements with two features instead of one are not very big but we believe they are important and promising. The dashed lines represent the minimum performances that the results of 1-1 cases should achieve. The graphs show that the more the feature being evaluated increases the more the results approach the dashed line.In other words, when two features are considered and a threshold reaches a point where its value is hard to achieve then the less it will be used in the algorithm, leaving the other feature to take the lead. Therefore the results in those situations start to behave as if there was only one feature present stable at the value marked by the dashed line. Figure 1a) shows an improvement of around 2% when α=1 while Figure 1b) have improvements of around4% when β1=200. Not in one moment cases considering only one feature perform better than cases with two features. For this reason, two features are better than one even if it is in a slight proportion. We expect more promising results when other features such as Content and common friends are added in future research.

On the other hand, Figure 2a and 2b show that the threshold γ does not makes a significant difference in the results. This means that we could automatically accept a recommendation if one or two of its features satisfies the new α and/or β1 rather than evaluating them again with γ. Nevertheless, we decided to keep γ because it may become significant when more features are added in future research and because it is a way to control how rigorous recommendations should be (i.e γ=1 when we want for force the algorithm to accept recommendations with the same amount of relevant features than u) .

6. DISCUSSION AND CONCLUSION

This paper proposes a way of recommending based on their followees. Non-reciprocal connections give space to innumerable ways of recommendations and challenges. In this study, two features were analyzed:popularity and activity. The authors have tried to identify and measure the impact of features influencing users to connect to others. A weighting algorithm based only on egocentric networks was proposed and discussed. For each user tested, the weighting algorithm takes into account the frequency of relevant features appearing in their followees profiles. A dataset with more than 800,000 connections was analyzed to test this weighting algorithm. Even with small improvements, we demonstrated that it is better to consider two features than only one when doing predictions because it help us to understand more the preferences of users when they choose connections. These preferences do not always go in hand with global rankings of prestige and activity hence we only analyzed the egocentric network. We demonstrated that weighting the impact of features in a personalized way could be an interesting , novel and personalized way of recommending connections. Although these experiments were carried out with Twitter, the algorithm and concepts can be applied in other environments. We look forward to continue with the analysis of other features such as location, content of tweets and connections in common. We are very enthusiastic about the potential improvements that considering more features could add to our results. We are also eager to compare our results with global ranking algorithms. In other future research, only reciprocal connections should be also studied.

7. REFERENCES

[1] R. Garcia. Hybrid methods for recommending connections in online social networks. Master Thesis, June 2010. UPF.

[2] D. Gayo-Avello. Nepotistic Relationships in Twitter and their Impact on Rank Prestige Algorithms. *Arxiv preprint arXiv:1004.0816*, 2010.

[3] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on WWW*, pages 591–600. ACM, 2010.

[4] U. McCann. UM Wave 4-Social Media Tracker: Power to the people. Universal McCann GmbH, Frankfurt am Main.

[5] M. Roth, A. Ben-David, D. Deutscher, G. Flysher, I. Horn, A. Leichtberg, N. Leiser, Y. Matias, and R. Merom. Suggesting friends using the implicit social graph. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 233–242. ACM, 2010.

[6] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora. Distance Matters: Geo-social Metrics for Online Social Networks.

[7] M. S. L. J. H. Sudheendra Hangal, Diana MacLean. All friends are not equal: Using weights in social graphs to improve search. In S. University, editor, *4th ACM Workshop on SONAM*, July 25, 2010.

[8] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos. Feature-Weighted User Model for Recommender Systems. *User Modeling 2007*, pages 97–106, 2009.

[9] D. Tunkelang. A Twitter Analog to PageRank (Authors' BlogOnline), January 2009.

[10] J. Weng, E. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 261–270. ACM, 2010.

[11] I. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Pub, 2005.