

Network-Aware Recommendations of Novel Tweets *

Noor Aldeen Alawad
Sapienza University of Rome
Rome, Italy
alawad@dis.uniroma1.it

Aris Anagnostopoulos
Sapienza University of Rome
Rome, Italy
aris@dis.uniroma1.it

Stefano Leonardi
Sapienza University of Rome
Rome, Italy
leonardi@dis.uniroma1.it

Ida Mele
Università della Svizzera Italiana
Lugano, Switzerland
ida.mele@usi.ch

Fabrizio Silvestri
Yahoo
London, UK
silvestr@yahoo-inc.com

ABSTRACT

With the rapid proliferation of microblogging services such as Twitter, a large number of tweets is published everyday often making users feel overwhelmed with information. Helping these users to discover potentially interesting tweets is an important task for such services. In this paper, we present a novel tweet-recommendation approach, which exploits network, content, and retweet analyses for making recommendations of tweets. The idea is to recommend tweets that are not visible to the user (i.e., they do not appear in the user timeline) because nobody in her social circles published or retweeted them. To do that, we create the user's ego-network up to depth two and apply the transitivity property of the *friends-of-friends* relationship to determine interesting recommendations, which are then ranked to best match the user's interests. Experimental results demonstrate that our approach improves the state-of-the-art technique.

Keywords

Tweet Recommendation, Content and Network Analysis.

1. INTRODUCTION

Online social networks allow people to connect to one another and to share information, opinions, and ideas. Twitter is a popular micro-blogging system where users can post short messages, called tweets, whose length is at maximum 140 characters. Tweets typically consist of personal information, status updates, news, or links to webpages or other web content (e.g., images and videos). Twitter users, by following one another, define a social graph, where nodes are users, and a direct edge (u, v) represents the fact that the user u follows the user v .

The tweets posted or retweeted by a user are shown on the user's profile page as well as on the timeline of her followers. These users can reply to the author of the tweet, or they can retweet the original

*This research was partially supported by the Google Focused Research Award "Algorithms for Large-Scale Data Analysis" and by the EU FET project MULTIPLEX 317532.

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGIR '16, July 17–21, 2016, Pisa, Italy.

© 2016 ACM. ISBN 978-1-4503-4069-4/16/07...\$15.00

DOI: <http://dx.doi.org/10.1145/2911451.2914760>

message, so that it is made visible to their followers, too. Retweets are very frequent, allowing the propagation of interesting information into the Twitter community, and, for this reason, the users follow news channels, favorite celebrities, or friends to obtain new information as soon as possible. However, the huge number of tweets posted everyday causes the problem of information overload, and recommendation systems can play a key role in solving this problem. In Twitter most of these systems are based on recommending users to follow [7], webpages to visit [17], or they consist in reranking the tweets appearing in the user's timeline [16]. A new line of research is based on recommending concealed tweets, by which we mean, tweets that are not posted or retweeted by anybody in the user's social circles. As a result, users have access to additional tweets that may be of interest. Consider, for example, the following scenario: A user is interested in computer science and everyday she checks the tweets of the computer scientists she is following. Some of these tweets are originally posted by them, whereas others are just the result of a retweet. When nobody has retweeted a message that is potentially interesting for the user, such message remains concealed from the user's eyes, because it does not appear in her timeline. Pennacchiotti et al. [10] tackled this problem by proposing tweets whose content matches the user's interests. We follow and extend their idea of recommending hidden tweets that are potentially interesting to the user, but instead of using only the content analysis, we propose to exploit the structure of the network around the user as well as the analysis of mutual retweets; we demonstrate that this yields an improvement in the quality of the recommendations.

2. RELATED WORK

Recommendation systems help the users to find products and services, music and videos as well as blogs and news articles, and they have been widely studied in the past [11]. With the rapid growth of Twitter, a lot of research has been focusing on analyzing Twitter data and the activities of its users for improving personalization and recommendations [8]. Depending on the type of recommendations, we can divide the recommendation tasks into three categories:

Followee recommendations: In [1], the authors proposed an approach which recommends the people to follow by analyzing the topology of the network around the user. Other studies have based their followee recommendations on the popularity and activity of Twitter users [3, 6]. Twittomender [7] is a famous system, which recommends followees based on the users' tweets and their relationships in the social graph. Although our system does not recommend people to follow but novel interesting tweets, there are some

similarities with these approaches, because we also exploit the network structure of the social graph.

Tweet-based content recommendations: Due to the lack of user-profiling information, recommending web content is often challenging, and Twitter has demonstrated to be a rich source from which is possible to obtain more information about web users. In particular, it has been proven that using Twitter data (e.g., tweets, retweets, hashtags of the user and of her friends) it is possible to improve the URL recommendations [4, 17]. Differently from these works, we do not recommend web content but focus on recommendations of tweets.

Tweet recommendations: Most of these approaches offer a ranking of tweets based on a query or they re-rank the tweets appearing in the user’s timeline. Yan et al. [16] rank tweets and their authors using hybrid networks, and Uysal and Croft [15] proposed to rank the incoming tweets based on their probability to be retweeted by the users. Our approach differs from these, because it aims at making recommendations of concealed tweets, which are tweets that do not appear in the user timeline, because they have not been posted or retweeted by anybody in the user’s social circles. Moreover, our approach is user-centric in the sense that the recommendations do not depend on a specific query but they rather match the user’s interests inferred from previous user’s (re)tweets. In this way, our recommendation algorithm is similar to [10] where the authors recommend unseen tweets based on the content similarity. Differently from them, we exploit not only the analysis of tweet content but also the structure of the network around the users (ego-network) and the users’ interest similarity which is mined from their common retweets. As we will explain in Section 3, we define a set of candidate tweets to recommend by creating the user’s ego-network with depth two and exploiting the transitivity property of the following–follower relationships among users. These candidate tweets are then ranked by using content-similarity measures plus the users’ interest similarity to best match the user’s interests.

3. METHODOLOGY

Our recommendation approach is based on recommending tweets that cannot be seen by the user, for example, because nobody in her circles wrote or retweeted them. Twitter users who subscribe to our system receive recommendations of novel tweets, which do not appear in their timeline but which are of potential interest.

For making recommendations we applied network, content, and retweet analyses. The idea is to use the network structure around the user as a pre-filtering step to find candidate tweets to recommend. Then, these tweets are ranked by using the content-similarity features and the number of common retweets, which is an indication of how much the users’ interests are similar.

To explore our method more deeply, when a user u subscribes to the service, the system uses the Twitter API to retrieve u ’s friends, friends of friends, and timeline (i.e., tweets and retweets). Given u ’s ego-network up to depth two, the recommendation algorithm exploits the transitivity property of the following–follower relationships. For example, assume that user u follows a set of users v_i who follow user z , and that u does not follow z , the (re)tweets of z are not visible to u , unless some v_i retweeted them. The idea is to use z ’s (re)tweets, which do not appear in u ’s timeline, as possible candidates for the recommendations. To weigh the importance of the tweets of z , we count the number of users v_i that are in between.

Network Analysis. The scenario described in the end of the previous paragraph can be mapped to the problem of finding open triangles in the ego-network of a user. Inspired by the MapReduce approach of Suri et al. [13] for counting triangles, we designed and implemented a MapReduce algorithm to find open triangles, so that

when user u follows v , and v follows z , the predicted link would be (u, z) , and recommendations would go from z to u as if u is actually one of z ’s followers. We also count the missing edges that close triangles so as to rank the nodes at distance two from the ego based on how many incoming links they have. Specifically, given the ego node u , let $\Gamma(u)$ be the set of u ’s friends (followees) and $\Gamma(\Gamma(u)) \setminus \Gamma(u)$ the set of u ’s friends of friends who are not friends of u . We define as the weight of user $z \in \Gamma(\Gamma(u)) \setminus \Gamma(u)$ to be the number of in-links $weight(z) = |\Gamma(u, z)|$ as described in Algorithm 1. Nodes are then ranked based on decreasing values of their weights.

Algorithm 1 MR-Counting-Open-Triangles(V, E)

```

1: /* Let  $G = (V, E)$  be the graph and  $(u, v) \in E$  the edge from  $u$  to  $v$ .
   Let  $ego$  be the id of the ego node and let  $\Gamma(v)$  be  $v$ ’s neighborhood */
2: Map 1: Input:  $key, (u, v); ego$ 
3:   if  $u = ego$  then
4:     emit  $key, (u, v)$  //  $key$  is the default input key
5: Reduce 1: Input:  $key, [(u, v1), (u, v2), \dots]$ 
6: // for each node at distance 2 from the ego, check cycles.
7:   for  $(u, v) \in values$  do
8:     for  $z \in \Gamma(v)$  do:
9:       if  $u \neq z$  then:
10:        emit  $(v, (u, z))$ 
11: Map 2: Input:  $v, (u, z)$ 
12: /* check if the edge  $(u, z)$  that closes a triangle is open. If so, then  $u$ 
   doesn’t follow  $z$  and cannot see  $z$ ’s tweets. */
13:   if  $(u, z) \notin E$  then:
14:     emit  $((u, z), 1)$ 
15: Reduce 2: Input:  $(u, z), [1, 1, \dots]$ 
16: /* sum the counts for each missing edge and emit a single key/value
   with the edge  $(u, z)$  and sum. */
17:    $sum \leftarrow 0$ 
18:   for  $i \in values$  do
19:      $sum = sum + i$ 
20:   emit  $(u, z), sum$ 

```

The first mapper identifies the set $\Gamma(ego)$. The first reducer identifies the nodes in $\Gamma(\Gamma(ego)) \setminus ego$. The second mapper identifies the set $\Gamma(\Gamma(ego)) \setminus (\Gamma(ego) \cup ego)$ and the final reducer counts for each node $z \in \Gamma(\Gamma(ego)) \setminus (\Gamma(ego) \cup ego)$ how many nodes v_i such that $ego \rightarrow v_i \rightarrow z$ exist.

Content Analysis. Tweets that best match the user’s interests can be discovered by applying the content similarity between a candidate tweet and the ego’s (re)tweets as well as the similarity between ego’s and candidate users’ timelines.

As content-similarity measures we used the *cosine similarity* and *Jaccard distance*, which are typically employed for making recommendations of tweets [2], news articles [12], and research papers [9]. Before applying the similarity measures, tweets were preprocessed to eliminate non-English content, stopwords, punctuations, and so on. We computed the similarity of tweets by using single terms and bi-grams. Moreover, one type of similarity is computed between the timelines of u and of $z \in \Gamma(\Gamma(u))$ and another similarity is between the candidate tweet originated from z and u ’s (re)tweets. This gives us an overall number of six content-similarity features: both cosine similarity and Jaccard distance for the tweets, the timelines, and the timelines using bigrams.

Retweet Analysis. We include another feature based on the number of common retweets, which provides an indication of the similarity between the interests of the user u and her neighbors at

distance two. Indeed, we noticed that users share on average 15 retweets, so we can use the number of mutual retweets to infer how close the users' interests are [5].

Ranking of Recommendations. We obtained the ranked list of tweets to recommend by using the pairwise comparison [14]. In more detail, we create a matrix of tweets and features (we assume that all the features have the same importance). We consider all pairwise combinations of candidate tweets and we compare them with respect to the features. A tweet beats another one if it has a better value for more features. At the end each tweet has a number of wins against the rest of the tweets, which induces a ranking among all the tweets. Tweets with higher number of wins are more likely to be relevant to the user's interests and are shown on the top of the ranked list of recommendations.

4. EXPERIMENTAL RESULTS

In this section, we present the experiment that we performed to validate our methodology and its results. As our approach does not recommend (re)tweets that are visible to the user, because it aims at recommending concealed tweets, we could not use retweets to assess if the recommendations are interesting or not. Therefore, we conducted a user study, where we proposed to real Twitter users our recommended tweets and collected their feedback.

4.1 User Study Evaluation

In the user study we involved 42 active Twitter users. The users could participate to our experiment by registering to our system using their screen name and email. Optional information, such as real name, gender, age range, Twitter and Internet usage, could be also provided. Participants of our user study were recruited from the list of the authors' followers. Most of them are young (between 20 and 30 years old) researchers, coming from different countries, affiliations, and research areas. The selected users are very active online (all of them use Internet everyday and 81% access Twitter on a daily basis). After the registration, the system retrieved the user's information needed to make the recommendations, and once these were ready, the user was notified by an email and could rate the list of recommended tweets. A total of 420 tweets from Oct. 7 to Nov. 17, 2015 were rated by our users. We compare the precision of our recommendations against a baseline approach for recommending concealed tweets. Since our aim is to recommend not visible tweets, we cannot compare our performance against algorithms that re-rank tweets appearing in the user's timeline. As a baseline, we used the approach presented by Pennacchiotti et al. [10], which exploits the content similarity among tweets, and, to the best of our knowledge, is the only work on recommendation of unseen tweets. We adopted their approach based on the cosine similarity, and for the evaluation we proposed the top-5 recommended tweets from our approach and the top-5 recommendations from the baseline. The two rankings were presented to the users, in a way that the user could not identify what system was used for creating the corresponding ranking. Following the same experiment of [10], our users could rate the proposed recommendations using a four-grade scale: *Excellent* (the tweet is very interesting/informative w.r.t. her interests), *Good* (the tweet is interesting/informative w.r.t. her interests), *Fair* (the tweet is somehow interesting/informative w.r.t. her interests), *Bad* (the tweet is not interesting/informative at all).

4.2 Assessing the Performance of the Recommendation System

For assessing the performance of our recommendation algorithm, we computed the following measures:

Precision@k (p@k): It is the percentage of relevant tweets found

in the top-k ranked tweets. To apply this measure, we casted the four-grade scale to a binary score (1 for interesting and 0 for uninteresting). In particular user's answers *Excellent*, *Good*, and *Fair* correspond to 1 and *Bad* to 0. We obtain similar results if we map *Excellent* and *Good* to 1 and *Fair* and *Bad* to 0.

Normalized Discounted Cumulative Gain (nDCG): It measures the performance of a recommendation system based on the graded relevance of the recommendations. We used this metric to assess the effectiveness of our recommendation algorithm using the actual nonbinary rates. It compares the ranking of tweets based on the recommendation scores and the ranking based on the user grades.

We calculated the average of these metrics over all the users who subscribed to the system. In Table 1 we show a comparison of our approach and the baseline for the average precision and nDCG of the top-k tweets (with $k = 1, \dots, 5$). Compared to the baseline we could observe an average improvement of 12.4% for Precision@k and of 1.6% for nDCG. We run t-test and could observe that the results were statistically significant at $p < 0.1$ for $p@1$ and at $p < 0.05$ for the other precisions.

Table 1: Comparison between our approach and the baseline.

		Precision				
		p@1	p@2	p@3	p@4	p@5
Our Approach		0.85	0.82	0.80	0.79	0.80
Baseline		0.73	0.67	0.67	0.68	0.69
		nDCG				
		@1	@2	@3	@4	@5
Our Approach		0.75	0.79	0.84	0.88	0.92
Baseline		0.74	0.77	0.82	0.86	0.91

We also computed the Reciprocal Rank (RR), which is the inverse of the ranking position of the first relevant tweet (i.e., the tweet which has been rated as *Excellent*, *Good*, or *Fair* by the user). The average RR over all users is of 91% for our approach and 83% for the baseline.

We report in Figure 1 the summary statistics with respect to the Likert scale judgements for both methods. We can observe that there is a larger number of tweets rated as *Excellent* and *Good*, while the number of tweets rated as *Bad* is lower compared to the baseline.

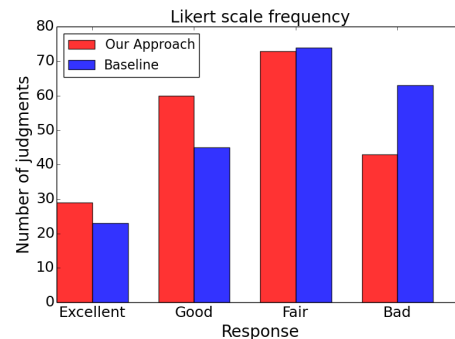


Figure 1: Comparison between our approach (red) and the baseline (blue) for the Likert scale.

Finally, in Figure 2 there are the percentages of tweets rated as *Excellent*, *Good*, *Fair*, and *Bad*, respectively. From these figures, it is clear that our method outperforms the baseline with higher percentage value.

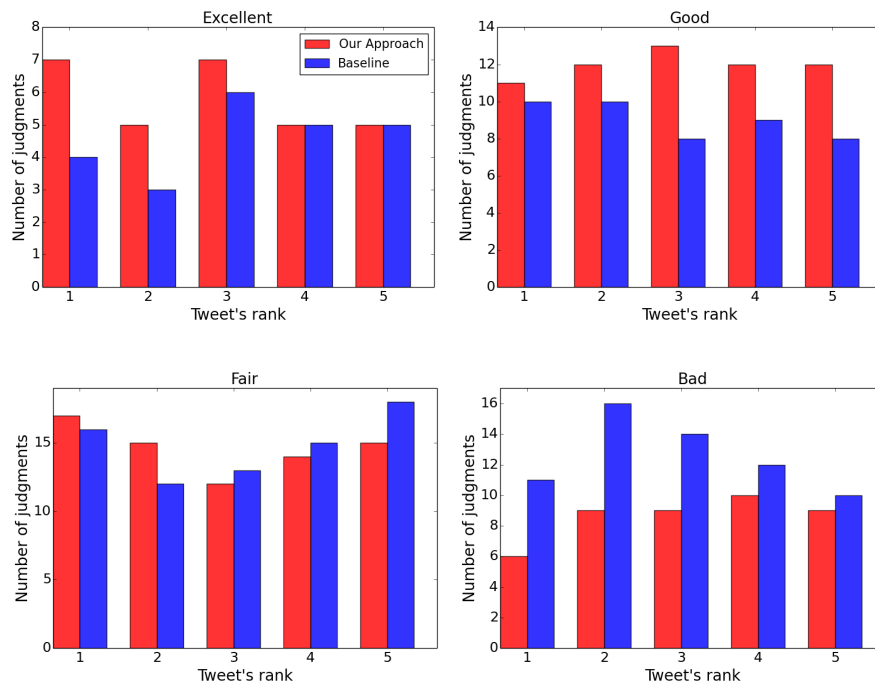


Figure 2: Comparison between our approach (red) and the baseline (blue) for each value of the Likert scale.

5. CONCLUSIONS

In this paper we have presented a novel approach for recommending concealed tweets. Our methodology uses the structure of a network with two-hop distance from the user, the content similarity, and the analysis of shared retweets in order to propose interesting tweets that otherwise would remain hidden from the user.

Experimental results corroborated that our recommendation system overcomes the existing approach for recommending unseen tweets with an improvement of 12.4% in the precision.

For future work, we would like to extend our study in order to consider also the feedback from the users. Finally, here we considered only a particular graph property, but we plan to try other approaches to find potential tweets (e.g., from the followers) and for weighting these tweets.

Acknowledgement

We would like to thank the authors of [10] for having provided the code of their recommendation algorithm.

6. REFERENCES

- [1] M. G. Armentano, D. L. Godoy, and A. A. Amandi. A topology-based approach for followees recommendation in Twitter. In *ITWP'11*.
- [2] S. Benzarti and R. Faiz. EgoTR: Personalized tweets recommendation approach. In *CSOC'15*.
- [3] H. Chen, X. Cui, and H. Jin. Top-k followee recommendation over microblogging systems by exploiting diverse information sources. *Future Generation Computer Systems*, 55, 2016.
- [4] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi. Short and tweet: experiments on recommending content from information streams. In *CHI'10*.
- [5] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu. Collaborative personalized tweet recommendation. In *SIGIR'12*.
- [6] R. Garcia and X. Amatriain. Weighted content based methods for recommending connections in online social networks. In *RSWeb'10*.
- [7] J. Hannon, K. McCarthy, and B. Smyth. Finding useful users on Twitter: Twittomender the followee recommender. In *ECIR'11*.
- [8] S. M. Kywe, E.-P. Lim, and F. Zhu. A survey of recommender systems in Twitter. In *SocInfo'12*.
- [9] A. B. Patel, N. B. Suthar, and J. S. Dhobi. Recommending top-n research papers (based on with, without and boolean items preferences: An user base collaborative filtering approach in mahout). In *ICCCIT'12*.
- [10] M. Pennacchiotti, F. Silvestri, H. Vahabi, and R. Venturini. Making your interests follow you on Twitter. In *CIKM'12*.
- [11] P. Resnick and H. R. Varian. Recommender systems. *Commun. ACM*, 40(3), 1997.
- [12] A. Sotsenko, M. Jansen, and M. Milrad. Using a rich context model for a news recommender system for mobile users. In *NRA'14*.
- [13] S. Suri and S. Vassilvitskii. Counting triangles and the curse of the last reducer. In *WWW'11*.
- [14] P. Tannenbaum and R. Arnold. *Excursions in modern mathematics*. Prentice Hall, 1995.
- [15] I. Uysal and W. B. Croft. User oriented tweet ranking: a filtering approach to microblogs. In *CIKM'11*.
- [16] R. Yan, M. Lapata, and X. Li. Tweet recommendation with graph co-ranking. In *ACL'12*.
- [17] N. Yazdanfar and A. Thomo. Link recommender: Collaborative-filtering for recommending urls to Twitter users. *ANT'13*.