Active Microbloggers: Identifying Influencers, Leaders and Discussers in Microblogging Networks

Conference Paper · October 2012

DOI: 10.1007/978-3-642-34109-0_12

Lamjed Ben Jabeur
Institut de Recherche en Informatique de Toulouse

Active Microbloggers: Identifying Influencers, Leaders and Discussers in Microblogging Networks

Lamjed Ben Jabeur, Lynda Tamine, and Mohand Boughanem

IRIT, University of Paul Sabatier,
118 Route de Narbonne, F-31062 TOULOUSE CEDEX 9, France
{jabeur,tamine,boughanem}@irit.fr

Abstract. This paper presents a social approach for identifying key actors in microblogging social network. In particular, we propose three specific link analysis algorithms called *InfRank*, *LeadRank* and *DiscussRank* that identify influencers, leaders and discussers, respectively. Conducted experiments on TREC 2011 Microblog dataset, show that the proposed algorithms outperform close microblogger ranking approaches.

1 Introduction

Microblogging services are characterized with an intensive activity. In the case of Twitter[1] which we mainly interest in, 340 million[2] tweets are published every day. Consequently, ranking microblogs by chronological order is no longer appropriate to access to the huge amount of information produced daily. In order to tackle this problem and access to interesting microblogs, tweets are ranked according to the importance of corresponding microbloggers in the social network.

In this context, popularity, authority and influence [1,2,3] has been addressed as the basic properties of important microbloggers. Meanwhile, new types of microbloggers has emerged with the promotion of collaboration and opinion sharing over microblogs. These microbloggers are called active and correspond to influencers as well as leaders and discussers. Influencers are network actors who are able to largely spread an information through the network. Leaders have the ability to motivate people and stimulate a community movement. Finally, discussers initiate valuable discussions around an interesting topic.

In this paper, we are interested in identifying active microbloggers and we propose a social network model that represents microbloggers using the social interactions between them such as following, retweeting and mentioning relationships. Moreover, we propose three different link analysis algorithms (*InfRank*, *LeadRank* and *DiscussRank*) that highlight key microbloggers in the network.

The remainder of this paper is organized as follows. Section 2 gives an overview of related work. Section 3 introduces the social network model for microbloggers. Section 4 details the proposed algorithms for identifying influencers, leaders and

[1] http://www.twitter.com/
[2] http://blog.twitter.com/2012/03/twitter-turns-six.html

L. Calderón-Benavides et al. (Eds.): SPIRE 2012, LNCS 7608, pp. 111–117, 2012.

discussers. Section 5 presents experimental results. Finally, section 6 concludes the paper.

2 Related Work

Previous approaches for ranking microbloggers have addressed three basic properties including popularity, authority and influence. Popularity is estimated based on either the number of published microblogs such as *TweetRank* metric or the number of followers such as *FollowRank* metric [1]. Authority in microblogs is approximated by applying *PageRank* algorithm on the followers network [2]. In the same aim, Pal et al. [3] propose to apply a probabilistic clustering method that uses a set of features from microblogging practices in order to identify topical authorities in the network. Influence in microblogs is addressed as a composite property that includes indegree (follower) influence, retweet influence and mention influence [4]. Influence is also assimilated to the topical authority of a microblogger in the social network. *TwitterRank* [5] algorithm is proposed in this context to identify topical authorities in the followers network.

Besides the previous properties, leaders and discussers have not been yet investigated as key actors in microblogging networks. Meanwhile, some research on traditional blogs has addressed similar properties. In order to discover leaders, some work has focused on propagation patterns in the social network [6]. Discussers have been addressed in [7] as network agitators who simulate discussion in blog threads.

We introduce in this paper a social model for identifying active microbloggers. Our approach is different in at least two respects from previous related work:

– We model the social network of microbloggers using a weighted multigraph that integrates followerships, retweets and mentions in the contrast of previous approaches using one or more binary social graphs [2,5].
– We investigate influencers, leaders and discussers as key microbloggers unlike previous works focusing on popularity, authority and influence [1,2,3,4,5].

3 The Social Network of Microbloggers

We propose to represent the social network of microbloggers using multigraphs. Unlike simple graphs, multigraph based representation allows multiple edges between nodes. Microbloggers can be simultaneously connected with several types of relationships including following, retweeting and mentioning associations.

3.1 Network Topology

The social network of microbloggers is represented using a directed, labeled and weighted multigraph $G := (U, E, \Sigma_E, \ell_E, \mathcal{O}, \mathcal{I})$ with:

- U is the set of microblogger nodes,
- $E = U \times U$ is the set of edges denoting relationships between microbloggers,
- $\Sigma_E = \{f, r, m\}$ is the alphabet of edge labels with f, r and m correspond respectively to following, retweeting and mentioning associations,
- $\ell_E : E \rightarrow \Sigma_E$ associates to each edge a label,
- $\mathcal{O} : U \times \Sigma_E \rightarrow U \times U \times \cdots \times U$ associates to each microblogger $u_i \in U$ the set of successor nodes with connecting edges are labeled by $l \in \Sigma_E$,
- $\mathcal{I} : U \times \Sigma_E \rightarrow U \times U \times \cdots \times U$ associates to each microblogger $u_i \in U$ the set of predecessor nodes with connecting edges are labeled by $l \in \Sigma_E$.

3.2 Relationship Weights

Let $T(u_i)$, $R(u_i)$, $M(u_i)$ respectively be the set of tweets of microblogger u_i, the set of tweets he has retweeted and the set of tweets where he has been mentioned in. A weight is assigned to each network edge according to its type as follows:

Following relationship: A followership edge is defined from a microblogger u_i to a microblogger u_j if the first microblogger follows the second one. Reinforced followership links via intermediary nodes confirms the interest of the first user in the followed microblogger. This association is weighted as follows:

$$w_f(u_i, u_j) = \frac{|\{u_i\} \cup \mathcal{O}(u_i, f) \cap \mathcal{I}(u_j, f)|}{|\mathcal{O}(u_i, f)|} \tag{1}$$

Retweeting relationship: A retweet edge is defined from a microblogger u_i to a microblogger u_j if there exists at least one tweet of u_j retweeted by u_i. Retweeting association would be as much reliable as microblogger u_i publishes retweets that belong to microblogger u_j.

$$w_r(u_i, u_j) = \frac{|T(u_j) \cap R(u_i)|}{|T(u_i)|} \tag{2}$$

Mentioning relationship: A mentioning edge is defined from a microblogger u_i to a microblogger u_j if there exists at least one tweet of u_i mentioning u_j. A microblogger u_i would communicate as much information to a microblogger u_j as he mentions him in his tweets.

$$w_m(u_i, u_j) = \frac{|T(u_i) \cap M(u_j)|}{|T(u_i) \cap \bigcup\limits_{u_k \in \mathcal{O}(u_i, m)} M(u_k)|} \tag{3}$$

4 Identifying Active Microbloggers

Based on the above social network model we propose in this section specific link analysis algorithms that identify each type of active microbloggers.

4.1 Influencers

Influencers have the ability to spread information over the microblogging network. They gain retweets for every published tweet. Highly followed microbloggers have more opportunity to be retweeted. Accordingly, we attribute to each microblogger an initial influence score based on his popularity $\mathcal{P}(u_i) = |\mathcal{I}(u_i, f)|$. The influence of a microblogger is, however, affirmed if he involves other good influencers in retweets. To rank microbloggers by their mutual influence, we propose a *PageRank*-like algorithm, called *InfRank*, that considers the social network of microbloggers. In particular, *InfRank* propagates influence score through retweet edges in respect to weights defined in formula 2. This process is repeated until convergence along with normalizing influence scores at each iteration.

Algorithm 1. InfRank

$k \longleftarrow 0$

foreach $u_i \in U$ **do** $Inf^k(u_i) = \frac{\mathcal{P}(u_i)}{|U|}$ *// initialization*

repeat

  $k \longleftarrow k + 1$

  foreach $u_i \in U$ **do**

    $Inf^k(u_i) = (1-d)\frac{\mathcal{P}(u_i)}{|U|} + d \times \sum_{u_j \in \mathcal{I}(u_i, r)} w_r(u_j, u_i)\frac{Inf^{k-1}(u_j)}{\mathcal{O}(u_j, r)}$

  end

  foreach $u_i \in U$ **do** $Inf^k(u_i) = \frac{Inf^k(u_i)}{\sum_{\forall u_j \in U} Inf^k(u_j)}$ *// normalization*

until *convergence* *// microblogger ranks never change*

4.2 Leaders

Leaders are advanced influencers who are able to create movements in the social network by stimulating replies and mentions. The number of followers, retweets and replies reflects the size their community, and thus the chance to be a leader. Consequently, an initial leadership score is attributed to a microblogger based on his social network attraction $\mathcal{A}(u_i) = |\mathcal{I}(u_i, f) \cup \mathcal{I}(u_i, r) \cup \mathcal{I}(u_i, m)|$. The leadership of a microblogger is enhanced if he is retweeted and mentioned by other leaders. Based on this idea, we propose a link analysis algorithm, called *LeadRank*, that highlights leaders in the social network. *LeadRank* propagates the leadership score through incoming retweet and mentioning associations respectively to weights defined in formulas 2 and 3. At each iteration, a new normalized leadership score is computed based on the product of both weighted sum of retweet predecessors scores and weighted sum of mention predecessors scores. We notice that this product emphasizes the two properties of leaders, namely influence (retweets) and ensured community motivation (mentions).

Algorithm 2. LeadRank

$k \longleftarrow 0$
foreach $u_i \in U$ **do** $Ldr^k(u_i) = \frac{\mathcal{A}(u_i)}{|U|}$ // *initialization*
repeat
  $k \longleftarrow k + 1$
  foreach $u_i \in U$ **do**
    $Ldr^k(u_i) = (1-d)\frac{\mathcal{A}(u_i)}{|U|} + d \times$
    $\left[\sum_{u_j \in \mathcal{I}(u_i,r)} w_r(u_j,u_i)\frac{Ldr^{k-1}(u_j)}{\mathcal{O}(u_j,r)} \times \sum_{u_j \in \mathcal{I}(u_i,m)} w_m(u_j,u_i)\frac{Ldr^{k-1}(u_j)}{\mathcal{O}(u_j,m)}\right]$
  end
  foreach $u_i \in U$ **do** $Ldr^k(u_i) = \frac{Ldr^k(u_i)}{\sum_{\forall u_j \in U} Ldr^k(u_j)}$ // *normalization*
until *convergence* // *microblogger ranks never change*

4.3 Discussers

Discussers do not influence the social network but discover interesting tweets and initiate valuable conversations around them. They are involved in conversations with different microbloggers and therefore receiving many replies. To identify network discussers, we consider mentioning relationships. Microbloggers already in interaction with many interlocutors are good candidates to be good discussers. Consequently, an initial discusser score is assigned to each microblogger based on his conversational activities $\mathcal{C}(u_i) = |\mathcal{I}(u_i,m) \cup \mathcal{O}(u_i,m)|$. Discusser importance increases if the microblogger is mentioning and being mentioned by good discussers. To highlight important discussers, we propose a link analysis algorithm, called *DiscussRank*. This algorithm is described next.

Algorithm 3. DiscussRank

$k \longleftarrow 0$
foreach $u_i \in U$ **do** $Desc^k(u_i) = \frac{\mathcal{C}(u_i)}{|U|}$ // *initialization*
repeat
  $k \longleftarrow k + 1$
  foreach $u_i \in U$ **do**
    $Desc^k(u_i) = (1-d)\frac{\mathcal{C}(u_i)}{|U|} + d \times$
    $\left[\sum_{u_j \in \mathcal{I}(u_i,m)} w_m(u_j,u_i)\frac{Desc^{k-1}(u_j)}{\mathcal{O}(u_j,m)} \times \sum_{u_j \in \mathcal{O}(u_i,m)} w_m(u_i,u_j)\frac{Desc^{k-1}(u_j)}{\mathcal{I}(u_j,m)}\right]$
  end
  foreach $u_i \in U$ **do** $Desc^k(u_i) = \frac{Desc^k(u_i)}{\sum_{\forall u_j \in U} Desc^k(u_j)}$ // *normalization*
until *convergence* // *microblogger ranks never change*

5 Experimental Evaluation

5.1 Experimental Setup

Topic Dataset. We used in these experiments the TREC 2011 Microblog dataset that includes about 16 million tweets. To identify active microbloggers for a particular topic, we selected 3 topics from main trends in the corpus period.

Table 1. Topic dataset statistics

#	Topic	Tweets	Microbloggers	Following	Retweets	Mentions
1	NFL Super Bowl	55,225	52,082	41,695	951	23,674
2	Egypt's Tahrir Square protests	53,047	36,571	154,628	27,712	12,976
3	State of the Union address	21,986	20,068	15,673	541	221
Mean		43,419	36,240	70,665	9,735	12,290

Baselines. We compare our algorithms to 3 baselines. *Followers* baseline ranks microbloggers by followers number. *F-PageRank* and *R-PageRank* compute a PageRank score on the social networks of followers and retweets, respectively.

Results Assessment. Inspired by the evaluation protocol in [3], we asked 2 regular Twitter users to rate the interestingness of the top 20 microbloggers. In order to study the impact of tweet content and the social profile of microblogger, this evaluation is conducted in 2 steps: *Anonymous evaluation AI* and *Non-anonymous evaluation ¬AI*.

5.2 Active Microblogger Precision

Table 6 presents Normalized Discounted Cumulative Gain (NDCG) values of different compared baselines and algorithms. This measure evaluates the retrieval effectiveness based on the position of interesting microbloggers in the result set. First, we notice that AI rating presents higher NDCG@10 and NDCG@20 values than ¬AI rating. Knowing the social context, annotators can make a final decision about microbloggers who discuss the topic but not socially interesting such as automatic tweets generators. In the case of AI evaluation, *R-PageRank* presents highest NDCG@10 and NDCG@20 values. *InfRank* algorithm outperforms the *LeadRank* algorithm. *Followers*, *F-PageRank* and *DiscussRank* are, however, less effective. Considering ¬AI evaluation which is the more significant one in these experiments, we note that *InfRank* and *LeadRank* algorithms

Table 2. Comparison of baseline effectiveness for AI and ¬AI evaluations

	AI		¬AI	
	NDCG@10	NDCG@20	NDCG@10	NDCG@20
Followers	0.248	0.262	0.198	0.230
F-PageRank	0.263	0.294	0.294	0.328
R-PageRank	**0.386**	**0.406**	0.301	0.378
InfRank	0.297	0.340	**0.345**	**0.398**
LeadRank	0.381	0.385	0.337	0.393
DiscussRank	0.155	0.191	0.175	0.244

present higher values. We conclude that influence is a primordial property of active microbloggers. Similar performances are shown by *R-PageRank* algorithm which investigates also microblogger influence with different interpretations.

6 Conclusion and Future Work

We proposed in this paper a weighted social network model for microbloggers. Furthermore, we proposed specific link-analysis algorithms that identify influencers, leaders and dicussers in the network. Experiments show that InfRank and LeadRank algorithms overpass compared baselines. In future work, we plan to integrate the proposed algorithms into a real-time content discovering system that focuses on active microbloggers streams instead of full network stream.

References

1. Nagmoti, R., Teredesai, A., Cock, M.D.: Ranking approaches for microblog search. In: Web Intelligence 2010, pp. 153–157 (2010)
2. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: WWW 2010, pp. 591–600 (2010)
3. Pal, A., Counts, S.: Identifying topical authorities in microblogs. In: WSDM 2011, pp. 45–54 (2011)
4. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring user influence in twitter: The million follower fallacy. In: ICWSM 2010 (2010)
5. Weng, J., Lim, E.-P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: WSDM 2010, pp. 261–270 (2010)
6. Goyal, A., Bonchi, F., Lakshmanan, L.V.: Discovering leaders from community actions. In: CIKM 2008, pp. 499–508 (2008)
7. Nakajima, S., Tatemura, J., Hara, Y., Tanaka, K., Uemura, S.: Identifying Agitators as Important Blogger Based on Analyzing Blog Threads. In: Zhou, X., Li, J., Shen, H.T., Kitsuregawa, M., Zhang, Y. (eds.) APWeb 2006. LNCS, vol. 3841, pp. 285–296. Springer, Heidelberg (2006)