A case study on the influence of the user profile enrichment on buzz propagation in social media: Experiments on *Delicious*

Manel Mezghani[1, 2], Sirinya On-At[2], André Péninou[2], Marie Françoise Canut[2], Corinne Amel Zayani[1], IkramAmous[1]and Florence Sedes[2]

[1] Department of computer science, Sfax University, MIRACL laboratory, Sfax, Tunisia
[2] Toulouse Institute of Computer Science Research (IRIT), University of Toulouse, CNRS, INPT, UPS, UT1, UT2J, 31062 TOULOUSE Cedex 9
mezghani.manel@gmail.com
{corinne.zayani,ikram.amous}@isecs.rnu.tn
{sirinya-onat, andre.peninou,marie-francoise.canut, florence.sedes}@irit.fr

Abstract. The user is the main contributor for creating information in social media. In these media, users are influenced by the information shared through thenetwork. In a social context, there are so-called “buzz”, which is a technique to make noise around an event. This technique engenders that several users will be interested in this event at a time *t*. A buzz is then popular information in a specific time. A buzz may be a fact (true information) or a rumour (fake, false information). We are interested in studying buzz propagation through time in the social network *Delicious*. Also, we study the influence of enriched user profilesthat we proposed [2] to propagate the buzz in the same social network. In this paper, we state a case study on some information of the social network *Delicious*. This latter contains social annotations (tags) provided by users. These tags contribute to influence the other users to follow this information or to use it. This study relies on three main axes: 1) we focus on tags considered as buzz and analyse their propagation through time 2)we consider a user profile as the set of tags provided by him. We will use the result of our previous work on dynamic user profile enrichment in order to analyse the influence of this enrichment in the buzz propagation. 3)we analyse each enriched user profile in order to show if the enrichment approach anticipate the buzz propagation. So, we can see the interest of filtering the information in order to avoid potential rumours and then, to propose relevant results to the user (e.g. avoid "bad" recommendation).

1. Introduction

In social media users are influenced by their information shared through the network. In a social context, there is so-called buzz that is a technique to make noise around an event. A buzz is popular information in a specific time. This technique engenders that several users will be interested in this event at a time *t*. A buzz may be a fact (true information) or a rumour (fake, false information).
Based on the definition of [5], a rumour is defined as "*an unverified proposition for belief that bears topical relevance for persons actively involved in its dissemination*".

According to [4], a rumour is characterized by its rapidly spread. However, rumour detection is a crucial problem since it requires additional background knowledge to verify information/proposition.
In this paper, we propose to study a buzz that could be a potential rumour. We are interested in studying the propagation through time of the buzz in the social network *Delicious*[1](more precisely a dataset of *Delicious* [3]). Also, we are interested in studying the influence of dynamic enrichment of users profiles proposed in [2], to propagate the buzz through time.

The dynamic enrichment approach considers the temporal dynamics of the social network. In fact, the user profile enrichment is done according to each period of time. It is not an accumulation of previous enrichment in previous periods. This enrichment approach takes into consideration the popularity, the freshness of information (a tag) and the similarity of users annotating (tagging) the same resource in a specific period of time.

In this paper, we make a case study on some information of the social network *Delicious* that contains social annotations (tags) provided by users. These tags contribute to influence other users to follow this information or to use it.

This study relies on three main axes:

1. Wefocus on tags considered as buzz and analyse their propagation through time.
2. We consider a user profile as the set of tags provided by him. We will use the result of our previous work on temporal user profile enrichment in order to analyse the influence of this enrichment in the buzz propagation.
3. We analyse each enriched user profile in order to show if the enrichment approach anticipate the buzz propagation.

This paper is structured as follows. First, we give an overview of the dynamic enrichment approach. Second, we detail the dataset used, we study some cases of buzz propagation through time with and without the enrichment approach, and also we analyseif the enrichment approach anticipates the buzz propagation. Finally, we conclude and give some perspectives.

2. Overview of the dynamic enrichment approach

In this section, we give an overview of our approach for enriching users profiles already detailed in [2]. The dynamic evolution of the user profile is treated by enriching users' interests with tags deemed relevant for each period of time. In fact, in social environment, the user consults the resources stored in the network, communicates and interacts with other users to find the information he needs. Enrichment in this context is done by analysing the environment of the user to detect relevant interests (relevant tags).

The relevance of an interest is usually calculated from the frequency of use of the tag at a given time. Frequency periodically varies. This change has already been treated by [1], through the concept of "temperature". This notion is interesting since it models the popularity of a term over time.

---

[1]www.delicious.com

The user profile is constructed in an implicit way, using the list of tags assigned by the user. The user profile is enriched with tags (considered as his interests) in each period of time in order to reflect the current interests of the user.

The first step consists in dividing the database in each period of time.The choice of thisperiodenables us to detect the evolution of the user interests between two successive periods. This latter, should be consistent with the quantity of data presented in the social network. By dividing the database, we obtain temporal information of the user activity in each period like his neighbours, his tags and the tagged resources.

The second step consists in calculating the temperature of each resource in a given period. In order to calculate this attribute, we propose a formula that takes into consideration several parameters: the freshness of a tag associated to the resource, the similarity of the users who tagged the resource and the number of tags associated with the resource (popularity). The temperature of the resource varies through time. It may increase or decrease. We consider that a resource is interesting if its temperature increases.

The third step consists in detecting the resources where temperature increases over time. After calculating the temperature of each resource, we consider only the resourceswhere temperature value is increasing between two periods of times (this reflect the interest of the user with this resource). However, in social networks that are characterized by the amount of the resources, we can have a lot of resources where temperature is increasing and then their treatment can be complex. So, in order to overcome such a problem, we should keep only the most relevant resources to the user. That's why we analyse the content of the resources and more precisely their metadata (we consider that the resources are semi-structured data). In our work, we use the metadata as the descriptors of the content of the resource, in order to filter the most relevant tagged resources. We attribute a weight for the tags associated with the resources. This weight is calculated according to the degree of correspondence of the tags with the metadata of the associated resource.

The fourth step consists in enriching the user profile with the tags associated with the resources. After calculating the weight of the tags associated with the most interesting resources, we enrich in this step the user profile with tags that reflect the best the user interests. In fact, the more the tag has a higher weight, the more it reflects the content of the resource and then, the more it reflects the user interests. So, we choose from the result of the previous step, tags that are more interesting to the user. A tag is stated as a potential interest if its weight is higher than a giventhreshold.

As a result of this approach, we have an enriched profile in each period of time.

3. Case study on *Delicious* dataset of buzz propagation through time

In this section, we first present the dataset used in our experiment. Second, we analyse the evolution of the top-10 buzz (popular tags) through time. Third, we analyse the influence of the enrichment approach on the top-10 buzz propagation. Finally, we analyse if the user profile enrichment anticipates buzz propagation.

3.1 *Delicious* dataset

The *Delicious* dataset contains social networking, bookmarking, and tagging information. The temporal interval of activity of the dataset varies between November 2003 and October 2010.It provides information about the user's friend relationships and the tagging relation information <U, T, R>. The users U are described through their ID (e.g. userID=8). The resources R are described through their ID, title and URL (e.g. 1 IFLA - The official website of the International Federation of Library Associations and Institutions http://www.ifla.org/). The tags T are described through their ID and value (e.g. 1 collection development). This dataset is extracted from [3]. We present some statistics of the data present in this dataset: 1867 users, 69226 URLs and 53388 tags. Also, the tagging behaviour is provided according to the time information. This behaviour implies that we know a tag is used in a specific period of time. An example of temporal tagging behaviour is shown in table 1.

Table 1: An example of the temporal tagging behaviour

userID	**bookmarkID**	**tagID**	**day**	**month**	**year**	**hour**	**minute**	**second**
8	1	1	8	11	2010	23	29	22

3.2 Buzz evolution tracking

In this section, we present the evolution of the selected tags considered as buzz on *Delicious* social network between the year 2003 and 2010. In this work, we consider the top-10 of the most popular tags on the whole dataset as the studied buzz.

Table 2: The top-10 of the most popular tags on *Delicious* between the year 2003 and 2010

Tag	Design	Tools	Video	Education	Webdesign	Web	Inspiration	Art	Web20	Google
Popularity	4060	2929	2236	2041	1907	1733	1723	1691	1653	1648

The evolution of each tag is presented as a graph of its popularity (number of use) along the temporal axis. In this study, we use the month granularity to study the evolution of each buzz. The visualization graph is presented in figure 1.

We can see that most of the studied tags represent the buzz characteristic: their popularity increases slightly in the beginning and then explodes at a time point and declines after that. We observe that, the most popularity period of all studied tags is around September 2010 - October 2010.

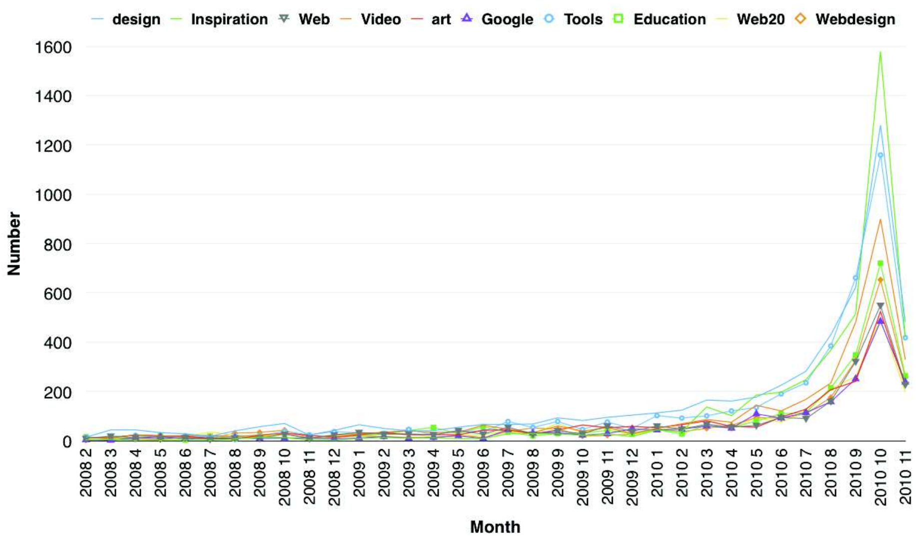
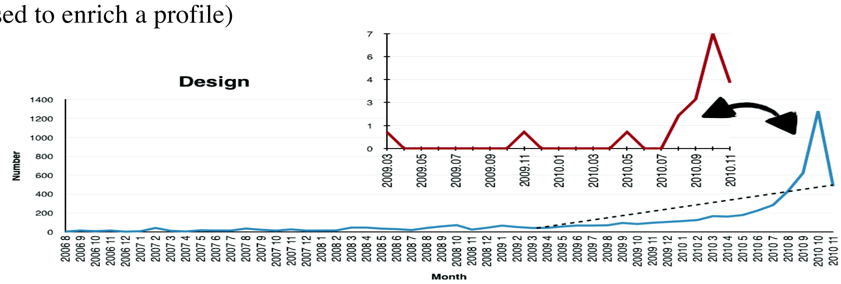


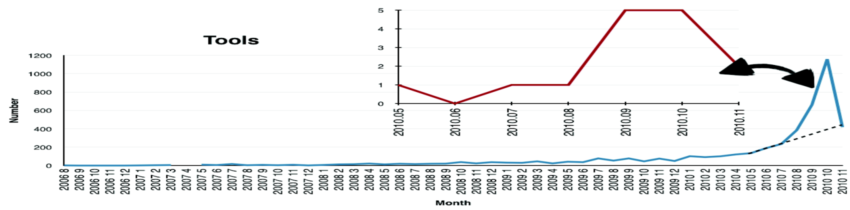
Fig.1. The evolution of the top-10 popular tags on *Delicious* between 2003 and 2010

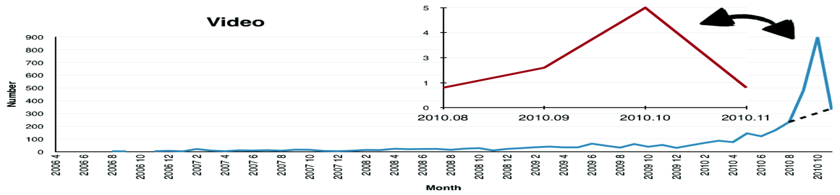
3.3 Analysis of the influence of user profile enrichment on buzz propagation

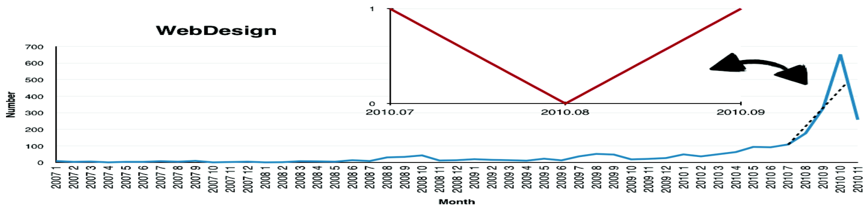
To analyse the impact of user profile enrichment on buzz propagation, we are interested in studying the correlation between the buzz propagation in the dataset and the buzz propagation in the enriched user profiles. In fact, we only analyse the result of the enrichment approach (not the whole enriched profile) with the buzz propagation.

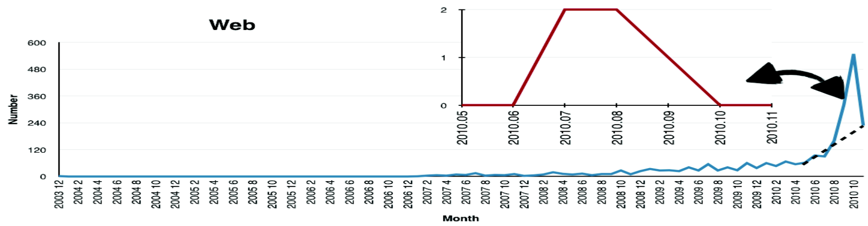
From the top-10 studied tags in the previous section, we found only 8 tags in the enrichment results (for all users in the dataset).The visualization graphs of these 8 tags is presented as follows in figure 2. It represents, for each tag, its popularity (number of use) in Delicious and its use in enriching profiles (number of times it is used to enrich a profile)

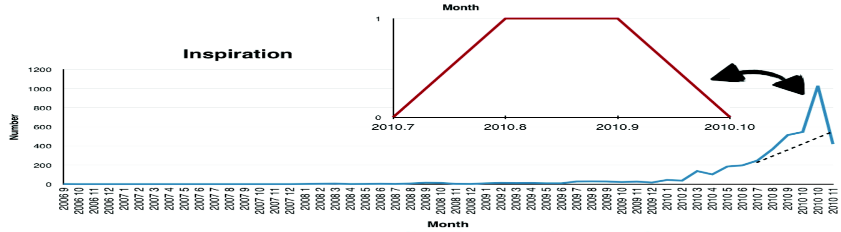


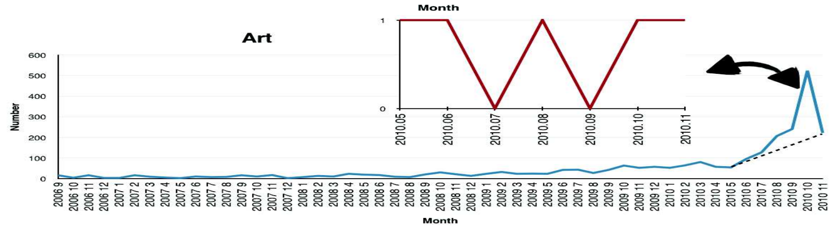

Tools
Number
Month


Video
Number
Month


WebDesign
Number
Month


Web
Number
Month


Inspiration
Number
Month


Art
Number
Month

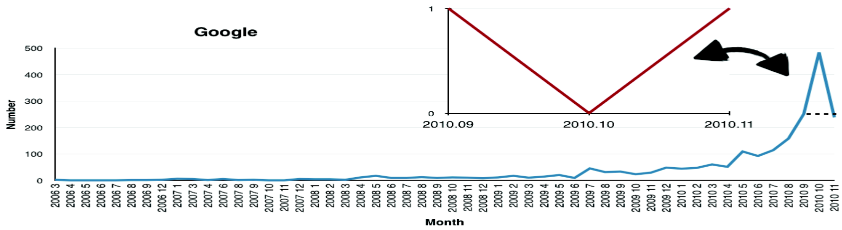


Fig.2. (Blue)The buzz propagation in the dataset and (Red) the buzz propagation in the enriched user profiles

The graphs above show that the studied tags are mostly retained in user profile after the enrichment process in the period in which they become popular. For example, the tag *Google* is retained in user profile between September 2010 - November 2010, the period in which the tag is more tagged by the whole users of this social network.

This analysis demonstrates that the popularity of the tags can be an important factor to the user profile enrichment process. If the tag is a buzz during a period, it has more chance to be extracted in the user profile enrichment process for this period.Thus, the user profile enrichment process can contribute to the buzz (as potential rumours) propagation in social networks.

3.4 Is user profile enrichment approach anticipating buzz propagation?

All along the previous analysis, we have analysed the buzz propagation in the whole network, independently of the user profile. In this section, we analyse each enriched user profile in order to show if the enrichment approach anticipate the buzz propagation. For each buzz found in the enrichment result, we detail the associated userID, the enrichment date, the number of occurrence of the tag (buzz) for the user before the enrichment date, the number of occurrence of the tag (buzz) for the user after the enrichment date and the date of the first use after enrichment by the user. These results are detailed in the tables 3, 4, 5, 6, 7, 8, 9 and 10.

Table 3: Analysis of the tag **design** according to each user

UserID	Enrichment Date	Occurrence before enrichment date	Occurrence after enrichment date	Date of the first use after enrichment
1094	8/10/2010	9	1	03/11/2010
1113	27/08/2010	11	51	29/08/2010
1113	29/08/2010	14	48	30/08/2010
16915	30/03/2009	1	6	16/04/2009
24802	10/11/2009	9	4	18/11/2009
8315	31/05/2010	1	2	24/06/2010
62070	20/09/2010	10	0	-
9960	28/09/2010	0	0	-
51543	30/09/2010	9	3	09/10/2010
2032	01/10/2010	6	0	-

8691	12/10/2010	32	38	13/10/2010
3233	21/10/2010	12	1	22/10/2010
1296	26/10/2010	0	0	-
11699	09/10/2010	0	3	01/11/2010
1701	09/10/2010	10	2	13/10/2010
15728	09/10/2010	29	17	11/10/2010
13222	03/11/2010	3	0	-
8452	04/11/2010	5	0	-
6067	04/11/2010	7	0	-

Table 4: Analysis of the tag **Tools** according to each user

UserID	Enrichment Date	Occurrence before enrichment date	Occurrence after enrichment date	Date of the first use after enrichment
8315	31/05/2010	1	20	01/06/2010
6120	31/07/2010	1	5	24/08/2010
46715	29/08/2010	1	6	12/10/2010
35745	16/09/2010	18	30	17/09/2010
11699	24/09/2010	2	34	26/09/2010
1328	29/09/2014	0	0	--
7396	19/10/2010	2	5	20/10/2010
2315	21/10/2010	7	2	27/10/2010
8554	22/10/2010	1	13	25/10/2010
70894	27/10/2010	12	6	01/11/2010
1505	29/10/2010	11	0	29/10/2010
13102	05/11/2010	8	2	06/11/2010
23135	06/11/2010	16	0	06/11/2010

Table 5: Analysis of the tag **Video** according to each user

UserID	Enrichment Date	Occurrence before enrichment date	Occurrence after enrichment date	Date of the first use after enrichment
74708	16/08/2010	7	9	23/08/2010
13084	14/09/2010	6	8	15/09/2010
4742	30/09/2010	2	4	06/10/2010
6796	12/10/2010	16	1	21/10/2010
8452	20/10/2010	4	0	--
11690	20/10/2010	7	0	--
8775	21/10/2010	2	1	21/10/2010
1701	21/10/2010	12	2	04/11/2010
12847	02/11/2010	2	0	--

Table 6: Analysis of the tag **Webdesign** according to each user

UserID	Enrichment Date	Occurrence before enrichment date	Occurrence after enrichment date	Date of the first use after enrichment
6120	23/07/2010	2	0	--
9660	28/09/2010	1	3	01/10/2010

Table 7: Analysis of the tag **Web** according to each user

UserID	Enrichment Date	Occurrence before enrichment date	Occurrence after enrichment date	Date of the first use after enrichment
13973	16/07/2010	2	12	02/08/2010
12506	05/07/2010	1	0	--
1113	27/08/2010	1	17	29/08/2010
1113	29/08/2010	3	15	31/08/2010
13084	14/09/2010	1	0	--

Table 8: Analysis of the tag **Inspiration** according to each user

UserID	Enrichment Date	Occurrence before enrichment date	Occurrence after enrichment date	Date of the first use after enrichment
1113	27/08/2010	7	22	29/08/2010
51543	30/09/2010	7	2	9/10/2010

Table 9: Analysis of the tag **Art** according to each user

UserID	Enrichment Date	Occurrence before enrichment date	Occurrence after enrichment date	Date of the first use after enrichment
31272	25/05/2010	8	5	09/06/2010
11962	30/06/2010	2	55	02/07/2010
10567	26/08/2010	1	1	09/11/2010
1701	27/10/2010	6	1	06/11/2010
8452	04/11/2010	4	0	--

Table 10: Analysis of the tag **Google** according to each user

UserID	Enrichment Date	Occurrence before enrichment date	Occurrence after enrichment date	Date of the first use after enrichment
1505	05/09/2010	2	1	13/09/2010
11853	05/11/2010	7	0	--

From this analysis, we notice that:

1. Regarding the occurrence before/after the enrichment date: it varies according to different cases. In fact, we notice that users who used the tag only before the enrichment are about 23.63 %, the users who used it only after the enrichment are about 1.81 %, the users who used it before and after the enrichment are about 69.09 % and the users who never used the tag and we have enriched their profile with this tag are about 5.45 %. So, we can conclude that the enrichment approach is somehow dependant with the previous activity of a user. However, the amount of these buzz found in the enrichment results is relatively low comparing to their popularity in the initial dataset.

2. Regarding the date of the first use of a tag after enrichment: this date aims to show the ability of the enrichment approach to anticipate the buzz. The bigger isthe interval between the enrichment date and the first use of the buzz, the buzz is

anticipated. According to these tables, we notice that the minimum value of anticipation is 0 day (we enrich the same day of a current activity) and is about 4,8 % of all cases. The maximum value of anticipation is 75 days (associated to the userID=10567 in table 9). The average value of anticipation is 9 days and the median value of anticipation is 5 days.

4. Conclusion

In this paper, we have made a case study on some information of the social network *Delicious*. This latter, contains social annotations (tags) that are provided by the user. These tags contribute to influence the other users to follow this information or to use it.

This study relies on three main axes:

1. we have focused on tags considered as buzz and we have analysed their propagation through time. In this analysis, we have noticed that the number of users in the network influences the propagation. The more active a user isin specific periods of time, the more the buzz is present in these periods.
2. we have considered a user profile as the set of tags provided by him. We have used the result of our previous work on temporal user profile enrichment, in order to analyse the influence of this enrichment in the propagation of the buzz. We have noticed that the enrichment process contributes to propagate the buzz in almost all the cases (8 tags about 10 were found in the enrichment result). Thus, the enrichment contributes to propagate the buzz in the network.
3. we have also analysed each enriched user profile in order to show if the enrichment approach anticipate the buzz propagation. So, we can see the interest of filtering the information in order to avoid potential rumours and then, to propose relevant results to the user (e.g. avoid "bad" recommendation). We have found that the enrichment approach is somehow dependant with the previous activity of a user. Also, the amount of these buzz found in the enrichment results is relatively low comparing to their popularity in the initial dataset. The anticipation varies from 0 day to 75 days. And the average is 9 days.

As perspectives, in order to reduce the buzz propagation, that may be potential rumours, we should take into consideration a buzz filtering process before applying our enrichment approach. Also, we plan to enlarge this case study more than 10 tags. Thus, to study the evolution of the other buzz and also the influence of the enrichment approach on the buzz propagation.

References

1. Manzat, A., Grigoras, R., Sèdes, F.: Towards a User-aware Enrichment of Multimedia Metadata. Workshop on Semantic Multimedia Database Technologies (SMDT 2010), Saarbrcken, Germany, Vol. 680, CEUR Workshop Proceedings (2010) 30-41.

2. Mezghani, M., Zayani, C-A.,Amous, I., Péninou A., Sèdes, F.: Dynamic Enrichment of Social Users' Interests. IEEE Eighth International Conference on Research Challenges in Information Science (RCIS). IEEE (2014) 1-11
3. Cantador, I., Brusilovsky, P., Kuflik, T.:2nd Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011). Proceedings of the 5th ACM conference on Recommender systems, RecSys, ACM, New York, NY, USA, 2011.
4. T. Hashimoto, T. Kuboyama, and Y. Shirota, "Rumor analysis framework in social media," in TENCON 2011 - 2011 IEEE Region 10 Conference, 2011, pp. 133–137.
5. Rosnow, Ralph L. and Alan J. Kimmel. 2000. "Rumor." Pp. 122–23 in Encyclopedia of Psychology, Vol. 7, edited by Alan E. Kazdin. New York: Oxford University Press.