Discovering most significant news using Network Science approach

Ilya Blokh[1] and Vassil Alexandrov[2]
[1]Perm State University, Russia,
[2]ICREA-Barcelona Supercomputing Centre, Spain.
ilyablokh@gmail.com, vassil.alexandrov@bsc.es

Abstract

The role of Internet mass media increased greatly in the recent years. Internet media attracts more and more audience coverage and thereby its role in forming of public opinion strengthens. Public organizations, political parties, companies as well as public figures, politicians and entrepreneurs are interested in any mentioning of them or of some specific facts concerning them in media including Internet media. They often trace a quantity of web-sites citing them, and different sources where some significant news could appear. Significant news finding could be also useful for common Internet uses since the amount of daily media publications is large and some news which are significant for someone could be missed. Thus significant (in some context) news identification is essential and actual problem.

Approaches for analyzing various types of news are widely different. From traditional search engines based on indexing to semantic methods based on idea of Semantic Web, keywords and keyphrases processing, NLP methods. Ontology based approaches have a remarkable prevalence too. In addition, some methods based on Data Mining technologies and machine learning used to resolve a problem of news search, interpretation and analysis. This paper is devoted to the use of Network Science approach for news analysis.

The origin of news provider could be some Internet mass-media (such as cnn.com, bbc.com and others) as well as a page in a social network or blog. In our research we decided to take Twitter as a typical representative of the social networking phenomenon as it is one of most subscribed social networks. Messages spread in Twitter after their appearance and a quantity of references to the message (tweet) appears. One of the spreading instruments of this social network is retweet mechanism. On the first stage of our research we focus on this mechanism and base our approach on small entities of Twitter - on tweets. We use network science approach to analyze messages and find significant news. Tweets act as nodes of network. Each tweet could be retweeted many times by some external twitter users. We define each retweet being one outgoing tie from a node. Thus for each tweet feed we have some degree distribution. Next we examine this distribution for scale-free property, define the significance and the threshold for it. Then we reveal a set of significant news from a given sample.

We interpret the news significance as a qualitative property of some piece of news and in the same way try to define some qualitative threshold to measure it. Significance analysis could help to characterize how different events impact on society or on separate social groups, to make research about events interconnection and to analyse news information spreading features. We define a quantitative metric of news significance as its retweet count for some time period. In the subsequent research we expect that this metrics will become more complex by aggregating information not only about retweet count but also include the information about number of user replies on the given tweet, number of users who put the tweet in favorites and some additional data.

We crawled data from Twitter accounts of the ones of the popular news agencies represented in social network: CNN, BBC, NYTimes, Mashable and TechCrunch. Collected data covers period from July 2014 up to January 2015 and includes information about 16500 news retweets. We looked at each agency news feed as at network and using Network Science apparatus we built and plotted degree distributions for each agency. It was determined in other research works that the whole Twitter network has scale-free property. Our goal was to specify degree distribution of sub-networks based on news messages retweets and check the presence of scale-free property. So we revealed that composed networks are scale-free and then supposed that the news is significant thus meeting the condition that the corresponding node in the network is a hub. It is crude initial measure and will be refined further taking into account a variety of additional factors. We discovered a number of significant news from initial data and presented them in the paper.

Afterwards we determine the future research scope. First we plan to make the significance property more complex and take into attention more factors and increase the number of data sources and the volume of processed data. Then we will describe the optimization approach for significant news finding and analyzing. In general we project to extend the

problem of significant news discovering and interpretation into the problem of world informational war analysis. Here we define informational war as a demonstration of multidirected regular actions forwarded into influencing at social opinion in some area and consisted in mass media publication's policy, in Internet media particularly.

REFERENCES

[1] Quint Randle: A historical overview of the effects of new mass media: Introductions in magazine publishing during the twentieth century. Volume 6, Number 9 - 3 September 2001.

[2] H. Kwak, C. Lee, H. Park, S. Moon.What is Twitter, a social network or a news media?, Proceedings of the 19th international conference on World wide web, 591-600, 2010.

[3] Albert-László Barabási (2013). Network Science.

[4] Gourab Ghoshal, Albert-László Barabási (2011). Ranking stability and super-stable nodes in complex networks.

[5] Dashun Wang, Zhen Wen, Hanghang Tong, Ching-Yung Lin, Chaoming Song, Albert-László Barabási (2011)/ Information Spreding in Context.

[6] G. Ghoshal, L. Chi, A.-L. Barabási. (2013) Uncovering the role of elementary processes in network evolution. Scientifc Reports 3, pp.1-8.