



Selected Papers of AoIR 2016:
The 17th Annual Conference of the
Association of Internet Researchers
Berlin, Germany / 5-8 October 2016

RETWEETING THE NEWS: TOWARDS A FORMALIZED APPROACH TO THE STUDY OF TWITTER NETWORKS

Frederik De Grove, Evelien D'heer & Pieter Verdegem

iMinds – MICT – Ghent University

Introduction

Twitter allows for the re-distribution and dissemination of mass media content. The latter is labeled “secondary gatekeeping”, and reflects social media users’ combined effort to generate visibility for news items (Singer, 2014). Whereas the selection of news content is predominantly related to journalists, news consumers’ online interaction with the news serve as a second evaluation of news content. In turn, users’ online behavior with news content is incorporated in the newsroom and influences journalists’ news judgment (Tandoc & Vos, 2015).

In this paper, we investigate the network processes behind the popularity metrics (i.e. retweet counts) for tweets coming from traditional news organizations.

We argue that a possible approach in understanding the structural processes behind popularity metrics lies in building a model in which retweet counts are regressed on network characteristics (e.g. centralization parameters). However, the challenge in building such a model lies in the kind of parameter distribution against which to evaluate parameter estimates. Since network characteristics such as centrality metrics are inherently tied to network size and hence to retweet counts, comparing parameter estimates against a normal probability distribution would lead to erroneous conclusions regarding the null hypothesis. Therefore, this paper explores and compares two possibilities: (1) comparing parameter estimates against those from randomly generated networks with similar density and size and (2) comparing parameter estimates against those from randomly collected networks on Twitter.

Methodology

Our data includes 418 tweets, sent out by eight different news organization from Belgium and the Netherlands that were retweeted at least ten times at most a 100 times. Each of these tweets contains a hyperlink, making reference to an article on website of the respective news organizations.

De Grove, F., D'heer, E. & Verdegem, P. (2016, October 5-8). *Retweeting the News: Towards a formalized approach to the study of Twitter networks*. Paper presented at AoIR 2016: The 17th Annual Meeting of the Association of Internet Researchers. Berlin, Germany: AoIR. Retrieved from <http://spir.aoir.org>.

First, we collected all users that retweeted these messages ($N=7728$). In order to reconstruct and understand the flow of retweets, we collected each of the users' personal follower networks. In this respect, we situate "micro-layer" practices, i.e. retweeting, within "meso-layer" structures (Bruns & Moe, 2014). Whereas most studies investigate communication networks based on replies or retweets, we account for users' stable set of connections through which communication *can* flow.

In order to predict the number retweets of mass media content, we build a regression model, including network parameters and control variables (e.g. kind of newspaper en country). In Figure 1 below, we present the variables that are included in the regression models. Data collection and analysis were conducted in R, using the packages TwitterR (Gentry, 2015) and igraph (Csardi & Nepusz, 2006).

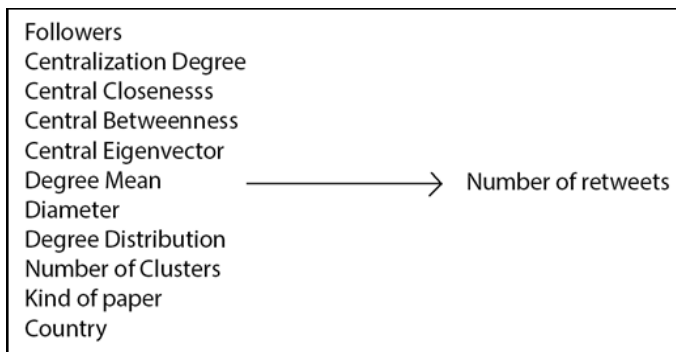


Figure 1 Predictors and dependent variable of the regression models

In order to test for the significance of the parameters, the regression models compare the data to randomly generated networks with the same size and density. This study was based 10.000 networks, randomly generated, varying in size ($N_{\min} = 10$, $N_{\max} = 100$) and density ($D_{\min} = 0.1$, $D_{\max} = 0.6$). Next, network characteristics such as centrality measures were computed on which network size is regressed.

The findings based on randomly generated networks

The results show that degree centralization^{***}, closeness centralization^{***} and betweenness centralization^{*} are significant predictors of the number of retweets news organizations' messages receive. In larger networks; the spread of information is mainly achieved through a few nodes (cf. degree centralization). However, information does not travel far. Large networks consist of closely connected nodes rather than dispersed ones (cf. closeness centralization). In the larger retweet networks, the information is spread less by bridges than in smaller networks (cf. degree betweenness). This is links up to the findings with respect to closeness centralization.

In short, the increase in the number of retweets messages receive, does not equal wider reach in the network. Further we found that the nodes that score high on closeness en degree centralization are more often elites (such as politicians or journalists) rather than non-elites. Hence, the spread of news content on social media does not run via "audiences" only.

The findings based on random sample data from Twitter

Despite the fact that network simulations have been randomly generated, all network characteristics significantly predict network size. What is more, they explain 90% of the variation in network size. Hence, this shows the need for a baseline that can be used to reliably evaluate the effect of network characteristics on Twitter. Such a baseline would allow us to detect and identify non-random processes such as secondary gatekeeping.

In particular, we propose to start from a random sample of Twitter messages for which we count retweets and calculate network measures (such as presented in Figure 1). These measures could serve as a baseline to compare our results, as well as other work on retweet networks. At the conference, we will be presenting preliminary results based on the data already collected. More in general, the construction of a reliable baseline for Twitter research is an important contribution to the formalization of the field of Twitter research.

References

Bruns, A., & Moe, H. (2014). Structural layers of communication on Twitter. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 15–28). New York: Peter Lang.

Csardi G, & Nepusz T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems* 1695. <http://igraph.org>

Gentry, J. (2015). *twitterR: R Based Twitter Client*. R package version 1.1.9. <http://CRAN.R-project.org/package=twitterR>

Singer, J. B. (2014). User-generated visibility: Secondary gatekeeping in a shared media space. *New Media & Society*, 16(1), 55–73.

Tandoc, E. C., & Vos, T. (2015). The Journalist is Marketing the News. *Journalism Practice*, 0(0), 1–17. <http://doi.org/10.1080/17512786.2015.1087811>