Are Most-viewed News Articles Most-shared?

Yangjie Yao and Aixin Sun

School of Computer Engineering
Nanyang Technological University, Singapore
yyao002@e.ntu.edu.sg axsun@ntu.edu.sg

Abstract. Despite many users get timely information through various social media platforms, news websites remain important mainstream media for high-quality news articles and comprehensive news coverage. Moreover, news websites are becoming well connected with the social media platform by enabling one-click sharing, allowing readers to comment on the articles, and pushing news update to social media through dedicated accounts. In this paper, we make the first step to analyze user behavior for news viewing, news commenting, and news sharing. Specifically, we focus on the sets of most-viewed, most-shared, and most-commented news published by a major news agency for about two months. Through topic modeling and named entity analysis, we observe that economy news is more likely to be shared and sports news is less likely to be shared or commented. News about health has higher chance of being shared, but does not attract large number of comments. Lastly, users are more likely to comment on than to share politics news.

Keywords: User behavior, News sharing, News commenting, News viewing, Popular news

1 Introduction

The popularity of social networking platforms (*e.g.,* `Twitter` and `Facebook`) is redefining the roles of information provider and information consumer and changing the way people access and receive information online. On the one hand, more users receive information pushed by other users through social platforms, which potentially reduces the number of direct visits to news websites. On the other hand, news websites remain important mainstream media for providing high-quality news articles written by professionals and offering comprehensive news coverage.

To be better connected with users and attract more visits, most news websites facilitate social interactions among their readers in at least three forms. One form of social interaction is to enable user discussion through comments to the news articles. Comments received from readers are maintained by the news websites and are often presented together with the news articles to the next readers [6]. Another form of social interaction is to minimize the effort for users to share a news article to her social networks. Many news websites provide a list of buttons; sharing a news article to Facebook, Twitter or other social networks

Fig. 1. The 4 categories and the 5 sections in most popular headlines

becomes a single click of a button. The last form of social interaction is to directly push news articles to the social networks through the news providers' accounts registered with the social networks. It is observed that mass media accounts (*e.g.,* CNN Breaking News, the New York Times, TIME) gain large followership in Twitter [9]. The link (often with a short description) of a news article then reaches more users in a social network though user re-sharing. The next question is: to what extent the user commenting and sharing mechanism influence news article viewership?

In this paper, we take the very first step and report a preliminary study on the most-popular news articles from a major news provider and try to answer the following questions: (i) are the most-viewed news articles most-shared, and vice versa? and (ii) are the most-viewed news articles most-commented, and vice versa? The answers to the above questions would help the news providers to better understand users' news reading behavior, so as to improve the effectiveness of news delivery to users through all possible channels including social platforms and news personalization or recommendation [4, 11, 12].

In the following, we first present the data collected for this study and then report the analysis based on topic modeling and named entity extraction.

2 Dataset

We collected the most-popular headlines published by Yahoo! News[1] for about two months from 15 April 2013 to 13 June 2013. Illustrated in Figure 1, the popular news headlines are categorized into most-popular, most-viewed, most-shared, and most-commented, for five sections: All, U.S., World, Science, and Health. For each kind of popular news (*e.g.,* most-viewed) in each section, maximum 100 news headlines are listed. We crawled the popular headlines and the full content of the news articles on daily basis at a fixed time. The number of distinct news articles collected for most-viewed/-commented/-shared in each section is listed in Table 1. We do not include the category "most-popular" in our following study because the meaning of popular here is not clearly defined. Observe from Table 1, the number of news falling under Health and Science sections is much smaller compared to the other three sections probably due to topic specificity.

[1] `http://news.yahoo.com/popular/`. Accessed on 20 June 2013.

Table 1. Number of distinct news articles under each category/section

Category	All	World	U.S.	Health	Science
Most-viewed	5472	4772	4901	1259	1435
Most-commented	5526	4897	4843	1188	1420
Most-shared	5034	4708	4503	1222	1403

Table 2. The conditional probability of news being most-viewed/-commented/-shared

Category	All	World	U.S.	Health	Science
P(most-viewed|most-commented)	0.67	0.86	0.84	0.98	0.97
P(most-commented|most-viewed)	0.68	0.89	0.83	0.92	0.96
P(most-viewed|most-shared)	0.69	0.82	0.84	0.98	0.98
P(most-shared|most-viewed)	0.63	0.81	0.75	0.95	0.96

The first row in Table 2 reports the conditional probability of a news article being most-viewed provided it is one of the most-commented articles in one of the five sections. Similarly, the conditional probabilities of being most-commented/most-shared are reported in the table. Observe that, the conditional probabilities reported under the Health and Science sections are above 90%. That is, for a news article, reporting a new finding in Health or Science area, if it is one of the most-viewed, very likely, it is one of the most-commented and the most-shared, and vice versa. One possible reason is that Health and Science are relatively topic-specific and the news articles are often about advices or new findings in these two areas with good support from scientific studies. Most users are not experts in Health or Science. Thus users have common background or common context in understanding the news. In other words, users' self-interests and inter-subjective interests[2] are likely to be the same.

For news articles falling under World and U.S. sections, the conditional probabilities are mostly over 80%. Under All section, the chance of a news article being most-viewed is below 70% even if it is one of the most-shared or most-commented article. That is, for a news article under this section, a good number of users read the article but do not share or comment it. A most-shared article in this section may not attract enough viewership to make it one of the most-viewed article. Compared with World and U.S., news articles in All section cover all happenings worldwide and cover various diverse topics. The diverse topics may affect users' behavior in viewing, commenting, and sharing the news articles. We therefore conduct our analysis mainly on the news articles in All section.

3 Analysis

We now analyze the news articles under All section and try to understand why not all most-viewed articles are not among most-shared/-commented or

[2] Inter-subjective interest refers to one user's prediction of other users' interest.

the most-shared/-commented articles are not among the most viewed. To begin with, we make the following assumptions.

- A news article's viewership consists of two groups of users: (i) users who discover the URL by themselves (*e.g.,* visiting the news website, news search), and (ii) users who click the URL from their social networks' feeds (*i.e.,* the URL is shared in the social networks).
- After reading a news article, a user may: (i) leave a comment on the news website, (ii) share the link in her social network with her description/comment of the link, and (iii) leave the page without commenting or sharing.
- Because we cannot access the exact number of views, number of shares, and number of comments each news article receives, all our analysis will be based on relative ranking. For example, if news article a_v is listed under most-viewed but not most-commented, and news article a_c is listed under most-commented but not most-viewed, we assume that a_v receives more views than a_c, and a_c receives more comments than a_v.

Let V, S, and C be the sets of news articles that are most-viewed, most-shared, and most-commented, respectively, in the All section. Next we perform topic modeling to analyze the topic distributions of the documents in All section.

3.1 Analysis by Topic Modeling

Topic modeling has demonstrated promising results in understanding the topic distribution of documents as well as in many prediction tasks [2]. Here, we are interested in finding out whether the topics of the news articles influence the viewing, commenting, and sharing behavior. More specifically, we are more interested in finding the topics for which the news articles are most-viewed but are less likely to be most-shared (or -commented), or the news articles are most-shared (or -commented) but are less likely to be most-viewed.

We adopt latent Dirichlet allocation (LDA) in our analysis. In LDA, a document in the collection is a distribution over a set of topics, and each topic is a probabilistic distribution over words. Given all the most-viewed/-shared/-commented documents in All section, we applied standard LDA model with following parameter setting: number of topics is set to 100[3], the Dirichlet prior on the per-document topic distribution $\alpha = 0.5$, the Dirichlet prior on the per-topic word distribution $\beta = 0.01$, the number of iteration is 1000.

Viewing vs Sharing With the result of topic modeling, each news article has a distribution over the 100 topics inferred from the news collection. For easy analysis, we assign each news article one topic (*i.e.,* the topic with the highest probability among the 100 topics for this news article)[4]. With the topic

[3] Similar results were observed by setting the number of topics to 50 or 200 in our experiments.

[4] We have also conducted analysis by assigning multiple topics to each news article weighted by the LDA results and similar results were observed.

Table 3. Topics and user preference of sharing and viewing

R_s	[Topic] and words for topics with higher user preference of sharing than viewing
0.81	**[economy]** company, million, percent, sales, billion, business, year, market, shares, price
0.75	**[economy]** percent, economy, year, market, rate, growth, month, price, job, economic
0.74	**[economy]** bank, europe, germany, eu, government, euro, switzerland, country, financial, greece
0.72	**[health]** study, research, drug, people, risk, health, patient, disease, blood, weight
0.71	**[energy]** oil, energy, gas, plant, water, company, power, industry, environmental, production
0.70	**[unknown]** france, italy, spain, paris, europe, beat, australia, britain, brazil, de
0.67	**[economy/politics]** worker, job, time, employee, union, company, labor, hire, pay, business
0.64	**[health]** restaurant, add, food, calorie, cup, cheese, fat, salt, minutes, pepper
0.63	**[economy/politics]** loan, student, rate, pay, debt, detroit, interest, financial, plan, payment
0.61	**[health]** virus, disease, health, infection, hospital, people, case, antibiotic, patient, infect
R_v	[Topic] and words for topics with higher user preference of viewing than sharing
0.77	**[crime]** castro, women, cleveland, berry, police, house, home, dejesu, knight, kidnap
0.76	**[sports]** final, match, nadal, set, title, champion, win, play, year, open
0.75	**[sports]** wood, shot, hole, garcia, tour, birdie, golf, play, par, putt
0.74	**[celebrity]** jackson, bieber, lohan, justin, virginia, rehab, aeg, lindsay, paris, cuccinelli
0.71	**[sports]** game, team, play, season, miami, james, points, final, nba, player
0.71	**[unknown]** reuter, edit, report, states, united, told, additional, writing, david, washington
0.71	**[celebrity]** dear, abby, husband, married, box, wedding, couple, phillip, mother, wife
0.70	**[politics]** sanford, colbert, carolina, south, busch, weiner, mark, campaign, politics, district
0.69	**[crime]** colorado, holmes, witherspoon, denver, arrest, trooper, driving, police, toth, insanity
0.69	**[crime]** trial, case, sentence, judge, prosecutor, defense, murder, death, prison, jury

assignment, we get the number of documents for each topic in the set V (most-viewed news articles) and in the set S (most-shared news articles) respectively. For each topic, we then compute its *user preference of sharing*. Let t_s be the number of news articles under topic t in set S, and let t_v be the number of news articles under the same topic in set V. The user preference of sharing for topic t is computed as the ratio $R_s = \frac{t_s}{t_s+t_v}$. The user preference of viewing for the topic t is computed in a similar manner $R_v = \frac{t_v}{t_s+t_v}$. Note that, if a news article is among both most-viewed and most-shared, the news article is counted in both t_s and t_v for its assigned topic t.

Table 3 reports the topics selected by the user preference of sharing and viewing. The upper half of the table lists the topics with higher user sharing

preference than viewing ranked by R_s in descending order; the lower half of the table lists the topics of higher user preference of viewing than sharing ranked by R_v in descending order. For each topic, the top-10 most relevant words are listed by their probabilities of belonging to that topic. The topic labels, in boldface in Table 3, are manually assigned based on the topical words and the content of the news articles in the topic. From Table 3, we make the following observations:

- The top-10 topics with higher preference of sharing and the top-10 topics with higher preference of viewing are significantly different.
- Users are more willing to share news articles under topics of economy (5 out of 10) and health (3 out of 10). The economic topics are about employment environment, European economics, and policies related to economic problems including worker rights and student loan.
- Users are less likely to share news articles under topics of sports (3 out of 10), crime (3 out of 10), and celebrity (2 out of 10). News articles under the three sports topics are about tennis match, golf match and NBA match respectively. The three crime cases are widely reported in newspapers.

Usefulness is one of the key motivations users use social networks [10]. A user in a social network therefore carefully selects who to make friend with and/or who to follow so as to receive useful information from the selected friends/followees. On the other hand, to be able to maintain or increase one's social capital [5,7], it is important for a user to provide useful information to her friends or followers. After reading a news article, a user shares this article to her friends/followers if she believes that this piece of information is useful to others. A piece of useful information in a social network feed has the potential to catch attention, attract new friends/followers, or strength social bonding with other users by initiating a conversation. From this point of view, it becomes reasonable that users are more willing to share news articles under topics of health and economy which are perceived to be more relevant to everyone's daily life or have impact to everyone's daily life in short term (*e.g.,* news about worker's right and employment environment).

A user's friends/followers in social networks usually have diverse interests. For example, to users who are not interested in sports, news updates on sports become less relevant or even spam to them. In particular, followee's informativeness is a major factor affecting the decision to unfollow in Twitter [8]. Sports news and crime news, in this sense, might not be useful to most other users in a social network unless in a domain-specific community formed by many users sharing the same interest. Another key issue a user has to consider before sharing a news article in social network is that the action of sharing reveals what she reads to all her friends/followers. Depending on her social status, a user may not share news about celebrity gossips for example.

Viewing vs Commenting Table 4 lists the top-10 topics with higher user preference for commenting and top-10 topics with higher user preference for viewing respectively, computed in a similar manner as that for Table 3. Note

Table 4. Topics and user preference of topic-commenting and topic-viewing

R_c	[Topic] and words for topics with higher user preference of commenting than viewing
0.73	**[politics]** united, states, country, meeting, mexico, talks, president, kerry, plan, america
0.71	**[politics]** tax, budget, cut, house, spend, bill, billion, government, republican, year
0.70	**[economy]** percent, economy, year, market, rate, growth, month, price, job, economic
0.70	**[politics]** sanford, colbert, carolina, south, busch, weiner, mark, campaign, politics, district
0.69	**[politics]** state, law, states, federal, bill, group, government, require, policy, pass
0.67	**[politics]** immigrate, bill, immigrant, senate, reform, republican, border, illegal, legislation, house
0.67	**[politics]** israel, iran, palestinian, nuclear, netanyahu, jerusalem, west, state, gaza, arab
0.66	**[politics]** gay, marriage, sex, rights, vote, bill, support, couple, state, lesbian
0.65	**[politics]** obama, president, house, white, bush, barack, administration, washington, america, republican
0.65	**[politics/economy]** loan, student, rate, pay, debt, detroit, interest, financial, plan, payment
R_v	[Topic] and words for topics with higher user preference of viewing than commenting
0.86	**[sports]** final, match, nadal, set, title, champion, win, play, year, open
0.79	**[sports]** wood, shot, hole, garcia, tour, birdie, golf, play, par, putt
0.76	**[science]** solar, moon, sun, space, photo, comet, image, planet, eclipse, earth
0.72	**[disaster]** river, water, flood, rain, snow, inch, people, area, weather, dam
0.71	**[health]** study, research, drug, people, risk, health, patient, disease, blood, weight
0.71	**[health]** restaurant, add, food, calorie, cup, cheese, fat, salt, minutes, pepper
0.71	**[sports]** game, team, play, season, miami, james, points, final, nba, player
0.69	**[science]** scientist, research, light, planet, star, particle, matter, space, earth, galaxy
0.69	**[technology]**apple, google, phone, iphone, device, app, microsoft, user, technology, company
0.67	**[family]** family, mother, children, father, son, daughter, parent, home, year, husband

that, the topics listed in the lower parts of Table 3 and 4 are different because the user preference for viewing R_v is computed differently, one is for viewing against sharing (Table 3) and the other is for viewing against commenting (Table 4).

Observe from Table 4, users have very different preferences for commenting and viewing (but not commenting) news articles. Among the top-10 topics attracted lots of comments, 9 of them are politics. For topics with higher user preference of viewing only, 3 are about sports, 3 are about science and technology, and 2 are about health. The results suggest that users are willing to express themselves through commenting on news articles about political issues (*e.g.,* government administration, tax and budget, immigration, and gay marriage). However, these political issues have less impact to most people's daily life, at least in short term. Different from sharing through social networks, commenting on news articles can be made anonymously and the comments are not pushed to the social network feeds. A user therefore has "more freedom" of expressing

her opinion about a news article without the worry about offending some of her friends/followers who may hold different opinions about the political issues. On the other hand, when sharing a news article (*i.e.,* its URL) in her social network, a user usually adds a description or her comments to the article. Such descriptions/comments are push to her friends/followers.

We also observe that many comments are organized into conversations through replying to existing comments. A news article in this case serves as a starting thread of a temporal forum facilitating user discussions. Like in most forums, users participating the discussions are not strongly connected through social relationships.

To summarize the key observations made from Tables 3 and 4 based on topic modeling:

- Sports news often attracts large number of views. However, users are unlikely to share or to comment on news articles about sports compared to news articles of other topics.
- Health news has higher chance of being viewed and shared, but relatively does not attract large number of comments.
- Economy news is more likely to be shared and politics news is the most commented among all news articles users read.

3.2 Analysis by Named Entity

A news article often reports an event involving people, organization, location, and time. In this section, we conduct discriminant analysis based on the named entities recognized from the news articles, to find out the named entities that may attract large number of views, shares, or comments.

We utilized the Stanford NLP package[5] to extract names of people, organizations, and locations, from the news articles. Recall that we use V, S, and C to denote the sets of news articles that are most-viewed, most-shared, and most-commented. We now partition the news articles into 4 groups.

- $V-S$: This is the group of news articles that are among most-viewed articles but not in the most-shared articles. There are 1999 news articles in this group, and 17715 named entities are extracted.
- $S-V$: This is the group of news articles that are among most-shared articles but not in the most-viewed articles. There are 1561 news articles in this group, and 16866 named entities are extracted.
- $V-C$: This is the group of news articles that are among most-viewed articles but not in the most-commented articles. There are 1757 news articles in this group and 17617 named entities are extracted.
- $C-V$: This is the group of news articles that are among most-commented articles but not in the most-viewed articles. There are 1811 news articles in this group and 14700 named entities are extracted.

[5] `http://www-nlp.stanford.edu/`. Accessed 20 June 2013.

Table 5. The top-20 most discriminative named entities and topics for $S - V$ and $V - S$, respectively

Most discriminative NEs for $S - V$		Most discriminative NEs for $V - S$	
Named entity and [type]	Topic	Named entity and [type]	Topic
[O] S&P	economy	[P] Amanda Berry	crime
[O] FactSet	economy	[P] Ariel Castro	crime
[O] Fed	economy	[P] Gina DeJesus	crime
[O] Labor Department	politics	[P] Berry	unknown
[O] Federal Reserve	economy	[P] Michelle Knight	crime
[P] Ben Bernanke	economy	[P] DeJesus	crime
[O] Reuters Health	health	[L] Qusair	politics
[P] Tanya Lewis	science	[P] Knight	unknown
[O] Thomson Reuters	unknown	[P] Roger Federer	sports
[O] IBM	science	[P] Abigail Van Buren	columnist
[O] IMF	economy	[L] IL	unknown
[O] University of Pennsylvania	unknown	[P] Pauline Phillips	columnist
[L] Bangalore	unknown	[O] Mount Morris	unknown
[O] UBS	economy	[P] Jeanne Phillips	columnist
[P] Mike Smith	unknown	[P] Deval Patrick	crime
[O] Dow Jones	economy	[O] Foreign Ministry	politics
[L] German	unknown	[L] Golan Heights	unknown
[O] European Central Bank	economy	[L] South Korean	unknown
[O] National Academy of Sciences	science	[P] Jo-Wilfried Tsonga	sports
[O] Mayo Clinic	health	[P] Abby	unknown

Based on the extracted named entities, each news article can be represented as a list of named entities contained in it. To find out which are the most discriminative named entities for identifying news articles in one group against another (*e.g.,* $V - S$ against $S - V$), many feature selection techniques can be applied directly [13]. We adopted *Odds Ratio* in our analysis for its effectiveness in many text classification tasks.

Table 5 and Table 6 list the top-20 most discriminative named entities for the two groups $S - V$ and $V - S$, and the two groups $C - V$ and $V - C$, respectively. The type of each named entity determined by the Stanford NLP package (*e.g.,* [P]erson, [L]ocation, and [O]ganization) is indicated in the front of the name entity. Based on the news articles in which the named entity appear, we manually assign each named entity a topic. Nevertheless, it is hard to identify the topic of some named entities, particularly the named entities referring to country names or locations. Another reason for not being able to identify a topic is that, a name entity extracted is not a full name (*e.g.,* Berry).

Observe from Table 5, the top-20 most discriminative named entities are mostly organizations for $S - V$ while the most discriminative named entities for $V - S$ are mostly persons. The topics of named entities are quite consistent with that in Table 3, with economy being the dominate topic covering nearly half of the top-20 named entities. Many of these named entities are from the finance sector such as Federal Reserve, Ben Bernanke, Dow Jones, IMF, and European

Table 6. The top-20 most discriminative named entities and topics for $C - V$ and $V - C$, respectively

Most discriminative NEs for $C - V$		Most discriminative NEs for $V - C$	
Named entity and [type]	Topic	Named entity and [type]	Topic
[P] Rubio	politics	[L] Space.com	science
[P] Lindsey Graham	politics	[P] Novak Djokovic	sports
[O] Republican Party	politics	[P] Woods	sports
[P] Harry Reid	politics	[O] Barcelona	sports
[P] John McCain	politics	[P] Iain Rogers	sports
[O] Tea Party	unknown	[P] Berry	unknown
[L] Palestinians	unknown	[P] Ariel Castro	crime
[P] Chuck Schumer	politics	[P] Mark Lamport-Stokes	sports
[O] Senate Judiciary Committee	politics	[P] Gina DeJesus	crime
[P] Netanyahu	politics	[P] Miriam Kramer	sports
[L] Americans	unknown	[P] Djokovic	sports
[O] Labor Department	politics	[P] Jo-Wilfried Tsonga	sports
[P] Schumer	politics	[O] Real Madrid	sports
[P] Mark Felsenthal	politics	[P] Michelle Knight	crime
[O] House Ways and Means Committee	politics	[P] Roger Federer	sports
[O] Pew Research Center	politics	[O] Bayern Munich	sports
[L] D-N.Y.	politics	[P] Mike Wall	science
[P] Roberta Rampton	politics	[O] Spurs	sports
[P] Rand Paul	politics	[P] DeJesus	crime
[P] Mahmoud Abbas	politics	[P] Knight	crime

Central Bank. Again, we argue that news articles related to these named entities have higher chance of affecting many people in short time and are perceived to be useful to many users for sharing. Science and health cover a quarter. The sad stories about the kidnappings of Amanda Berry, Gina DeJesus, and Michelle Knight[6] gained large viewership but not many sharing.

Let us look at the named entities listed in Table 6. Almost all the named entities attracted large number of comments are from the politics topic. For named entities that attract large viewership but not commenting are mostly sportsman. The two observations are consistent with that from Table 4. Users are willing to comment on politics issues more freely (or even with anonymous ids) without sharing the comments with their friends/followers. On the other hand, news articles about sports gain a large readership but receive relatively fewer comments.

To summarize, the observations made from the analysis of named entities are consistent with the observations made from the analysis using topic modeling.

[6] `http://en.wikipedia.org/wiki/Kidnappings_of_Amanda_Berry,_Gina_DeJesus,_and_Michelle_Knight`. Accessed 20 June 2013

4 Related Work

User behavior understanding and analysis is a major research topic [5,7]. Particularly the studies on motivation of the use of social networks are related to our work to help to understand the possible reasons that a user would or would not share a news article after reading it. On the other hand, user behavior analysis on social platforms (*e.g.,* `Twitter` and `Facebook`) has attracted significant research interests [3,14]. However, social platform is much more complicated with many more factors (*e.g.,* number of friends, strength of the relationships, degree of activeness) affecting users' behavior. In our study, we are more focused on the textual content of the news articles.

The most related work to our study is the analysis of the relationship between different user actions (view, share, comment) reported in [1]. The authors found that the number of times people sharing a news article is related to the number of times people viewing this news article although the correlation is not very strong. This finding is consistent with our findings that users selectively share news articles depending on perceived usefulness of the topic of the news articles. The authors also reported that comparing to view action and share action, view action and comment action are even less correlated. We show in our study that politics topic receives large number of comments and sports topic receives large number of views but less commenting. In [1], the authors divided the news articles into different categories and found that the correlations between view action and other actions are diverse among categories. However, as the categories are predefined by news publishers, the number of categories is limited and may cover all news articles in fine granularity. In our study, we use topic modeling to infer the topics from the collection of news articles. Another major difference between our study is that, we use the most-viewed, most-shared, and most-commented news articles which are believed to be more representative for the view, share, and comment actions. The news articles used in [1] were randomly selected.

5 Conclusion

In this paper, we collected two months of most-viewed, most-shared, and most-commented news articles from a major news agency. Through topic modeling and named entity analysis we tried to answer the question: are most-viewed news articles most-shared or most-commented, and vice versa? Our analysis reveals that the sharing and commenting behavior from users is largely affected by the topic of news. Specifically, sports news articles receive large viewership but are less likely to be shared or commented; politics news articles are more likely to receive large number of comments; users like to share news articles about health and economy. We believe these findings are useful for news agencies in determining the best news promotion strategies to enlarge their readership. The findings are also helpful in the design of news personalization and recommendation systems. Although the lack of exact numbers of views, shares, and comments of the news articles in the data collection is considered as a limitation of this study, we believe the findings remain valid.

References

1. D. Agarwal, B.-C. Chen, and X. Wang. Multi-faceted ranking of news articles using post-read actions. In *Proceedings of the 21st ACM international conference on Information and knowledge management (CIKM)*, pages 694–703. ACM, 2012.
2. D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
3. P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun. Who should share what?: item-level social influence prediction for users and posts ranking. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 185–194. ACM, 2011.
4. G. De Francisci Morales, A. Gionis, and C. Lucchese. From chatter to headlines: harnessing the real-time web for personalized news recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining (WSDM)*, pages 153–162. ACM, 2012.
5. N. Ellison, C. Steinfield, and C. Lampe. The benefits of facebook "friends:" social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168, 2007.
6. M. Hu, A. Sun, and E.-P. Lim. Comments-oriented document summarization: understanding documents with readers' feedback. In *Proceedings of ACM SIGIR conference on Research and development in information retrieval*, pages 291–298. ACM, 2008.
7. K. Johnston, M. Tanner, N. Lalla, and D. Kawalski. Social capital: The benefit of facebook friends. *Behaviour and Information Technology*, 32(1):24–36, 2013.
8. H. Kwak, H. Chun, and S. Moon. Fragile online relationship: a first look at unfollow dynamics in twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1091–1100. ACM, 2011.
9. H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web (WWW)*, pages 591–600. ACM, 2010.
10. K.-Y. Lin and H.-P. Lu. Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior*, 27(3):1152 – 1161, 2011.
11. J. Liu, P. Dolan, and E. R. Pedersen. Personalized news recommendation based on click behavior. In *Proceedings of the 15th international conference on Intelligent user interfaces (IUI)*, pages 31–40. ACM, 2010.
12. S. O'Banion, L. Birnbaum, and K. Hammond. Social media-driven news personalization. In *Proceedings of the 4th ACM RecSys workshop on Recommender systems and the social web*, pages 45–52. ACM, 2012.
13. F. Sebastiani. Machine learning in automated text categorization. *ACM Comput. Surv.*, 34(1):1–47, 2002.
14. Z. Xu, Y. Zhang, Y. Wu, and Q. Yang. Modeling user posting behavior on social media. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 545–554. ACM, 2012.