Mainstream Media Behavior Analysis on Twitter: A Case Study on UK General Election

Zhongyu Wei[1], Yulan He[2], Wei Gao[3], Binyang Li[1], Lanjun Zhou[1], Kam-fai Wong[14]
[1]The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
{zywei, byli, ljzhou, kfwong}@se.cuhk.edu.hk
[2]School of Engineering & Applied Science, Aston University, Birmingham, UK
y.he@cantab.net
[3]Qatar Computing Research Institute, Qatar Foundation, Doha, Qatar
wgao@qf.org.qa
[4]Key Laboratory of High Confidence Software Technologies, Ministry of Education, China

ABSTRACT

With the development of social media tools such as Facebook and Twitter, mainstream media organizations including newspapers and TV media have played an active role in engaging with their audience and strengthening their influence on the recently emerged platforms. In this paper, we analyze the behavior of mainstream media on Twitter and study how they exert their influence to shape public opinion during the UK's 2010 General Election. We first propose an empirical measure to quantify mainstream media bias based on sentiment analysis and show that it correlates better with the actual political bias in the UK media than the pure quantitative measures based on media coverage of various political parties. We then compare the information diffusion patterns from different categories of sources. We found that while mainstream media is good at seeding prominent information cascades, its role in shaping public opinion is being challenged by journalists since tweets from them are more likely to be retweeted and they spread faster and have longer lifespan compared to tweets from mainstream media. Moreover, the political bias of the journalists is a good indicator of the actual election results.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Text analysis*; I.7 [**Document and Text Processing**]: General; J.4 [**Social and Behavioral Sciences**]: [Sociology]

General Terms

Social Media Analysis

1. INTRODUCTION

With the development of social media platforms, mainstream media organizations, e.g. newspapers and TV media, have played an active role in trying to get their readers and audience engaged and involved and strengthening their influence through social media tools. Although they are not the most followed users on Twitter, mainstream media still drives the discussions by creating most of the trending topics [**?**]. The important role of mainstream media has attracted a number of researches analyzing its characteristics and behaviors in social media. Studies have provided a brief picture about the basic ecological habit of mainstream media on Twitter including what functions they used, when they tweet, and why they tweet [**?**]. Others also look into the way mainstream media extends its influence on Twitter [**?**, **?**].

In this paper, we perform an investigation on tweets data[**?**] (over 900,000) collected for 8 weeks leading to the United Kingdom's General Election of 2010[1] to analyze the mainstream media behaviors and study whether they still have strong influence in shaping public opinion. We focus on the following questions: how to quantify media bias on Twitter; how the information originated from mainstream media propagates and diffuses in the Twittersphere; how the mainstream media compare with the most influential individuals in terms of social influence. In summary, our contribution is three-fold.

Firstly, we propose an empirical measure to quantify mainstream media bias based on sentiment analysis. Existing research on media bias analysis in social media mainly uses the quantitative measures [**?**, **?**] such as the number of media outlet citations or mentions of some political parties or politicians. Nevertheless, such measurements only reveal the existence of bias, but not whether a media outlet supports or attacks a particular political party. Our results show that media bias measured by sentiment analysis is a better indicator than the pure quantitative measures based on media coverage of various political parties.

Secondly, we analyze the diffusion pattern of information disseminated from mainstream media. We found that the opinion leaders of different mainstream media overlap little.

[1]There are three main parties in the UK General Election 2010, i.e., Conservative, Labour, and Liberal Democrat. We categorized the tweets relevant to different parties by comparing against a list of keywords and hash tags manually constructed.

24th ACM Conference on Hypertext and Social Media
1–3 May 2013, Paris, France

Furthermore, journalists play a vital role in extending the influence of mainstream media to wider public.

Lastly, we compare the mainstream media with other types of users in terms of social influence. Our analysis shows that mainstream media is still good at seeding prominent information cascades. However, individual journalists appear to be more effective information disseminator since tweets published by them are more likely to be retweeted and their tweets also spread faster and have longer life span compared to tweets from mainstream media.

2. MAINSTREAM MEDIA BIAS ANALYSIS

In this section, we propose a method for evaluating media bias base on sentiment analysis and present our findings in comparison with the actual media bias as perceived by audience. We focus on TV media only since their tweets volume is much larger than other media outlets such as newspaper organizations.

2.1 Media Bias Survey

The Media Blog conducted a survey to analyze the TV media bias between 19th and 23rd April 2010, just before the UK's General Election on 6th May 2010. 96% of respondents said they believe they have seen clear bias within the UK media. Respondents were also required to evaluate the party slant of four major TV media organizations in the UK, namely, BBC news, Channel 4 news, ITV news and Sky news. The result of the survey[2] is depicted in Figure 1.

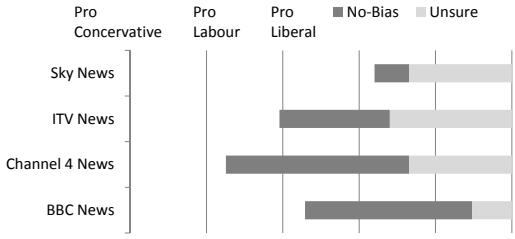

Figure 1: Survey result of TV bias on the UK political parties.

As can be observed in Figure 1, Sky News shows the strongest bias that nearly two-third of respondents said Sky News displays a clear pro-Conservative bias in its reporting, followed by ITV News with 28.8% of respondents believing that it shows pro-Conservative bias. Although many respondents hold that Channel 4 News and BBC News did not display clear bias, still a number of interviewers believe both of them reveal bias to different parties to some extent. We can also perceive that the Conservative Party receives most bias from TV media.

2.2 Media Bias Evaluation based on Sentiment Analysis

We selected Twitter feeds under each of the four major TV media outlets in the UK and extracted tweets published by them from the political tweets data. The volume of tweets varies for different TV media as can be seen in Table 1. BBC News having the most Twitter feeds generates the largest number of tweets, while ITV News published merely 112 tweets. It can be also observed that simply using media coverage quantities as bias measure does not generate meaningful results since all the four TV media outlets publish most tweets about the Labour party.

Table 1: Twitter Feeds distribution of the TV Media.

TV Media	BBC News	Channel 4 News	ITV News	Sky News
Feeds	9	2	3	4
Tweets	732	662	112	224
Conservative	286	258	29	114
Labour	**498**	**356**	**48**	**175**
Liberal	168	154	15	55

There are over 61% tweets contain mentions of more than one party. For example, the tweet, "*Labour have copied many Tory policy ideas, some of which they have previously denounced!*", expresses negative opinion to Labour, but neutral polarity to Conservative. Therefore, the traditional lexicon-based sentiment analysis methods might fail in this task. In our research, we employed OpenAmplify[3] for entity-level sentiment extraction from tweets. We compared the performance of entity-level sentiment analysis using a lexicon-based method[4] and OpenAmplify. Our results on the randomly selected 500 tweets show that the lexicon-based method only gives an accuracy of 54% while OpenAmplify significantly outperforms the former with 74% accuracy obtained. For each party, we count the total number of positive and negative mentions relating to this specific party across all the tweets. We define the media bias of a media outlet i towards a party j as:

$$\text{Media Bias}_{ij} = \frac{C_{ij}^{\text{pos}} + 1}{C_{ij}^{\text{neg}} + 1} - 1 \qquad (1)$$

where C_{ij}^{pos} and C_{ij}^{neg} denotes the total number of positive and negative tweets from a media outlet i towards a party j. Media Bias takes value 0 if there is no bias. And it is positive for positive bias and negative vice versa.

As can be seen from Figure 2, Sky News has clear pro-Conservative bias. The bias of Sky News can also be perceived by its negative bias value for the Labour party, since Labour is the principal opponent of Conservative in the UK General Election. Although ITV news shows a strong positive bias for Liberal, it is less statistically meaningful since there are only 9 tweets mentioning Liberal published by ITV news. For the other two parties, ITV displays pro-Conservative and anti-Labour bias which is similar to the survey results. Channel 4 News is neutral about Conservative, slightly positive about Labour, and more positive about Liberal. The rank of these three parties based on the bias value for Channel 4 is exactly the same as what the survey results reveal. BBC News is more neutral and displays least bias. This also correlates well with the actual survey results.

3. MEDIA INTERMEDIARIES ANALYSIS

[2] http://goo.gl/E83Hv

[3] http://www.openamplify.com/insights

[4] Based on the sentiment lexicon from http://www.cs.pitt.edu/mpqa/

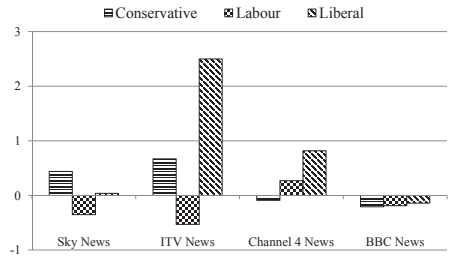

Figure 2: TV media bias based on sentiment results.

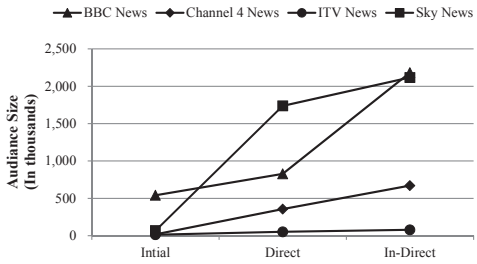

Figure 3: Change of audience size of TV Media with information diffusion. (Initial: the follower number of each TV Media; Direct: audience size of each media after retweeting of followers; In-Direct: audience size of each media after all the retweetings)

Wu et al. [**?**] has found considerable support of the "two-step flow" theory that a substantial amount of information originated from mainstream media is transmitted to masses through intermediaries. In this section, we want to find out who are those intermediaries for TV medias to spread the influence and how important they are in extending media influence to masses.

3.1 Intermediaries Distribution

In our research, information propagation is identified by tracking the inherent repost behavior of Twitter as well as the rebroadcasting of a URL that has been previously posted. Besides, we introduce *audience size* to identify the influence of a mainstream media which is the number of users who can receive the messages originated from the target mainstream media. For all the retweeted messages, we further classify them into two categories, *direct retweets* and *indirect retweets*. *Direct retweets* are messages retweeted by the followers of the source, otherwise, messages are classified as *indirect retweets*. Intermediaries are defined as those who retweet messages from a source user directly. Intermediaries contribute to the diffusion of social influence of mainstream media by spreading tweets from media to their own followers and hence significantly enlarge the audience size of mainstream media.

Through direct and indirect retweeting, the audience size of TV Media on Twitter increases dramatically. As can be seen from Figure 3, by direct retweeting, the audience size of Sky News increases by 24.16 times and becomes the largest of all the four TV media at this stage. The second fastest growth rate of the audience size by direct retweeting is 19.15 for Channel 4 News. BBC News has the most followers, however, direct retweeting only triples its audience size. Nevertheless, indirect retweeting significantly increases the audience size for BBC News with 1,300,000 more readers reached. The audience size of ITV news does not change much through retweeting due to the small volume of seed tweets(refer to Table 1).

3.2 Intermediaries Contribution to Media Influence

To measure the contribution of an intermediary in extending media influence, we consider both of his/her retweet volume and the number of his/her followers who do not follow the source media directly. Let C_{ij}^{RT} denotes the total number of tweets retweeted directly by an intermediary j from source media i, F_x denotes the set of followers of a user x, the contribution of j in extending the influence of source media i is defined as:

$$\mathrm{Support}_{ij} = C_{ij}^{\mathrm{RT}} \times \|\{k|k \in F_j \text{ and } k \notin F_i\}\| \quad (2)$$

where $||.||$ denotes the size of target set. The top 10 most influential intermediaries for each TV media are selected based on the support value and are depicted in Figure 4. Each node denotes an intermediary. The size of a node represents the popularity of the intermediary calculated by the number of followers. The thickness of an edge represents the number of direct retweets. It can be observed that the number of overlapping users of the top 10 intermediaries among different TV Media is only three. Indeed, considering all the intermediaries of the four TV Media, the overlap is also little, with only 148 users out of 2158 users retweeting messages from more than one TV media outlet.

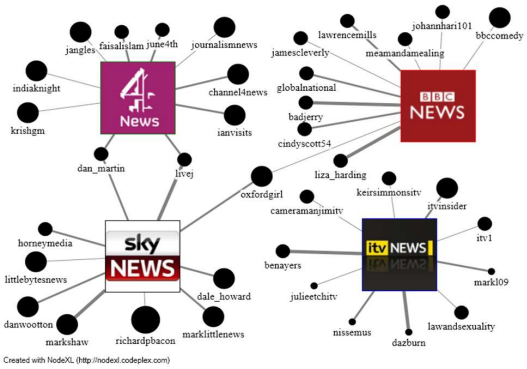

Figure 4: The top 10 intermediaries for 4 TV Media.

We classify a total of 37 most influential intermediaries into different user types manually and the distribution is shown in Figure 5. The majority is journalists followed by bloggers. We did a further analysis on journalists and see whether they are affiliated with the TV media organizations. The identity of users are confirmed manually by combining the information from their Twitter homepage and Wikipedia page as well as the TV media staff page. Result showed that over half of the journalists who serve as intermediaries for these TV media organizations are actually not affiliated with them. What is more interesting is that, although Sky News has the most journalists as its intermediaries (6 out of 10) among all the TV media, none of them work for Sky (one of them even work for BBC Radio). Furthermore, the growth

rate of the audience size of Sky News is the fastest through direct retweeting. Hence, we can speculate that journalists play an important role for extending influence on Twitter for mainstream media.

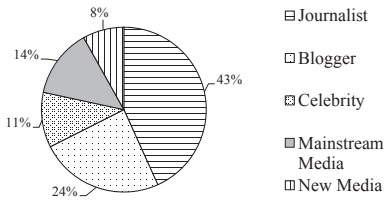

Figure 5: The distribution of the most influential intermediaries.

4. MEDIA INFORMATION DIFFUSION

In this section, we compare information diffusion patterns from different categories of sources and analyze the characteristics of the mainstream media outlets to understand if they still play a dominant role in shaping public opinion during the UK's General Election.

4.1 Information Diffusion by Source Category

We first extract Twitter users who have their tweets retweeted directly (by their followers) for more than 100 times. This results in 110 users. We then manually classify them into 6 categories including, mainstream media, celebrity, journalist, new media, political party (including politicians), and others. We pool all the tweets within one category together and analyze information diffusion patterns of each category.

The user account distribution of different categories is shown in Figure 6. Celebrity is the biggest category with 26 users including actors, singers, famous writers, etc., followed by Journalist and Mainstream Media. Twitter accounts in the New Media category are mainly online news feeds such as Politics Home UK[5]. Party category includes accounts created by political parties or their supporters during the election and accounts of member of Parliament of different parties and other political activists. The Other category consists of users who tweet a lot and are retweeted by a considerable number of times. As it is hard to extract common characteristics from users in this category, we exclude it from our subsequent analysis.

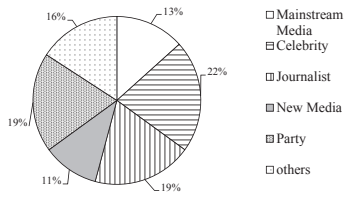

Figure 6: User distribution of various categories.

Supposing a single information cascade is generated by a seed tweet followed by all of its retweets, we calculate the distribution of information cascades by source category as shown in Figure 7. It can be observed that most information cascades are originated from media (including Mainstream Media and New Media) and Party.

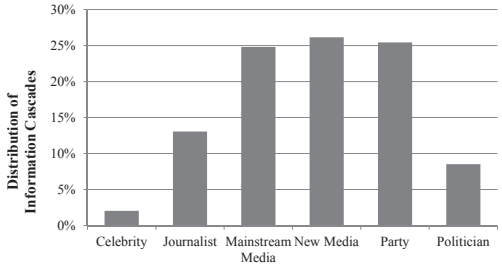

Figure 7: Distribution of information cascades by source category.

Table 2 summarizes the information diffusion attributes of tweets from different categories. *Retweet rate* is the retweet to tweet ratio. It measures how well users from certain category engage with their audience. *Average Retweet times per tweet* calculates how many times a tweet has been retweeted on average. *Lifespan* is the average time lag in hours between the first appearance of a seed tweet and the last appearance of its retweet.

As can be seen from Table 2, New Media publishes highest number of tweets. Nevertheless, both its retweet rate and average retweet times per tweet are the lowest among all the categories here. Journalist has the highest retweet rate followed by Mainstream Media indicating they maintain good engagements with their audience since over 81% of their tweets have been retweeted. We also notice that the retweet rate has no correlation with the number of followers. Journalist has much lower average follower number compared to Celebrity and Mainstream Media. Yet, it has a higher retweet rate than those two categories. This is in contrast with existing findings [**?**] that the more followers one has, the more likely his/her tweets gets retweeted.

Celebrity has exceptionally high value of average retweet times per tweet. However, only 2% of information cascades are originated from them as seen from Figure 7. Hence, they don't appear to have strong influence during the UK's general election. Journalist has the second highest value of average retweet times per tweet, followed by Party. Also, Tweets originated from Celebrity have the longest lifespan on average followed by Party and Journalist. The total retweet times for the Party category is significantly higher than those of other categories, indicating that people tend to retweet a lot from users accounts associated with political parties or politicians during the election period.

Figure 8 shows the information diffusion speed of the five categories. For all the categories, half of the diffusion diminishes within one hour. More than 80% of diffusions ceases in the second hour. Tweets originated from New Media and Journalist spread faster than those from other categories. Tweets from Party and Celebrity reach masses with the lowest speed.

4.2 Bias Analysis of Different Categories

In this section, we analyze the bias of different categories based on our proposed bias measure in Equation 1. Figure 9

[5]http://www.politicshome.com/

Table 2: Information diffusion attributes of tweets from different categories.

Category	Avg. Follower No.	Tweets	Retweeted Tweets	Retweeted times	Retweet Rate	Average Retweet Times per Tweet	Lifespan (hours)
Celebrity	**458944**	162	124	7670	0.765	**61.855**	**64.0**
Journalist	9277	948	791	9731	**0.834**	12.302	22.3
Mainstream Media	66060	1856	1502	11476	0.809	7.64	11.02
New Media	4184	**4588**	**1583**	7750	0.345	4.896	13.8
Party	7564	2033	1482	**17768**	0.729	11.989	52.4

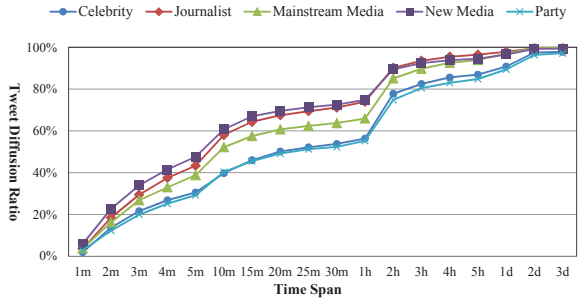

Figure 8: Information diffusion speed of various categories. Tweet diffusion ratio is the percentage of the number of retweets within a time period over the total number of retweets.

shows the bias analysis result of different categories. As we can see, Celebrity and Party display strong pro-Labour bias. We did a further analysis on the individual user accounts and found that the majority users are either associated with the Labour Party or declared to support Labour. Indeed, among the top 10 most retweeted accounts relating to political parties, 7 are from the Labour Party. From which we can see that Labour has spent significant efforts on the election campaign. Nevertheless, Labour still failed to win the last general election.

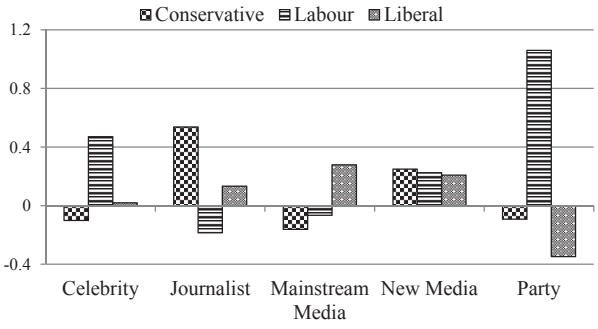

Figure 9: Bias of various categories.

Mainstream media outlets appear to be critical that they have slightly negative polarity towards both Labour and Conservative. On the other hand, new media exhibits positive bias towards all the three parties. Journalists show strong pro-Conservative and anti-Labour bias. They also display positive bias towards Liberal. We can observe that the bias exhibited by the Journalist category correlates well with the actual UK election results, in which, the percentage of seats obtained by Conservative increases by 3.7, that of Labour drops by 6.2 and that of Liberal also grows slightly by 1.0.

5. CONCLUSION

In this paper, we have studied media behavior on Twitter during the UK's General Election in 2010. Our findings show that while most information flows are originated from media, they seem to lose their dominant position in shaping public opinion during the UK's general election. The Journalist category appears to have the strongest influence as it has the highest retweet rate and gets the highest average retweet times per tweet. Its bias also reflects the actual political landscape in real world. The analysis result also reveals that journalists play an important role in enlarging the audience of mainstream media and extending media influence in social web.

6. ACKNOWLEDGEMENT

This work is partially supported by General Research Fund of Hong Kong (No. 417112).

7. REFERENCES

[1] J. An, M. Cha, K. Gummadi, and J. Crowcroft. Media landscape in twitter: A world of new conventions and political diversity. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 1–8, 2011.

[2] S. Asur, B. Huberman, G. Szabo, and C. Wang. Trends in social media: Persistence and decay. In *5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 434–437, 2011.

[3] T. Groseclose and J. Milyo. A measure of media bias. *The Quarterly Journal of Economics*, 120(4):1191–1237, 2005.

[4] Y. He, H. Saif, and Z. Wei. Quantizing opinions for political tweets analysis. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*, 2012.

[5] J. Holcomb, K. Gross, and A. Mitchell. *How mainstream media outlets use twitter*, 2011. http://www.journalism.org/analysis_report/how_mainstream_media_outlets_use_twitter.

[6] Y. Lin, J. Bagrow, and D. Lazer. More voices than ever? quantifying media bias in networks. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 193–200, 2011.

[7] B. Suh, L. Hong, P. Pirolli, and E. Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *IEEE Second International Conference on Social Computing (SocialCom)*, pages 177–184, 2010.

[8] S. Wu, J. Hofman, W. Mason, and D. Watts. Who says what to whom on twitter. In *Proceedings of the 20th international conference on World wide web (WWW)*, pages 705–714, 2011.