Visualizing Media Bias through Twitter

Jisun An
University of Cambridge

Meeyoung Cha
KAIST

Krishna P. Gummadi
MPI-SWS

Jon Crowcroft
University of Cambridge

Daniele Quercia
University of Cambridge

Abstract

Traditional media outlets are known to report political news in a biased way, potentially affecting the political beliefs of the audience and even altering their voting behaviors. Therefore, tracking bias in everyday news and building a platform where people can receive balanced news information is important. We propose a model that maps the news media sources along a dimensional dichotomous political spectrum using the co-subscriptions relationships inferred by Twitter links. By analyzing 7 million follow links, we show that the political dichotomy naturally arises on Twitter when we only consider direct media subscription. Furthermore, we demonstrate a real-time Twitter-based application that visualizes an ideological map of various media sources.

Introduction

Media influence has been widely studied in cultivation theory, which holds that the popular media like newspapers have power to influence people's view of the world and set their day-to-day norms. It is also well known that mainstream newspapers today have bias in selecting what to report and in choosing a slant on a particular report. Over 70 percent of Americans admit such bias (PewResearch 2004), and a number of studies have confirmed that left and right leaning news media consistently refer to different think-tanks in their stories (Milyo and Groseclose 2005; Gentzkow and Shapiro 2010).

Exposure to biased news information have several important consequences. It may increase intolerance of dissent and foster more ideological segregation of political and social issues (Glynn et al. 1999). Furthermore, it can affect the political beliefs of the media audience and could ultimately alter voting behavior (Vigna and Kaplan 2007). Therefore, tracking bias in everyday news and building a platform where people can receive balanced news information are important. Unfortunately, existing studies on identifying media bias have been restricted to examining a small set of news outlets, due to challenges in gathering and analyzing a huge amount of appropriate data (Milyo and Groseclose 2005; Gentzkow and Shapiro 2010). As a step toward building a such platform, we propose a novel model for inferring bias of news media outlets in real-time from the way Twitter users subscribe and disseminate news articles.

With the advent of social media services, news media outlets have started publishing on social networking sites. Likewise Internet users have moved from scanning traditional mediums such as newspapers and television to using the Internet, in particular social networking sites, to find news. In the popular microblogging site Twitter, users actively follow a wide set of news sources, form interpersonal networks, and propagate interesting news articles to their peers. These media subscription and interaction patterns, which had previously been hidden behind media corporations, poses as a new opportunity to understand media supply and consumption across society.

Social media provide an opportunity for researchers to examine how different sources report different angles on the same event and how the news consumers react to that. Conducting a similar study on the traditional media would have been difficult, as it would have required extensive surveys to gather the required data. By contrast, since all interactions in social media are recorded online and are often made publicly accessible, gathering and aggregating data—processes that are largely automated—can yield a view of an ideological separation of media sources (An et al. 2011).

In this work, we investigate a methodological issue: can we draw a valid ideological map of news media based on users' subscription and interaction patterns. In order to answer this question, we focused on 24 major U.S. based news outlets in Twitter and their aggregate 7 million followers. We created a distance model based on the co-subscription relationships and mapped the news media outlets along a single dimensional dichotomous political spectrum. Based on the distance measure, we also built a real-time Twitter-based application that visualizes an ideological map of various media sources.

Our data analysis revealed extreme polarization among media sources, indicating that the political dichotomy naturally arises on Twitter in the media subscription patterns of users. The political ideological map in user subscription networks was strikingly similar to that proposed in previous work (Milyo and Groseclose 2005), which assigned a ADA (Americans for Democratic Action) score for each media outlet by manually investigating the think-tank citations of its news articles.

Copyright © 2012, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Methodology

The Twitter dataset

We obtained the Twitter data published in a previous work (Cha et al. 2010), which comprises the following three types of information: profiles of 54M users, 1.9B directed follow links among these users, and all 1.7B public tweets that were ever posted by the collected users. For the analysis, we identified a set of news media sources by consulting: (1) *http://newspapers.com* website, which listed top 100 news papers in the U.S. by circulation; and (2) Twitter's "Browse Interest" directory at *http://twitter.com/#!/who_to_follow/interests/news*. From these two lists, we searched news providers, including main stream news outlets as well as individual journalists and anchors as it is also known that they have distinctive set of audience and play a prominent role as news providers.

We only considered U.S. based news media sources, and that left us with 24 media sources in news category. Those media sources examined are listed in Table 1 with their political bias. We mapped the political leaning of media sources into three groups, left-wing (liberal), center, and right-wing (conservative), using a number of public data including a seminal paper (Milyo and Groseclose 2005) and web resources such as *http://www.left-right.us/about.html* in order to use them as a gold standard.

Leaning	News media sources
Left	nytimes, washingtonpost, nprnews, nightline, theearlyshow, nprscottsimon, davidgregory, ariannahuff, terrymoran, jdickerson, maddow, nprpolitics, todayshow, huffingtonpost,
Center	andersoncooper, cnnbrk, richardpbacon, jackgraycnn, GMA
Right	foxnews, washtimes, usnews, chicagotribune

Table 1: Political leaning of news media sources

Then we obtained all follow links to media sources and corresponding tweets. The resulting dataset includes 24 media sources that have a total of 7,782,104 subscribers. Some media sources were extremely popular and had millions of followers like the New York Times (1,755,740) while other media sources have fewer followers, e.g., NPR News (116,834), Fox News (100,272), and U.S. News (4,747). Among all subscribers of those 24 media sources, we only considered active users for the analysis by filtering out users having less than 10 tweets for last three months.

Generating an ideological map

We present a novel but preliminary algorithm that generates an ideological map of media sources through Twitter network. The basic idea is to determine a position of one media source on a one dimensional space by considering its distances to other media sources. The distance between them may be inferred from their co-subscribers. Hence there are two major parts in this algorithm; how to measure a distance between two media sources and how to align them in a line.

We previously proposed a measure of *closeness* between two media sources (An et al. 2011). There, we calculated the fraction of common audience. The intuition behind this is the closer two media sources are the more their audiences overlap. While this metric tells us about a relative distance to other media source, it does not yield an actual distance of two media sources given one dimensional space.

Let A represent the media of interest and $\{B_1, B_2, \cdots, B_n\}$ be the set of n other media sources for which we would like to measure the closeness from A. Then, the closeness value of A and B_i is defined as:

$$\mathrm{c}(A, B_i) = \frac{|A \cap B_i|}{|A \cup B_i|} \tag{1}$$

The distance value of A and B_i in an one dimensional space S is defined as:

$$\mathrm{d}(A, B_i) = k \cdot \left(\frac{\mathrm{c}(A, B_i)}{\sum_{j=1}^{n} \mathrm{c}(A, B_j)} \right)^{-1} \tag{2}$$

where k is a constant value determined based on the given space S.

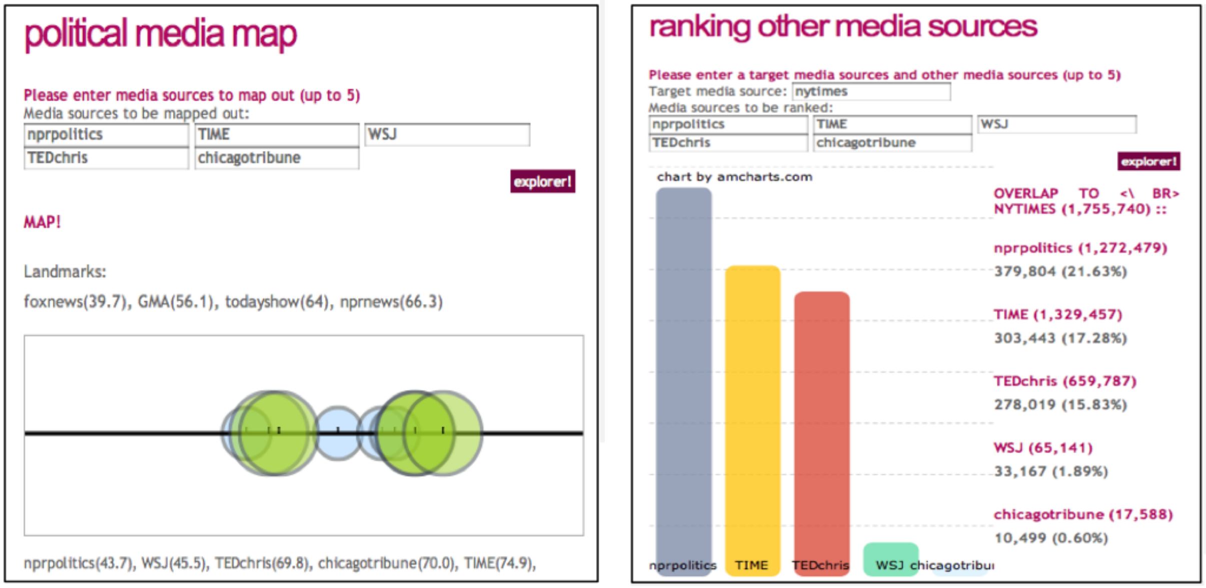
To predict a position of a media source on a given space, we apply a global network positioning (GNP) algorithm (Ng and Zhang 2002). GNP is a peer-to-peer and coordinates-based approach that models the Internet as a geometric space. It characterizes the position of any host in that space with a set of geometric coordinates. In GNP, a small set of hosts called Landmarks firstly compute their own coordinates in a chosen geometric space (e.g., a 1-dimensional Euclidean space), then any remaining host computes its own coordinates relative to the Landmarks. By considering a media source as a host, we are able to apply the GNP algorithm to determine coordinates of media sources on 1-dimensional Euclidean space.

Given the coordinates of the N Landmarks L_i in the geometric space S, each media source now derives its own coordinates. To do so, a media source M measures the distances to those Landmarks L_i with $Eq.(2)$ (ideal media-to-Landmark distances - $d(M, L_i)$). By minimizing the overall error between ideal $d(M, L_i)$ and euclidean distances, M can determine its own coordinates. For error measurement function, we take the mean squared error (MSE).

Results

We implemented our proposed algorithm and developed a web-based application that visualizes a political spectrum of various media sources in Twitter at *http://bit.ly/mediaexplorer*. As we mentioned before, our algorithm requires Landmarks and their coordinates. Rather computing their coordinates in advance, we used a well-known media bias measure, named ADA (Americans for Democratic Action) score, which is calculated based on various quantities such as the number of times a media outlet cites various think-tanks and other policy groups (Milyo and Groseclose 2005).

The ADA score is scaled from 0 to 100, where 0 means strongly conservative and 100, strongly liberal. For instance, ADA scores are 39.7 for Fox News and 73.7 for NYTimes.



(a) Political dichotomous map

(b) Ranking media sources

Figure 1: Screenshots of the Media Explorer webpage (*http://bit.ly/mediaexplorer*)

Our application uses the same scale with ADA score. Out of 18 news sources reported in (Milyo and Groseclose 2005), we used four of them as Landmarks for our application; Fox News (39.7), GMA(56.1), Today Show (64), and NPR News (66.3), which are known to have not changed their political leanings since their scores have reported.

The application shows the political coordinates of few other media sources (e.g., NPR Politics, TIME, WSJ, TED Chris, Chicago Tribunes) on the map depicted as green circles (Figure 1(a)). The blue circles on the map are those four Landmarks. Note that an ideological map of any news media sources can be generated automatically from our proposed methodology on inferring media bias while only 18 media sources have been examined in (Milyo and Groseclose 2005). Our application also shows, for a given media outlet, the list for the most similar media outlets based on the closeness measure. Figure 1(b) shows an example result we obtained for New York Times, where NPR Politics, TIME, and TED Chris come out as the top three closest media.

Finally, we test the effectiveness of our algorithm through comparison between our predicted positions and that in ADA's list. Out of 18 news sources with ADA scores reported in (Milyo and Groseclose 2005), we found 10 of them in our dataset. We used two of them as Landmarks of our algorithm, Fox News (39.7) and Today Show (64), and positioned 8 media sources remained on a one dimensional space scaled from 0 to 100. In the future, we will explore automatic ways of selecting the optimal (number of) Landmarks. For now, we select those two media sources simply because they belong to opposite sides of the political spectrum.

We measure how well ADA's list and our algorithm's list (Figure 2) are correlated. To do so, we opt for two widely-used correlation measures: Spearman's Rank Order correlation (r_s) and Pearson product-moment correlation coefficient (r_p). We find high correlations between the two lists: correlation coefficients are as high as .44 ($p > 0.1$) for r_s and .51 ($p > 0.1$) for r_p (their statistical significance is low simply because the number of media sources in the lists is low - it is 8).

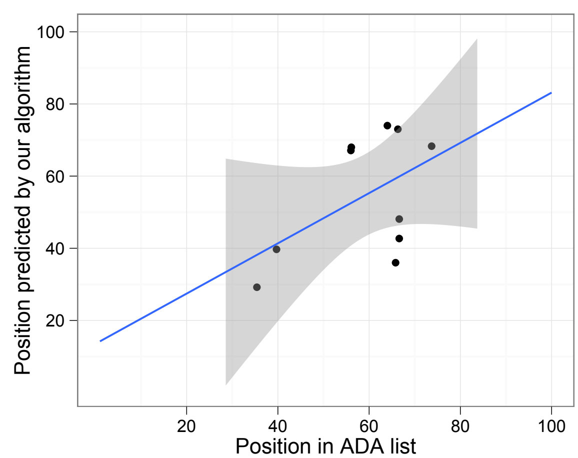


Figure 2: Comparison between our predicted position and that in ADA's list. The grey band includes one standard error of the prediction.

A pictorial political map of those 10 media sources is shown in Figure 3 along with their coordinates. We observed a strong tendency of known political dichotomy where NPR News and New York Times, which are known to be left-slanted, are positioned to one side and Washington Times, Fox News, and U.S News, which are known to be right-slanted, are positioned on the other side. However we also

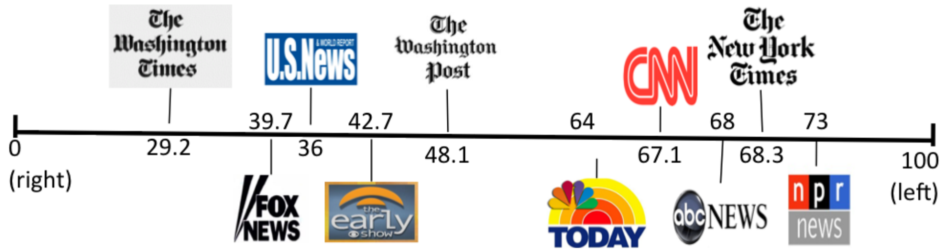


Figure 3: An ideological map of 10 news media sources. The coordinate of each media source is given along with its name.

found a few exceptions; Washington Post and Washington Times, known to have conflicting political preferences, lined up close to each other – possibly due to regional proximity.

Related work

It is known that republicans and democrats read different newspapers and books and geographically sort themselves by choosing to live in completely different areas (Bishop 2008). Such media slant is important because it can change people's beliefs in, for example, who they should vote for (Vigna and Kaplan 2007). Furthermore exposure to biased information can result into negative societal consequences such as intolerance of dissent, political segregation, and group polarization (Glynn et al. 1999).

Group polarization happens not only in the real world but also online. Blogs reflecting different political views rarely link to each other (Adamic and Glance 2005) and online news consumption is biased, much like offline consumption (Gentzkow and Shapiro 2011). To date, several studies have studied how people exchange political content in Twitter (Livne et al. 2011; Conover et al. 2011; Yardi and danah boyd 2010). This work builds upon these existing studies by extracting the overall media landscape from user activities in Twitter.

Conclusion

We proposed a novel algorithm that generates a political dichotomy map of media sources on Twitter, which is based on gathering online data and aggregating it via a closeness measure. The ideological map of a particular issue can be created in real time in conjunction with a public stream of tweets from Twitter. Extending this work, we are currently examining how news media sources of different political slants cover the same news story by conducting topic classification on news articles that are shared on Twitter.

Individuals might need to access to a pool of multiple points of view against which they can contrast their own values and belief as it helps them shape their eventual opinion. In the future, we hope to build a real-time platform that helps people receive balanced news information based on the model we proposed here. Nonetheless, we deemphasize the potential benefit of such political diversity because not everyone prefers to receive diverse political opinions (Munson and Resnick 2010). Hence different strategies are required to assist heterogeneous individuals when news aggregators plan to increase opinion diversity.

References

Adamic, L. A., and Glance, N. 2005. The political blogosphere and the 2004 U.S. election: Divided they blog. In *Proceedings of the ACM SIGKDD International Workshop on Link Discovery*.

An, J.; Cha, M.; Gummadi, K.; and Crowcroft, J. 2011. Media landscape in Twitter: A world of new conventions and political diversity. In *Proceedings of the ICWSM*.

Bishop, B. 2008. *The Big Sort: why the clustering of like-minded America is tearing us apart*. New York, New York: Houghton Mifflin Company.

Cha, M.; Haddadi, H.; Benevenuto, F.; and Gummadi, K. 2010. Measuring user influence in Twitter: The million follower fallacy. In *Proceedings of the ICWSM*.

Conover, M. D.; Ratkiewicz, J.; Francisco, M.; Goncalves, B.; Menczer, F.; and Flammini, A. 2011. Political polarization on Twitter. In *Proceedings of the ICWSM*.

Gentzkow, M., and Shapiro, J. M. 2010. What drives media slant? evidence from U.S. daily newspapers. *Econometrica Econometric Society* 78(1):35–71.

Gentzkow, M., and Shapiro, J. M. 2011. Ideological Segregation Online and Offline. *Quarterly Journal of Economics*.

Glynn, C. J.; Herbs, S.; OKeefe, G. J.; and Shapiro, R. Y. 1999. *Public Opinion*. Boulder CO: Westview Press.

Livne, A.; Simmons, M. P.; Adar, E.; and Adamic, L. 2011. The Party is Over Here: Structure and Content in the 2010 Election. In *Proceedings of the ICWSM*.

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

Munson, S., and Resnick, P. 2010. Presenting Diverse Political Opinions: How and How Much. In *Proceedings of the ACM CHI*.

Ng, T. S. E., and Zhang, H. 2002. Predicting internet network distance with coordinates-based approaches. In *Proceedings of the INFOCOM*.

PewResearch. 2004. The 2004 Political Landscape. *Washington, D.C.*

Vigna, S. D., and Kaplan, E. 2007. The Fox News effect: Media bias and voting. *Quarterly Journal of Economics*.

Yardi, S., and danah boyd. 2010. Dynamic Debates: An Analysis of Group Polarization over Time on Twitter. *Special Issue on Persistence and Change in Social Media. In Bulletin of Science, Technology and Society* 30(4).