

Identification and Analysis of Media Bias in News Articles

*Felix Hamborg¹, Norman Meuschke¹,
Akiko Aizawa², Bela Gipp¹*

¹University of Konstanz, Germany
[first.lastname]@uni-konstanz.de

²National Institute of Informatics, Tokyo, Japan
aizawa@nii.ac.jp

Abstract

Depending on the news source, a reader can be exposed to a different narrative and conflicting perceptions for the same event. Today, news aggregators help users cope with the large volume of news published daily. However, aggregators focus on presenting shared information, but do not expose the different perspectives from articles on same topics. Thus, users of such aggregators suffer from media bias, which is often implemented intentionally to influence public opinion. In this paper, we present NewsBird, an aggregator that presents shared and different information on topics. Currently, NewsBird reveals different perspectives on international news. Our system has led to insights about media bias and news analysis, which we use to propose approaches to be investigated in future research. Our vision is to provide a system that reveals media bias, and thus ultimately allows users to make their own judgement on the potential bias inherent in news.

Keywords: news aggregation; framing; content analysis; media bias

In: M. Gäde/V. Trkulja/V. Petras (Eds.): Everything Changes, Everything Stays the Same? Understanding Information Spaces. Proceedings of the 15th International Symposium of Information Science (ISI 2017), Berlin, 13th–15th March 2017. Glückstadt: Verlag Werner Hülsbusch, pp. 224–236.

1 Introduction

The Internet has allowed people to gather knowledge, form their own views, and engage in society (Mossberger, Tolbert & McNeal, 2007). Unrestricted access to unbiased information is crucial for understanding current topics. News articles are the primary source of such information, and thus of particular importance. However, the coverage of media outlets often exhibits a bias, commonly called *media bias*, e.g., due to influencing factors such as the ownership and the source of income of the media outlet, as well as political interference, lobbying, or ideological focus (University of Michigan, 2014). Not only developing countries, but also developed nations are subject to media bias, e.g., in the USA six corporations control 90% of the media (Insider, 2014), which results in a high chance of media manipulation (Esser, Reinemann & Fan, 2001). Trust in media is at a historical low. Less than half of the readers in USA trust media and think it is objective (GALLUP, 2015). Table 1 shows different headlines of two news articles reporting on the same event from 7 November 2014, during the Ukraine crisis.

Table 1: Different headlines about the same event

Publisher	Headline
(CNBC, 2014)	Tank column crosses from Russia into Ukraine: Kiev military
(RT, 2014)	Moscow to Kiev: Stick to Minsk ceasefire, stop making false invasion claims

While Western media, such as CNBC, reported that Russian tanks crossed the Ukrainian border, Russian media, such as RT, primarily portrayed these reports as false claims or did not report about the event. The content and tone of the two articles differs just as strongly as the two headlines suggest. We assume that readers' perceptions of the actual event will differ significantly depending on which article they read.

Also, consider the following headlines of the *same article* published in the New York Times (NYT). The NYT author has covertly altered the article, including its headline, from mainly complimentary (first row) to criticizing back then U.S. president candidate Sanders (second row, which is online now).

Table 2: Different headlines of the same article

Publisher	Headline
(Steinhauer, 2016a)	Bernie Sanders Scored Victories for Years Via Legislative Side Doors
(Steinhauer, 2016b)	Via Legislative Side Doors, Bernie Sanders Won Modest Victories

The author also rewrote several sentences, which initially had a positive tone, to subsequently reflect a dismissive tone. For instance, they removed a complimentary quote on his “very successful strategy”, and replaced it with a paragraph criticizing Bernie Sanders.¹

Reading articles that portray different perspectives on an event helps broaden the narrow and distorted perspective induced by media bias. News aggregators, such as Google News, enable readers to quickly get an overview of the large news landscape. However, most readers only consult a small subset of the available set of sources (Newman, Levy & Nielsen, 2015). Reasons include the overwhelming number of sources, language barriers, or out of habit. Established systems provide no support for showing the different perspectives within related articles. Recent efforts aim to fill this gap and reduce the effects of media bias, but suffer from practical limitations, such as being fine-tuned to one news category, or relying heavily on user input (Munson, Zhou & Resnick, 2009; Park et al., 2009).

After giving an overview of media bias, news aggregation, and news analysis, we introduce research questions that are motivated by weaknesses of state-of-the-art systems to support users’ awareness of media bias and its mitigation. Afterwards, we discuss approaches that address these questions. Finally, we present *NewsBird*, an aggregator that focuses on identification and presentation of shared and different information to reduce media bias effects.

¹ The early version was preserved by the Wayback Machine (Murphy, Hashim & O’Connor, 2007) available at <https://archive.org/web/>.

2 Related work

One of the main effects of media bias is the change of people's awareness and perception of topics (Siemens, 2014), which becomes critical for public issues, such as elections (Bernhardt, Krasa & Polborn, 2008). Reasons for biased news coverage range from intentional bias, e.g., governments and companies influencing publishers in their favor (Besley & Prat, 2002), to rather unintentional bias, e.g., due to the influence of *news values* (Harcup & O'Neill, 2001). One popular example is the NYT's preference for U.S. president candidate Hillary Clinton, which may have resulted for instance in the changes illustrated in table 2. Due to the journalistic requirement of objectivity, media bias is typically rather inconspicuous (Park et al., 2009), e.g., bias is embodied through story selection or placement (front-page story vs. on another page) or word choice ("coalition forces" vs. "invasion forces").

The workflow of news aggregators typically includes article crawling and extraction, summarization of related articles, and visual presentation (Evans, Klavans & McKeown, 2004). However, none of the established systems focuses on revealing differences among related articles (Park et al., 2009). Thus, their users may remain unaware of media bias and possibly different perspectives (Bui, 2010; Wanta, Golan & Lee, 2004).

Recent research efforts have attempted to fill this gap, and aim to reduce the effects of media bias. For instance, *NewsCube* divides a group of related articles using extracted keywords and hierarchical clustering. The resulting subgroups represent different semantic propositions of the main topic (Park et al., 2009). Another approach makes use of users' feedback, who set articles in relation to one another (Park et al., 2011). *Sidelines* counts links from pre-categorized blogs, e.g., liberal vs. conservative, pointing to an article to determine its political orientation. Afterwards, it selects politically opposing articles (Munson et al., 2009).

While these and other systems can reduce the effects of media bias by broadening readers' understanding of news topics, they suffer from several practical limitations. First, they are restricted to the analysis of one news category, i.e., politics (ibid.; Park et al., 2009). Additionally, some of them rely on manually built knowledge bases (Munson et al., 2009; Park et al., 2011).

A major reason for the difficulty of aggregators to show the effects of media bias is that current NLP methods perform especially poorly in identi-

fyng semantic differences in news (Park et al., 2009). In summary, classic NLP techniques that commonly rely on statistics “[...] are just a first step towards natural language understanding” (Cambria & White, 2014). This makes the analysis of news challenging, since semantic differences are often encoded subtly due to journalistic objectivity (Gauthier, 1993).

3 Research questions

This section derives the motivation and goals of our research from the findings of our literature review, which was briefly summarized in section 2. Afterwards, we discuss approaches to address our goals. While the proposals from this section are goals and future work, section 4 describes the current system.

3.1 Research motivation

The fundamental question that motivates our research is how a news analysis system can help users to become aware of media bias and thus reduce the effects of unidentified media bias. While news aggregators do not focus on exposing media bias, aggregators are still a first step in this direction given that they allow users to quickly get an overview of the news landscape. We envision a system that reuses news aggregation design and methods, but is additionally capable of the following:

- (T1) Reveal which news outlets portray which perspectives on what topics.
- (T2) Abstract and aggregate higher level knowledge from (T1), e.g., present possible factors that influence coverage. For instance, different interests in different countries or who is the owner of an outlet.
- (T3) To identify copy-editing and undocumented article changes: enable temporal and contextual analysis of topics to reveal which publishers are influenced by others and how.
- (T4) Provide a platform for users to collectively view and read different perspectives on topics, set publishers in relation, and judge their objectivity, currentness, etc. Such relations could express publishing (dis-)similarity or copy-editing flows.

3.2 News overview

Figure 1 depicts a visualization for a single topic that addresses (T1) from section 3. The upper section briefly summarizes the news topic similar to established news aggregators. The lower section of the view shows the most contrasting perspectives from articles on the selected topic. For geo-based news, such as international topics, a world map (right) visualizes from where these views originate. For other news categories, we will research further supporting views. Information related to (T2) will also be displayed in this view to help users identify the context of this topic and the presented perspectives, e.g., the news outlet that portrays a critical perspective on a green energy topic is owned by a company affiliated with the oil industry.

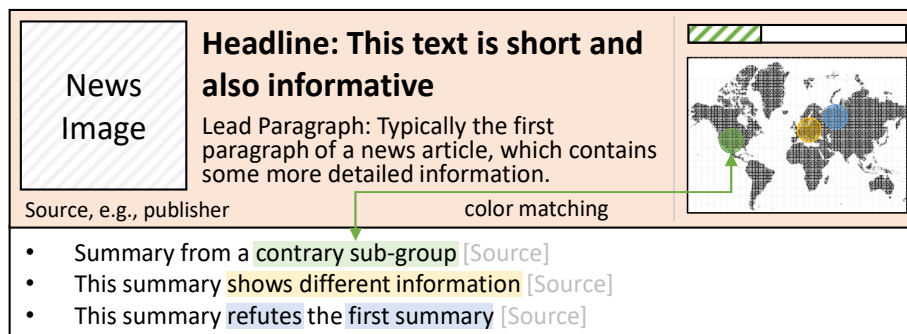


Fig. 1 News topic view for broadened news understanding

3.3 Analysis of news development and relations

We propose a second visualization depicted in figure 2 to analyze temporal and contextual relations between articles, such as if one article is mainly copy-edited from other previously published articles, and if yes, from which article, and which sections were copied or changed. The visualization also shows how an article's content and tone changes over time (cf. semi-transparent popup in fig. 2). This way, our system would help to identify cases of undocumented changes, as shown in table 2. To prove such changes, we plan to use decentralized trusted timestamping (Gipp, Meuschke & Gernandt, 2015).

The visualization in figure 2 addresses (T3). Such dependencies are not visible in current aggregators. The visualization would enable users to explore the origins of an article, i.e., from which an article may have been derived. Each article is represented with a colored box. Dependent articles, e.g., copy-edited from another, are connected with a line and colored in a similar shade.

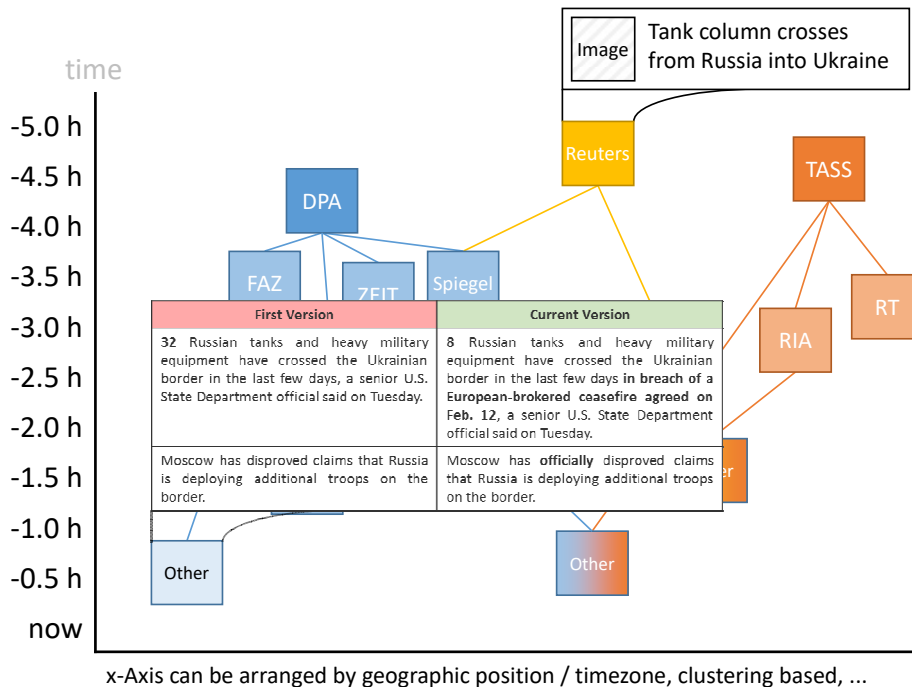


Fig. 2 Visualization for temporal and contextual analysis

3.4 General design of the news analysis

To acquire the information necessary for previously described visualizations, we propose a design that enables the system to analyze any news category, such as international news, finance, etc. While following the typical aggregator workflow (cf. sect. 2), we insert an additional task after topical grouping: *group refinement*. The goal of this task is to refine a group of related articles into subgroups that have a maximized expected diversity, i.e., represent different perspectives on the topic. Our hypothesis is that for each news category there is at least one *dimension* that achieves high diversity among resulting subgroups created by that dimension, e.g., in international news or conflicts often the involved countries have different perspectives on the topic as was demonstrated in table 1. We want to investigate for other categories which dimensions are effective in splitting articles into diverse subgroups.

Figure 3 depicts the system's group refinement workflow, which takes a group of related articles and finds subgroups that have a maximized expected diversity, estimated by the *d-score*. First, our system generates subgroup

candidates for each dimension. The system estimates the subgroups' diversity using the d-score, for which it uses user-provided relations and a method we will term *document similarity analysis* (DSA). DSA measures the semantic (dis-)similarity between two documents by comparing features such as extracted keywords, syntax, hyperlinks, images, and other elements common to news. We will gather user-provided relations directly in our system by facilitating such feedback (Park et al., 2011). Users can provide publisher and article relations, and quantified information on news, such as rate their objectivity and currentness (T4) (cf. *ibid.*). Finally, our system selects subgroups that yield the highest d-score. The results are then presented in one of the presented visualizations.

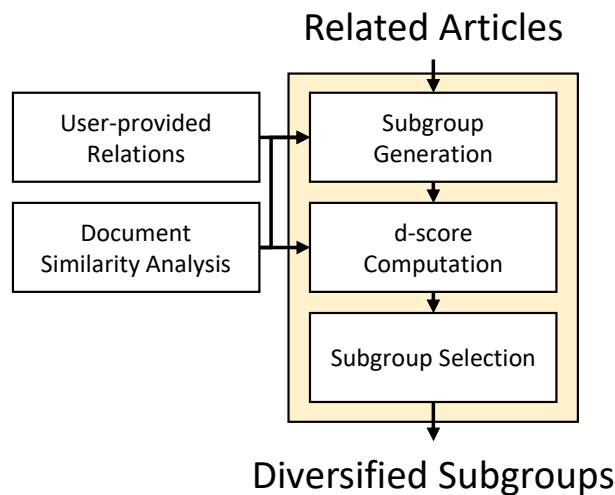


Fig. 3 Group refinement workflow

4 *NewsBird*: Different perspectives in international news

We developed *NewsBird* as a news aggregator that additionally addresses the goals described in section 3.1. Our aggregator shows both the shared and the opposing perspectives expressed in topically related news articles. *NewsBird* currently focuses only on international news, which we consider a prime example of media bias as displayed in table 1 and figure 4. We plan to general-

ize the design to support further news categories, and add visualizations that help users investigate the effects of media bias and get broad news understanding (cf. sect. 3). Therefore, we will use the findings we obtained during the development of the prototype and from an expert case study.

Figure 4 depicts NewsBird’s design, which is built around the idea of analyzing and visualizing news topics in a matrix to reveal their differences. In a onetime or regular process, the first two stages *data gathering* and *articles extraction* are performed to create or update the database by insertion of news articles. We used a non-disclosed dataset that contains 1.6 million articles gathered from almost 4,000 publishers from over 100 countries in October and November 2014. The dataset stems from the European Media Monitor (Atkinson & Goot, 2009). In future, we plan to use news-please to be able to aggregate and analyze articles reporting on further events (Hamborg et al., 2017).

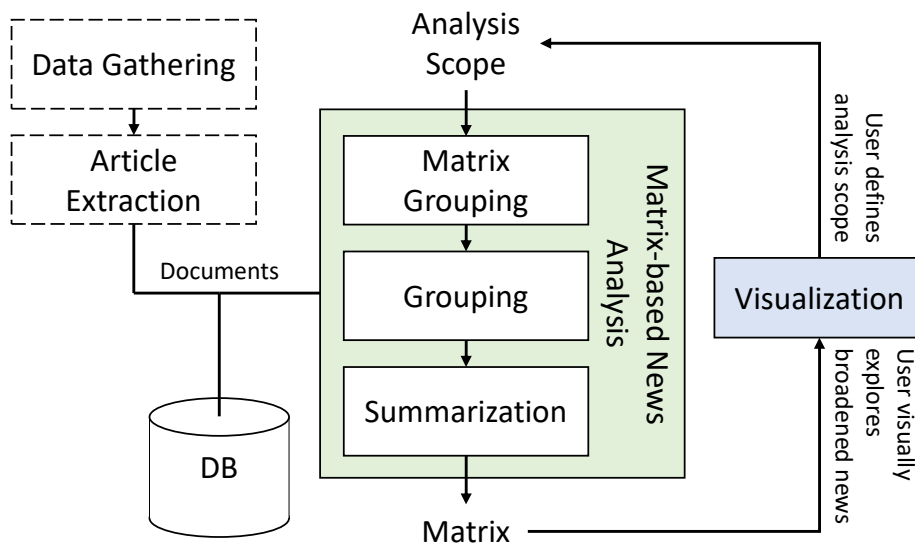


Fig. 4 Analysis design of NewsBird

First, in the *visualization* users define the *analysis scope*, which contains query date and two matrix dimensions including their specific values. To support the analysis of international news, NewsBird’s dimensions are the *publisher* and *mentioned country*. For instance, the cell of publisher country Russia and mentioned country Ukraine, hereafter denoted with RU-UA, contains all articles that have been published in Russia and mention Ukraine. This way, NewsBird reveals what is stated in which country about another

country, e.g., RU-UA contains articles a Russian news reader might read about the Ukraine.

The first analysis task is *matrix grouping*, which spans a matrix over both dimensions, creates corresponding cells, and queries articles for each cell. Afterwards, the *grouping* task finds groups of related articles, i.e., which report on the same topic. NewsBird finds these topics and subsequently calculates which articles are related by performing LDA on the articles across all matrix cells. The *summarization* generates summaries for 1) each topic and each cell 2) on all the cell's topics, and 3) for each of the topics of the cell. NewsBird uses the tokens' TF-IDF weights and cumulates them on sentence level.

Finally, NewsBird's *visualization* presents these analysis results to the user. Similar to established news aggregators, NewsBird allows users to get a news overview by showing them a list of topics and related articles. The key difference is our matrix visualization, which enables a comparative news exploration by showing both shared and opposing information in parallel as depicted in figure 5.

		Mentioned Countries			
		UA	RU	GB	DE
Publisher Countries	RU	Foreign Policy Adviser Says Russia Committed to Peace Process in East Ukraine	Ukraine Crisis, Sanctions Against Russia Not on G20 Agenda in Australia: Russian Sherpa	Cameron Says Britain Will Pay Only Half of \$2.6 Bln EU Surcharge	Berlin wall: the symbol of Cold War as an art object
	GB	Ukraine crisis: Kiev accuses Russia of military invasion after 'tanks cross border'	Tank column crosses from Russia into Ukraine – Kiev military	Cameron has warned there will be a „major problem“ if Brussels insists on Britain paying its \$2.6 bn	Fall of the Berlin Wall: „Our tears of frustration turned to those of joy“
	DE	Kyiv calls Berlin amid Russian incursion reports	Kyiv: 32 tanks enter Ukraine from Russia	Britain allowed to halve EU budget bill	Germany's east still lags behind
	US	Ukraine accuses Russia of sending in dozens of tanks	Ukraine accuses Russia of sending in dozens of tanks	Britain finds deal with EU over controversial bill	AP WAS THERE: The Berlin Wall crumbles

Fig. 5 Comparative news exploration in NewsBird

Figure 5 illustrates the efficacy of NewsBird's visualization on the exemplary topic of the Ukraine crisis from 7 November 2014, which was introduced in table 1. Note how the matrix reveals that media in the analyzed Western countries (rows) primarily reports that Russian tanks have invaded

Ukraine (in both Russia and Ukraine columns). However, the cell RU-UA shows that Russian media's most frequent topic involving the Ukraine is Russia's commitment to the peace process in the Ukraine.

We conducted an expert case study with one participant where we collected observational and think aloud data as well as information from eye gaze tracking and a follow-up interview to assess the subjective experience. Given the early stage of this project, we were mainly interested in whether users find NewsBird helpful to become aware of media bias and understand resulting differences in news coverage. We used three real-world news scenarios, including the Russian tank invasion on 7 November 2014, which we briefly introduced in section 1. We found that NewsBird enables users to get an overview of the current news situation similarly to established news aggregators. Furthermore, the user reported that using NewsBird he became aware of media bias and how this can lead to differences in news coverage for the first time.

5 Conclusion

In this paper, we introduced NewsBird, a news aggregator that aims to address practical limitations of established systems. NewsBird uses a two-dimensional matrix spanned by publisher and mentioned countries. This way, related articles are split into subgroups that have a large expected diversity and reveal differences in news coverage. While the current prototype is limited to the category of international news, our expert case study has shown that NewsBird already enables users to identify the effects of media bias and thus broaden users' news understanding. Furthermore, NewsBird has helped us identify directions for further research. The main goal will be the generalization of the current workflow to further news categories. We proposed multiple concepts for investigation, such as visualizations that reveal article and publisher dependencies, and a method to show temporal and contextual changes in news articles. To implement these aims, we will investigate and enhance NLP methods, e.g., to extract entities from news texts and measure article similarity by using both text and non-text features of news articles. This will help users in identifying and comparing media bias in current news topics, and thus ultimately reduce the effects of media bias in news coverage.

References

- Atkinson, M., and E. Van der Goot (2009): Near real time information mining in multilingual news. In: *Proceedings of the 18th international conference on World Wide Web* (pp. 1153–1154). ACM.
- Bernhardt, D., S. Krasa, and M. Polborn (2008): Political polarization and the electoral effects of media bias. In: *Journal of Public Economics*, 92 (5), 1092–1104.
- Besley, T. J., and A. Prat (2002): Handcuffs for the grabbing hand? Media capture and government accountability. CEPR Discussion Paper.
- Bui, C. (2010): How online gatekeepers guard our view: News portals' inclusion and ranking of media and events. In: *Global Media J.*, 9 (16), 1–41.
- Cambria, E., and B. White (2014): Jumping NLP curves: a review of natural language processing research. In: *Computational Intelligence Magazine, IEEE*, 9 (2), 48–57.
- CNBC (2014): Tank column crosses from Russia into Ukraine: Kiev military. <http://www.cnn.com/id/102155038>
- Esser, F., C. Reinemann, and D. Fan (2001): Spin Doctors in the United States, Great Britain, and Germany Metacommunication about Media Manipulation. In: *The Harvard International Journal of Press/Politics*, 6 (1), 16–45.
- Evans, D. K., J. L. Klavans, and K. R. McKeown (2004): Columbia newsblaster: multilingual news summarization on the Web. In: *Demonstration Papers at HLT-NAACL 2004* (pp. 1–4). ACL.
- GALLUP (2015): Americans' Trust in Media Remains at Historical Low. <http://www.gallup.com/poll/185927/americans-trust-media-remains-historical-low.aspx>
- Gauthier, G. (1993): In defence of a supposedly outdated notion: The range of application of journalistic objectivity. In: *Canadian Journal of communication*, 18 (4), 497–505.
- Gipp, B., Meuschke, N., & Gernandt, A. (2015): Decentralized Trusted Timestamping using the Crypto Currency Bitcoin. In: *Proceedings of the iConference 2015* (to appear). Newport Beach, CA, USA. <http://ischools.org/the-iconeference/>
- Hamborg, F., Meuschke, N., Breitingger, C., & Gipp, B. (2017): news-please: A Generic News Crawler and Extractor. In: M. Gäde, V. Trkulja and V. Petras (eds.): *Everything Changes, Everything Stays the Same? Understanding Information Spaces*. Proceedings of the 15th International Symposium of Information Science (ISI 2017), Berlin, 13th–15th March 2017. Glückstadt: Hülsbusch, pp. 218–223.
- Harcup, T., & O'Neill, D. (2001): What is news? Galtung and Ruge revisited. In: *Journalism studies*, 2 (2), 261–280.

- Insider, B. (2014): These 6 Corporations Control 90% of the Media in America. <http://www.businessinsider.com/these-6-corporations-control-90-of-the-media-in-america-2012-6>
- Mossberger, K., Tolbert, C. J., & McNeal, R. S. (2007): *Digital citizenship: The Internet, society, and participation*. MIT Press.
- Munson, S. A., Zhou, D. X., & Resnick, P. (2009): Sidelines: An Algorithm for Increasing Diversity in News and Opinion Aggregators. ICWSM.
- Newman, N., Levy, D. A., & Nielsen, R. K. (2015): Reuters Institute Digital News Report 2015. Available at SSRN 2619576.
- Park, S., Kang, S., Chung, S., & Song, J. (2009): NewsCube: delivering multiple aspects of news to mitigate media bias. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 443–452). ACM.
- Park, S., Ko, M., Kim, J., Choi, H., & Song, J. (2011): NewsCube 2.0: An Exploratory Design of a Social News Website for Media Bias Mitigation. In: *Workshop on Social Recommender Systems*.
- RT (2014): Moscow to Kiev: Stick to Minsk ceasefire, stop making false ‘invasion’ claims. <http://rt.com/news/203203-ukraine-russia-troops-border/>
- Siemens, G. (2014): Connectivism: A learning theory for the digital age. http://www.itdl.org/journal/jan_05/article01.htm
- Steinhauer, J. (2016a): Bernie Sanders Scored Victories for Years via Legislative Side Doors. <https://web.archive.org/web/20160314164825/http://www.nytimes.com/2016/03/15/us/politics/bernie-sanders-amendments.html?partner=rss&emc=rss>
- Steinhauer, J. (2016b): Via Legislative Side Doors, Bernie Sanders Won Modest Victories. <https://web.archive.org/web/20160314164825/http://www.nytimes.com/2016/03/15/us/politics/bernie-sanders-amendments.html?partner=rss&emc=rss>
- University of Michigan (2014): News Bias Explored – The art of reading the news. <http://umich.edu/~newsbias/index.html>
- Wanta, W., Golan, G., & Lee, C. (2004): Agenda setting and international news: Media influence on public perceptions of foreign nations. In: *Journalism & Mass Communication Quarterly*, 81 (2), 364–377.