

Article

Redefining Media Agendas: Topic Problematization in Online Reader Comments

Olessia Koltsova ^{1,*} and Oleg Nagornyy ²

¹ Laboratory for Internet Studies, National Research University Higher School of Economics, 192148 St. Petersburg, Russia; E-Mail: ekoltsova@hse.ru

² Qiwi Bank, 117648 Moscow, Russia; E-Mail: nagornyy.o@gmail.com

* Corresponding author

Submitted: 13 December 2018 | Accepted: 29 March 2019 | Published: 9 August 2019

Abstract

Media audiences representing a significant portion of the public in any given country may hold opinions on media-generated definitions of social problems which differ from those of media professionals. The proliferation of online reader comments not only makes such opinions available but also alters the process of agenda formation and problem definition in the public space. Based on a dataset of 33,877 news items and 258,121 comments from a sample of regional Russian newspapers we investigate readers' perceptions of social problems. We find that the volume of attention paid to issues or topics by the media and the importance of those issues for audiences, as judged by the number of their comments, diverge. Further, while the prevalence of general negative sentiment in comments accompanies such topics as disasters and accidents that are not perceived as social problems, a high level of sentiment polarization in comments does suggest issue problematization. It is also positively related to topic importance for the audience. Thus, instead of finding fixed social problem definitions in the reader comments, we observe the process of problem formation, where different points of view clash. These perceptions are not necessarily those expressed in media texts since the latter are predominantly "hard" news covering separate events, rather than trends or issues. As our research suggests, problematization emerges from readers' background knowledge, external experience, or values.

Keywords

audience; issue problematization; online media; reader; Russia; sentiment analysis; social problems; topic modeling

Issue

This article is part of the issue "Public Discussion in Russian Social Media", edited by Olessia Koltsova (Higher School of Economics, Russia) and Svetlana Bodrunova (St. Petersburg State University, Russia).

© 2019 by the authors; licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

Readers' comments on the news within online media are increasingly used both as a source of audience feedback by media organizations and as a new type of empirical data by media scholars. Unlike audience surveys and general public opinion polls, comment sections of media websites give users the ability to express themselves, in any form they choose, on issues they consider important. The consequences of such a new form of public expression for the processes of media agenda formation, set-

ting, and problematization has yet to be fully understood. Traditional poll-based studies in agenda setting find a vast array of evidence of the alignment between agenda salience in media and the importance of the respective issues for audiences (Scheufele & Tewksbury, 2006). At the same time, a theoretical tradition stemming from Stuart Hall's concepts of hegemonic and oppositional decoding of media messages (Hall, 1980) focuses on audiences' abilities to resist media frames within given agendas. Building upon the ideas from agenda-setting research, Hall's approach, and from certain theories of

social media, we can formulate a number of further assumptions. In particular, it is plausible, that through their comments, readers may be redefining the level of importance of agendas offered by media, as well as reframing some of them as problematic, thus altering or even subverting the professional definitions of the respective issues. As, according to constructivist approaches (Spector & Kitsuse, 1977), social problems are results of collective issue problematization, reader comments may actually serve as a source of popular perceptions of social problems.

In our research, we seek to extract these perceptions from large numbers of reader comments and to interpret them qualitatively. We investigate whether issue salience in professional media content and its importance for the commenting audience are aligned, whether general negative sentiment in comments or comment polarization indicates issue problematization by audiences, and what issues exactly are framed as social problems by readers. To do so, we focus on a set of regional online newspapers in Russia—a type of media that is loosely controlled so that discrepancies between media content and the audience's perceptions are likely to be manifest, but are unlikely to be suppressed. We apply a range of methods, from innovative automatic text mining to traditional qualitative text interpretation.

2. Social Problem Formation by the Public and the Media

A social problem can be most broadly defined as an undesirable situation or condition that characterizes society as a whole or some of its parts and that can be eliminated only if a collective effort is made. Early approaches to social problems define them as a social pathology that objectively undermines social health, and that should be diagnosed and treated (Smith, 1911), which means diagnosed by experts. Development of a more relativistic vision of social issues has led researchers to regard the public and its opinion as the source of problem definition. Still, it has been unclear how widely an issue has to be recognized as a problem to be considered as such by sociologists. Lauer (1976) observes that most of the time the number of people is to be “considerable”, while often the public is narrowed to “issue specific groups”, “strategically placed groups” or just “individuals and groups”, as in Spector and Kitsuse (1977).

Conceptualization of the role of media in the process of issue (de)problematization can be conventionally reduced to two major approaches. In the first, media are seen as powerful but not quite legitimate actors of problem definition. They are described as able to influence the public, a legitimate source of problem definition—both through agenda setting and framing (Scheufele & Tewksbury, 2006). The second, radically constructivist approach views media as an integral element of problem formation and existence. Thus, Hilgartner and Bosk (1988) view social problems as re-

sults of efforts of those who promote them to stay on the public agenda reinforced by inter-media linkages and constrained by the limited “carrying capacity” of various public arenas where such problems emerge. This approach makes it obvious that different factions of the public may hold different opinions on the existence and the features of social problems.

The advent of the Internet has added new layers of complexity to this picture (Zhou & Moy, 2007). Nowadays, public opinion is no longer locked within the datasets collected by pollsters. It has flooded into blogs, social network sites and, most importantly, to forums of online media. By commenting on media messages, readers get a chance to apply oppositional decoding, thus altering news framing, and influence the opinions of others (Kim, 2015; Lee & Jang, 2010), however, the evidence in this sphere is mixed and some studies have found that comments have no effect on reader opinions (Steinfeld, Samuel-Azran, & Lev-On, 2016). At the same time, numerous pieces of research have shown that social media can undermine media's monopoly over agenda setting (Sayre, Bode, Shah, Wilcox, & Shah, 2010; Wallsten, 2007), but it should be noted that this was demonstrated mostly for professional—or nearly professional—social media content. Lay comments are less influential and less likely to alter media agendas. However, since each comment influences its nearest neighbours and the amount of comments is huge, their cumulative capacity may be enough to affect public opinion directly, bypassing the agenda-setting cycle and the regular media. Thus, public perceptions of an issue's importance and their problematization of it may diverge from those set by media professionals without the latter being fully aware of it; this is especially plausible in societies with partially controlled media environments.

Of course, audiences' opinions, particularly those expressed in comments, do not necessarily mirror the distribution of opinions over the general population. It is known that commenters may differ from non-commenters by their psychological (Wu & Atkin, 2017) and demographic features (Stroud, Van Duyn, & Peacock, 2016). However, earlier research of letters to the editor that had dominated reader feedback before the Internet found that those letters tended to reflect public opinion, especially on salient issues, as found in McCluskey and Hmielowski (2012). Simultaneously, the authors found that comments present a wider range of opinions than traditional letters, and thus should be a better proxy for public opinion. But the most important point is that even though comments, just like regular media content, might be unrepresentative, they do play an important role in social problem definition and public opinion formation (Henrich & Holmes, 2013). In particular, critical comments have been found to shift opinions of readers away from the opinion expressed in media messages (Lee & Jang, 2010) and more generally to alter readers' opinions by influencing their perceptions of others' opinions about the messages (Kim, 2015). Thus, readers'

comments are not only important sources of knowledge about the audience's perception of existing social problems but also an instrument of social problem formation. It is also important that the agenda-setting, frame-building, problem-setting, and opinion-forming roles of comments have been found in different societies far beyond simply the "West", including such diverse countries as China (Zhou & Moy, 2007), South Korea (Lee & Jang, 2010), India (Jayachandran, 2015), as well as Post-Soviet countries (Toepfl & Litvinenko, 2018).

3. Approach

Our aim in this study is to identify which issues covered by media are important for audiences and which of them are problematized by the audiences, based on their comments. A traditional way to do so would be to perform a standard manual content analysis (Krippendorff, 2003). In this case assessors (coders) would have to answer the following questions: (1) what social issue is a given media text devoted to?; (2) is this issue problematized?; (3) does this issue attract much public attention in the form of comments; and (4) does a given comment problematize this issue? A severe limitation of this approach is that it does not scale to the volume of Internet content. We, therefore, employ a semi-automated approach that includes topic modeling, sentiment analysis, and a number of specially developed indices. Specifically, we extract agendas from news texts and compare their salience in news to the level of attention paid to them by the audience, as expressed in the volume of comments; we also determine the prevailing polarity of comments for each agenda item and the balance between negative and positive comments. Our methodological procedure is as follows.

First, media agendas or topics covered in media texts are discovered with topic modeling. Although, to our knowledge, media studies has mostly used supervised methods to detect topics (Flaounas et al., 2013; Scharkow, 2011), we apply an unsupervised approach (Blei, Ng, & Jordan, 2003) following a few examples of its application to media research (Koltsova & Shcherbak, 2015). The unsupervised approach reveals the latent topic structure that is not known to a researcher beforehand, while supervised approaches have to be guided by prior human knowledge submitted by researchers in various forms, such as keywords or manually labelled texts. As we are interested in finding all topics in a large media collection (not just a few of special interest), the unsupervised approach is a natural choice.

Second, based on topic modeling output, we calculate the salience of each topic in the entire collection of media texts thus elucidating media professionals' perception of newsworthiness. Third, based on knowledge of the proportion of each topic in each news item and the number of comments to each news item, we calculate an index of a topic's importance for the commenting audience. We then compare which topics are impor-

tant for media professionals, and which are important for readers.

Fourth, we apply a lexicon-based version of sentiment analysis (Thelwall, Buckley, Paltoglou, Cai, & Kappas, 2010) to detect the presence of both negative and positive sentiment in each comment, and then detect the prevalent sentiment in the comments related to each topic. We suggest that the prevalence of negative sentiment, especially in those topics which are important for audiences, is a first step to reveal problematized issues and therefore social problems. Although negative words may indicate grief or fear expressed in relation to an accident or a disaster (Thelwall, Buckley, & Paltoglou, 2011) and thus do not always indicate social problems, the latter, however, can hardly be discussed without some negative vocabulary. Therefore, although a researcher's judgement is needed to find social problems among other negative topics, still, this task only becomes feasible when thousands of texts are reduced to a hundred of topics of which only several dozen are negatively commented upon.

Fifth and finally, we calculate a polarization index of each topic—which indicates the ability of a topic to arouse controversial discussion among readers—by assessing the variance in the sentiment of comments related to a given topic. While prevailing negative sentiment may be the first signal of a problem, it may also—albeit not always—indicate an issue whose problematic character is a matter of consensus and is not questioned. However, sentiment polarization may indicate a lack of consensus on whether the issue should be treated positively or negatively, that is, whether it is a problem or not.

Although all proposed indices are only proxies of a latent phenomenon called "problematized issue", they are the only ways to conduct such research at scale. Furthermore, armed with the knowledge about most typical texts in each topic, we supplement our quantitative research with a qualitative examination of readers' comments related to all topics of interest. We use a set of simple interpretative procedures developed in our earlier research (Koltsova, 2011). Most broadly, these procedures constitute a type of discourse or frame analysis of news adapted for analysis of reader comments. It, in turn, builds upon some relatively old ideas of Van Dijk (1988) and Bell (1998) and alters them so as to be able to identify such structural elements of problem-oriented texts as blame attribution, victims, beneficiaries, other actors, problem causes and their consequences.

4. Hypotheses

Agenda-setting theory states that the more an issue is covered by media, the more important it will be for the public (McCombs & Reynolds, 2002). Furthermore, media professionals deliberately aim to satisfy the audience's interest which suggests that the amount of attention to a topic given both by the media and by the pub-

lic should be aligned. At the same time, some research finds that media professionals often have vague ideas about their audiences' demographic features, interests (Atkin, Burgoon, & Burgoon, 1983), and the content of their comments (Ürper & Çevikel, 2014) which suggests that the alignment of media and public attention still needs further analysis. Our first hypothesis may thus be formulated as follows:

H1: Topic salience in media texts will be positively related to topic importance for the audience.

Previous research (Liu, Zhou, & Zhao, 2015; Weber, 2014) also finds that negative news items (e.g., those addressing damage rather than success) get more comments than positive ones. To our knowledge, there is no research exploring the relationship between the volume of comments and their sentiment which would reveal whether readers are more inclined to leave negative comments. Ksiazek, Peer and Zivic (2015) show that, contrary to their assumptions, civil comments prevail over hateful comments. However, despite the possible overall prevalence of positive or neutral comments, the share of negative comments might be higher among highly commented—that is, more important—topics. Therefore, our second hypothesis is as follows:

H2: The more negatively a topic is perceived by the commenting audience, the higher its importance for the audience.

We have not found any research investigating whether the importance of an issue for the audience is related to a topic's being controversial. Such a finding would be highly novel, which leads us to our last hypothesis:

H3: The more controversy the topic arouses in readers' comments, the more important it is for readers.

5. Data

In this research, we study social problems at the regional level as we hypothesize that more specific problems may emerge at this level (McCombs & Funk, 2011). Also, we focus on Russia as a country in which the media are partially controlled as we expect that it is here that divergence between topic importance set by media and by readers should diverge most visibly. In Russia in 2014, the media that were most likely to combine content control and absence of comment moderation could be found mostly at the regional level. We use the data from the Omsk region, a typical Russian area in South Siberia. When ranked along with other Russian regions, it finds most of its important socio-demographic indices in the second quartile which includes population size (Russian Federal State Statistics Service, 2014a), per capita income (Russian Federal State Statistics Service, 2018), consumer expenditure, employment (Russian Federal

State Statistics Service, 2014b), and in the third quartile for ethnic diversity index (Timonin, Ryazancev, & Tikunova, 2011).

We define an Omsk media outlet as a website registered as such in the region, targeting only an Omsk audience and having a certain level of penetration—no less than 10,000 unique users per month. According to the Agency of Regional Research, the leading Omsk marketing organization, 18 such online outlets we registered in Omsk as of June 2014, the time closely preceding data collection. We focus on four of them that make up 65% of all the visits: BK55 (27% of visits), Omsk-inform (16.7%), Gorod55 (12%), and NGS Omsk (9.2%). Since Omsk regional media are not polarized politically, we believe that audience size is a sufficient selection criterion. On the whole, we follow a procedure typical for sampling media organizations in media studies.

Our sample thus includes all news items and respective comments from Gorod55 (6,302 news items; 67,195 comments), BK55 (14,078 items; 120,015 comments), NGS Omsk (4,780 items; 47,231 comments), and Omsk-inform (8,727 items; 23,666 comments) for the entire year from September 2013 to September 2014. The entire collection comprises 33,887 news items and 258,107 comments. On average, the four sources published 116 news items per working day and 33 news items per holiday. Distribution of comments per news item is, as expected, uneven, but not particularly skewed. Around 80% of news items (26,783) got at least one comment. The average number of comments per news item was 7.6, with Gorod55 taking the lead with 10.7 and Omsk-Inform lagging behind with only 2.7. It thus can be seen that the production leader and the feedback leader are not the same organization. The subsample for qualitative analysis includes 50 news items with the highest probabilities from each of 50 identified topics and all related comments (2500 news texts and 17,390 comments in total).

6. Topics and Their Salience in News

To extract topics from news texts, we used the Gensim software (Řehůřek, 2010) implementing the Latent Dirichlet Allocation (LDA) algorithm (Blei et al., 2003). While LDA limitations are discussed elsewhere (Maier et al., 2018), we employed a metric by Arun, Suresh, Madhavan and Murthy (2010) to overcome one of them, namely to avoid an arbitrary choice of the number of topics. Having obtained 19 topic solutions in increments of 5 ranging from 5 to 100 topics, we found several minima of Arun's et al. (2010) measure and selected one of them (which corresponds to 50 topics) based on manual topic assessment.

Topic modeling may be viewed as a type of fuzzy clustering which clusters both words and documents into a given number of topics so that each word and each text may occur in multiple topics. The search for clusters is undertaken based solely on information about word co-occurrence in texts, and the only information

needed by the algorithm is the number of topics. The output of the algorithm is two tables which sort words and texts, respectively, by the probability of their belonging to each topic. Most probable words in a topic (“top words”) are useful for topic interpretation (e.g., “theatre, play, culture, festival” or “Russia, Ukraine, against, sanction”). In our case, the topics were labelled based both on top words and top texts, by two researchers who then agreed on the labels. LDA always yields a certain proportion of uninterpretable topics (e.g., those crystallized around pejoratives), but in our case, we obtained only one completely uninterpretable topic and four topics whose interpretation posed some difficulties (marked *). This is a sign of the high quality of both the data and the solutions.

By summing the probabilities of all the texts in a given topic, we obtained an index of the topic’s salience in the collection and thus assessed how widely this topic is covered by journalists compared to other issues. The list of topics sorted by this index is presented in Table 1. When interpreting it, one should bear in mind that due to topic modeling instability less salient topics may appear not in all solutions obtained from the same data with the same algorithm parameters. However, the more salient the topic, the more often it appears and the less its salience score fluctuates. Also, since the calculation of other metrics is based on topic weights from the same so-

lution, the relationship between topic salience and other metrics should not be distorted.

As expected, local and regional topics outnumber national and international topics. The most salient topics are also regional, however, topic 6 due to the fact it relates to the Ukrainian crisis stands particularly high in this list, too. Not surprisingly, accidents and disasters also occupy the three top positions. The topics can be divided into event-driven and issue-driven (including problem-driven). Since our data is news, issue-driven topics do not necessarily contain texts that discuss general trends, causes or consequences of social phenomena as such; they mostly group together events related to a certain issue or a problem, for instance, schooling or economic crime. Issue-driven topics outnumber those centred around a single event, which is natural because a single event has to be as salient as an entire “issue” to compete to be detected as a separate topic. The Sochi Olympic Games has the highest salience among event-driven topics, however, it is the only non-regional topic in this category. Two resonant local events appear in this list: the resignation and subsequent arrest of the Omsk vice-mayor Yury Gamburg, and the murder of the locally famous boxer Ivan Klimov. In relation to the latter, many news items and comments share a belief in that his murder was connected to his conflict with the local “Gipsy Baron” Yan Lebedovoy and was perhaps committed or

Table 1. Topics ranked by salience (max = 0.0478; min = 0.0032; mean = 0.0198; stdev = 0.0315).

1. Car accidents	25. Schools, orphanages & child charit
2. Criminal news	26. Urban events & openings
3. Fires	27. Theatres & festivals
4. Local authorities: appointments, resignations & statements 1	28. Beauty contests & their winners
5. Regional taxes & fuel prices	29. Real estate: construction
6. Russia, Ukraine & US international relations	30. Control & regulation of enterprises
7. Trials on economic crime	31. Macroeconomic events: currency rates & oil prices
8. Sport: hockey	32. Holidays & VIP weddings
9. *Urban development (misc)	33. Movies and movie stars
10. Omsk region industrial development	34. Housing: heating
11. Public transport and traffic	35. Ads of banking services
12. Urban landscaping & greening	36. Libraries, literature & art
13. Regional parliament activities	37. *Regional elections and misc.
14. Accidents with children	38. Yury Gamburg resignation
15. Olympic Games 2014 & Omsk athletes	39. Missing person announcements
16. Police actions drug, alcohol & counterfeit money crimes	40. Street & bridge reconstruction & maintenance
17. Local authorities: appointments, resignations & statements 2	41. Concerts
18. Stray dogs & dog hunters	42. Hockey
19. Weather	43. Ivan Klimov’s murder
20. Abridgements of traffic law	44. Housing & the case of disabled Akhmetov
21. Arbitration court and the Mostovik case	45. Car sales
22. Urban demography & housing payments	46. *Military holidays
23. Education	47. *Uninterpretable
24. IT & military high tech	48. *Omsk media, plants & animals
	49. Crimea accession
	50. *NATO warships in the Black Sea, Russian rocket launch & contests

ordered by someone from the local Roma community. Neither event seems to be very important when judged solely by its salience in the collection. However, further analysis of the comments brings both of them back to our attention.

7. Importance of News Topics for Readers

News item importance for (or popularity among) readers is usually measured through the number of views or clicks, but clicks contain no data for further analysis of the sentiment of readers' feedback. This can be most easily assessed through the number of comments (Shoemaker, Johnson, Seo, & Wang, 2010). However, a comment is an attribute of a news item, and no news item belongs to any one single topic entirely, which is why we propose an index of topic importance:

$$importance_t = \frac{\sum_{d=1}^D prob_{dt} \times qcomments_d}{prob_t}, \quad (1)$$

where $importance_t$ —importance index of topic t , D —number of documents (texts), d —document index, $prob_{dt}$ —probability of topic t in document d , $qcomments_d$ —number of comments on document d and $prob_t$ is a salience of topic t . This formula normalizes the absolute volume of topic-related comments by the mean probability of a topic in all texts to penalize the influence of the long tail of a topic's distribution over texts. Also, for better representation, we normalize all the importance scores which were obtained to the range 0–100%: $importance_norm_t = importance_t / max(importance)$. A potential limitation of this metric is that it does not take into account the number of unique commentators involved in the discussion. Following Ksiazek, Peer and Lessard (2016) who differentiate between user-content and user-user types of interaction in commenting, it can be assumed that some lengthy comment threads may, in fact, be discussions among a few readers who, additionally, may diverge from the topic of the news item. Unfortunately, the data on commenters' identity is unavailable in our dataset.

Ten most commented and ten least commented topics are presented in Table 2.

It would be logical to expect that readers turn to regional and local media for regional or local topics. However, the leading topic in our importance list is centred around the relations between Russia, Ukraine and the US in the context of the Ukrainian crisis. As we saw before, this topic is also one of the most salient. The latter fact could have been explained by the bias of local journalists who could be copying the propagandistic agenda of the national media. Nevertheless, this topic is much discussed. A shift of audience's preferences from local to national to international news had already been noticed a few decades ago (Bogart, 1989), especially among younger, urban, male, and more educated consumers. A more recent study has found that online audiences consume more national and international news than traditional audiences (Tewksbury, 2003), which means that our findings are broadly consistent with the existing trends detected by other methods.

Another observation concerns the two local topic-forming events—Gamburg's arrest and Klimov's murder—both of which were much discussed, despite being moderately covered. Here, we can assume the media's intention to play down important local issues. Overall, there is no correlation between topic salience in texts and topic importance to readers (Spearman's rank correlation = 0.076, p-value = 0.598). This means that H1 has to be rejected. This goes against the main claim of the agenda-setting theory, however, it is in line with the recent research on the ability of user texts to influence public agendas. Although unlike autonomous social media posts studied e.g., by Wallsten (2007), news comments cannot generate entirely new agendas, they can signal which issues constitute an agenda in audiences' opinions.

8. Negatively and Positively Perceived Topics and Issue Problematization

Sentiment analysis is a methodologically difficult task, and currently, these methods mostly take the form of simple trinary classification into neutral messages and messages containing either generally negative or generally positive sentiment, without detecting the objects of

Table 2. Topic importance for readers (max = 100%; min = 20.7%).

Ten most commented topics	Ten least commented topics
Russia, Ukraine and the US international relations	Housing: heating
Stray dogs & dog hunters	Movies and movie stars
Yury Gamburg resignation	Urban events & openings
Local authorities: appointments, resignations & statements 1	Olympic Games 2014 & Omsk athletes
*Urban development (misc)	Education
Abridgements of traffic law	Concerts
Libraries, literature & art	*Regional elections and misc.
Crimea accession	Fires
Ivan Klimov's murder	Ads of banking services
Regional taxes & fuel prices	Car sales

these sentiments and without differentiating between different types of either negative or positive emotions. Accuracy of these methods is quite high for the English language but varies greatly for other languages which is currently an important limitation. Due to the scarcity of sentiment analysis instruments for the Russian language, we used PolSentiLex lexicon as it was the only one available at the time of the data analysis. Its quality was later shown to be comparable both to a more recent lexicon and to some other approaches (Alexeeva, Koltcov, & Koltsova, 2015). A lexicon is a list of words each of which is assigned a sentiment score by multiple assessors (coders) which may vary from negative to positive along a predefined scale. We submitted PolSentiLex lexicon into the well-known SA freeware SentiStrength (Thelwall et al., 2010) and calculated the overall sentiment score of each comment. This was done by averaging the negative and positive sentiment text scores each of which was equalled to the sentiment score of either the most negative or the most positive word in the text, respectively.

The aggregated sentiment scores of all comments related to a given topic was computed similarly to the topic's importance score, but instead of the number of comments the formula contains the mean sentiment score of all comments on a given news item:

$$sentiment_t = \frac{\sum_{d=1}^D prob_{dt} \times sent_d}{prob_t} \quad (2)$$

This index was also normalized to the range of 0–100%.

The ten most positively commented topics and the ten most negatively commented topics are presented in Table 3.

It is not surprising that the most positive emotions are related to holidays, entertainment, sports and culture. At the negative end, all topics except Klimov's murder are related to disasters and crimes framed as separate cases. They arouse general negative emotions, such as grief, fear, and anger, without bringing audiences to problem definitions. Klimov's murder, however, arouses heated discussions around police inaction, its alleged corrupt links to the Roma community accused of organized

crime, and emotional attempts to defend Roma people from ethnic hate speech.

A major observation is that importance scores and sentiment scores do not correlate (Spearman's rank correlation = 0.131, p-value = 0.365); that is, H2 is not supported. We find all types of topics: important positive (Crimea accession), important negative (Ivan Klimov's murder), unimportant positive (weather), and unimportant negative (fires). Moreover, positive comments prevail over negative ones (26% against 8%) which is in line with the findings of Ksiazek et al. (2015) about the prevalence of civil comments over hateful ones.

It is also clear that the negativity of comments is not a sufficient indicator of social problems. However, as all definitions of social problems state that a problem is an issue relatively widely perceived as a problem, we calculate an integral additive index reflecting both a topic's negativity and its importance for audiences. Table 4 shows topics that are both most important and most negatively commented. Those of them that are still event-driven (accidents and crimes) are now less common, and quite a number of others, marked with italics, easily reveal their problem-driven nature after a brief look at the comments. The stray dog topic reflects tensions between defenders of animal rights and those who try to "clear" the city by killing dogs to protect people, especially children, from the danger. The NATO warships topic is overwhelmed with hate speech towards "Americans" and is driven by Russia-US political tensions. Finally, the topic of Russia-Ukraine relations produces the most polarized and diverse comments that embrace the entire spectrum of audience's modes of decoding, as defined by Hall (1980)—from hegemonic to oppositional. They can be grouped into the following types: (1) hostile to Ukrainians (prevailing); (2) hostile to Putin because of the war with Ukrainians; (3) hostile to Putin because of insufficient war with Ukrainians; (4) supportive of Putin; (5) hostile to separatists in the Eastern Ukraine; (6) supportive of separatists. While in the NATO warships topic most comments identify the source of the problem in a similar way, in the Ukrainian topic one can see competing definitions of the problem, which leads us further to calculate a polarization index.

Table 3. Prevailing sentiment of the comments (max = 100%; min = 22.74%).

Ten most positively commented topics	Ten least positively commented topics
Olympic Games 2014 & Omsk athletes	Abridgements of traffic law
Hockey	Ads of banking services
Beauty contests & their winners	Police actions drug, alcohol & counterfeit money crimes
Theatres & festivals	Accidents with children
Street & bridge reconstruction & maintenance	Missing person announcements
Sport, hockey	Car sales
Holidays & VIP weddings	Car accidents
Public transport and traffic	Fires
Weather	Criminal news
Local authorities: appointments, resignations & statements	Ivan Klimov's murder

Table 4. Top ten topics ranked jointly by negative sentiment and importance.

Topics	Joint index of sentiment and importance
<i>Ivan Klimov's murder</i>	144.56%
<i>Russia, Ukraine & US international relations</i>	137.62%
<i>Stray dogs & dog hunters</i>	130.30%
Criminal news	124.61%
Abridgements of traffic law	117.89%
Yury Gamburg resignation	113.70%
Car accidents	108.62%
<i>*NATO warships in the Black Sea, Russian rocket launch & contests</i>	107.74%
Accidents with children	100.98%
<i>Macroeconomic events: currency rates & oil prices</i>	99.03%

9. Topics' Polarity as Best Proxies for Social Problems

Controversy in user comments has been studied manually (Weber, 2014), through supervised machine learning (Mishne & Glance, 2006), and based on a combination of sentiment analysis and issue detection (Sriteja, Pandey, & Pudi, 2017), the latter being the most relevant approach for us. Thus, to evaluate how different the sentiment of comments is, we choose standard deviation as a common measure of variation in data and obtain the following formula:

$$polarisation_t = \frac{\sum_{d=1}^D prob_{dt} \times std_d}{prob_t}. \quad (3)$$

In this equation, std_d is the standard deviation of the comments' sentiment scores in document d . We normalize the index to the range 0–100% as before.

Top ten topics ranked by polarization score are shown in Table 5. First, it can be seen that many of these topics are also among the most important ones (Spearman's rank correlation 0.877, p -value = 4.757e-16), and H3 is confirmed. Topics that arouse controversy attract more attention and more comments. Second, now most topics in this list are discussed as social problems—that is, the undesirable situations demanding collective effort (marked with italics). Three topics are related to local authority resignations and appointments, and they arouse

discussion about the fairness of such decisions and corruption. An unexpected case, the “Libraries, art and literature” topic is dominated by a story of a sixteen-year-old girl who was denied access to *The Financier* novel by Theodore Dreiser in one of the local libraries because the book contains scenes of sex. The heated discussions provoked by this story raise the problem of boundaries regarding what is permissible for children.

The Crimea topic is a most interesting case, but it also illustrates a limitation of topic modeling that sometimes brings together several interpretable, but unrelated subtopics. Although Crimea news items dominate this topic, three smaller clusters of news attracted the majority of comments that were both more negative and more polarized which makes this topic as a whole high in both negativity and polarization scores. However, the subtopic truly related to Crimea demonstrates a slightly different pattern. As suggested by manual analysis of comments to 21 Crimea-related news articles that appear among top 50 texts of this topic, the modest volume of Crimea-related comments might indicate readers' uncertainty and inability to forecast the consequences of such a tectonic shift in politics. Moreover, of 82 comments, we find only two that clearly frame the Crimea event as a matter of national pride. However, no extreme negative sentiment is observed in critical comments either, which contributes most to the overall positive tone of this subtopic. Critical comments' authors are mostly

Table 5. Ten most polarized topics.

Topic	Polarization index
<i>Stray dogs & dog hunters</i>	100.0%
<i>Local authorities: appointments, resignations & statements 1</i>	97.2%
<i>Yury Gamburg resignation</i>	95.3%
<i>Russia, Ukraine & US international relations</i>	94.8%
<i>Ivan Klimov's murder</i>	93.1%
Abridgements of traffic law	91.9%
<i>*Urban development (misc)</i>	91.4%
<i>Libraries, art and literature</i>	91.3%
<i>Crimea accession</i>	88.9%
<i>Local authorities: appointments, resignations & statements 2</i>	86.8%

concerned with the possible reallocation of public funds from their region to the newly acquired territory as well as with the anticipated increase in prices at Crimean resorts (both of which actually happened).

10. Conclusion

In this article we have examined which topics are important and which get defined as social problems by commenting audiences of regional media, focusing on a country with a partially controlled media. We have obtained several interrelated findings.

First, we have shown that the volume of attention to topics demonstrated by the media and by their audiences diverge. Possible causes of the divergence include: censorship or erroneous editorial policies that prevent audience's interests from being taken into consideration, difference between general and commenting audiences, and a chance that the volume of comments might measure a type of importance different from that captured by opinion polls (that are traditionally used to verify agenda-setting theory). In any case, this finding contradicts the main claim of the agenda-setting theory about the ability of media to tell the public "what to think about" (Cohen, 1963, p. 13). It is still an open question whether re-evaluation of topic importance by readers transfers directly into user-generated agenda-setting process, in particular, because the size of comment readership is less known than the size of news readership. As mentioned before, what has been studied so far is mostly the ability of social media content to influence the agenda of traditional media, not public opinion itself (as the classical agenda-setting research design would demand), furthermore this has been related to professional social media content only (Groshek & Groshek, 2013; Jang, Park, & Lee, 2017; Sayre et al., 2010). Also, re-evaluation of topic importance is not equal to autonomous agenda building as it only chooses between the offered issues. This process can be called agenda reweighting and it is closer to Hall's oppositional decoding (Hall, 1980), although, while Hall talks about an audience's redefinition of news frames, here we deal with the redefinition of news importance. New theories of agenda formation and spread are thus needed to account for the changed cycle and agents of media content production and dissemination.

Our second finding is that while the overall sentiment of comments does not correlate to a topic's importance for audiences, the level of sentiment polarization in those comments does. It means that commenting activity evolves there where discussion begins. This leads us to the third and most important finding. While the prevailing negative sentiment alone, and even combined with high topic importance, does not usually indicate topic problematization, high the level of sentiment polarization does. Although initially, we expected to find both problems whose problematic status is not challenged and those whose status is debated, through qualitative analysis of comments we instead found that problem for-

mulation takes place mostly when it is contested. That is, instead of the nomination of fixed problems we observe the process of their definition which potentially can contribute to the debate between approaches that view public opinion as either a static condition or a fluid process.

Closely connected to this is our fourth finding. Although the news mostly reports separate events and seldom formulate them as problems, we see that the media do not need to frame events as reflections of broader problems for audiences to problematize them. Readers make their own conclusions and generalizations bringing their background knowledge into the discussion—thus, although the news does not frame Klimov's murder as a potential ethnic conflict, many comment threads do. This finding is especially important in the context of Russian partially controlled media. This may be correct in regard to the national media and national agendas, as, indeed, we see a striking lack of discussion on Crimea and unanimous hate towards the US which is in line with the national media framing. However, at the regional level, many inconvenient questions are raised by readers that include corruption, police malpractice, censorship in literature, and ethnic tensions. The less problematization of such issues is found in media, the more important is the public availability of such discussions. This observation can have implications for the understanding of social problem definition and public opinion formation in semi-controlled environments. In such environments, governments are seldom able to control user-generated content as efficiently as they control professional content, which creates a visible gap between the two. This makes at least some segments of audiences consult non-institutionalized content for definitions of social problems thus creating alternative paths of public opinion formation.

Acknowledgments

This article is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University).

Conflict of Interests

The authors declare no conflict of interests.

References

- Alexeeva, S., Koltcov, S., & Koltsova, O. (2015). Linis-crowd.org: Leksichesky resurs dlya analiza tonalnosti socialno-politicheskikh textov na russkom yazyke [Linis-crowd.org: A lexical resource for sentiment analysis of sociopolitical texts in Russian]. In *Kompiuternaya lingvistika i vychislitelniye ontologii: Sbornik nauchnykh statey. Trudy XVIII obyedinennoy konferentsii "Internet i sovremennoye obschestvo"* [Computational linguistics and com-

- putational ontologies: A collection of scientific articles. Proceedings of the XVIII Joint Conference "The Internet and modern society" (pp. 25–34). Saint Petersburg: ITMO University. Retrieved from <http://openbooks.ifmo.ru/ru/file/2203/2203.pdf>
- Arun, R., Suresh, V., Madhavan, C. E. V., & Murthy, M. N. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. In M. J. Zaki, J. X. Yu, B. Ravindran, & V. Pudi (Eds.), *Advances in knowledge discovery and data mining* (pp. 391–402). Heidelberg: Springer. https://doi.org/10.1007/978-3-642-13657-3_43
- Atkin, C. K., Burgoon, J. K., & Burgoon, M. (1983). How journalists perceive the reading audience. *Newspaper Research Journal*, 4(2), 51–63. <https://doi.org/10.1177/073953298300400206>
- Bell, A. (1998). The discourse structure of news stories. In A. Bell & P. Garrett (Eds.), *Approaches to media discourse*. Oxford: Blackwell.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Bogart, L. (1989). *Press and public: Who reads what, when, where, and why in American newspapers*. Sussex: Psychology Press.
- Cohen, B. (1963). *The press and foreign policy*. New York, NY: Harcourt.
- Flaounas, I., Ali, O., Lansdall-Welfare, T., De Bie, T., Mosdell, N., Lewis, J., & Cristianini, N. (2013). Research methods in the age of digital journalism: Massive-scale automated analysis of news-content: Topics, style and gender. *Digital Journalism*, 1(1), 102–116. <https://doi.org/10.1080/21670811.2012.714928>
- Groshek, J., & Groshek, M. C. (2013). Agenda trending: Reciprocity and the predictive capacity of social networking sites in intermedia agenda setting across topics over time. *Media and Communication*, 1(1), 15–27. <https://doi.org/10.17645/mac.v1i1.71>
- Hall, S. (1980). Encoding/decoding. In *Culture, media, language: Working papers in cultural studies* (pp. 128–137). London: Hutchinson.
- Henrich, N., & Holmes, B. (2013). Web news readers comments: Towards developing a methodology for using on-line comments in social inquiry. *Journal of Media and Communication Studies*, 5(1), 1–4. <https://doi.org/10.5897/JMCS11.103>
- Hilgartner, S., & Bosk, C. L. (1988). The rise and fall of social problems: A public arenas model. *American Journal of Sociology*, 94(1), 53–78.
- Jang, S. M., Park, Y. J., & Lee, H. (2017). Round-trip agenda setting: Tracking the intermedia process over time in the ice bucket challenge. *Journalism*, 18(10), 1292–1308. <https://doi.org/10.1177/1464884916665405>
- Jayachandran, J. (2015). Outrage, debate or silence: An analysis of reader comments and online rape news. In N.-C. Schneider & F.-M. Titzmann (Eds.), *Studying youth, media, and gender in post-liberalisation India: Focus on and beyond the "Delhi gang rape"* (pp. 45–77). Berlin: Frank & Timme GmbH.
- Kim, Y. (2015). Exploring the effects of source credibility and others' comments on online news evaluation. *Electronic News*, 9(3), 160–176. <https://doi.org/10.1177/1931243115593318>
- Koltsova, O. (2011). Coverage of social problems in St. Petersburg press. In C. von Feilitzen & P. Petrov (Eds.), *Use and views of media in Sweden & Russia: A comparative study in St. Petersburg & Stockholm* (p. 325). Huddinge: Södertörns högskola.
- Koltsova, O., & Shcherbak, A. (2015). 'LiveJournal Libra!': The political blogosphere and voting preferences in Russia in 2011–2012. *New Media & Society*, 17(10), 1715–1732. <https://doi.org/10.1177/1461444814531875>
- Krippendorff, K. H. (2003). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.
- Ksiazek, T. B., Peer, L., & Lessard, K. (2016). User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments. *New Media & Society*, 18(3), 502–520. <https://doi.org/10.1177/1461444814545073>
- Ksiazek, T. B., Peer, L., & Zivic, A. (2015). Discussing the news. *Digital Journalism*, 3(6), 850–870. <https://doi.org/10.1080/21670811.2014.972079>
- Lauer, R. H. (1976). Defining social problems: Public and professional perspectives. *Social Problems*, 24(1), 122–130. <https://doi.org/10.2307/800329>
- Lee, E.-J., & Jang, Y. J. (2010). What do others' reactions to news on internet portal sites tell us? Effects of presentation format and readers' need for cognition on reality perception. *Communication Research*, 37(6), 825–846. <https://doi.org/10.1177/0093650210376189>
- Liu, Q., Zhou, M., & Zhao, X. (2015). Understanding news 2.0: A framework for explaining the number of comments from readers on online news. *Information & Management*, 52(7), 764–776. <https://doi.org/10.1016/j.im.2015.01.002>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., . . . Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2/3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- McCluskey, M., & Hmielowski, J. (2012). Opinion expression during social conflict: Comparing online reader comments and letters to the editor. *Journalism*, 13(3), 303–319. <https://doi.org/10.1177/1464884911421696>
- McCombs, M., & Funk, M. (2011). Shaping the agenda of local daily newspapers: A methodology merging the agenda setting and community structure perspectives. *Mass Communication and Society*, 14(6), 905–919. <https://doi.org/10.1080/15205436>

2011.615447

- McCombs, M., & Reynolds, A. (2002). News influence on our pictures of the world. In *Media effects: Advances in theory and research* (pp. 1–18). Mahwah, NJ: Lawrence Erlbaum.
- Mishne, G., & Glance, N. (2006). *Leave a reply: An analysis of weblog comments*. Paper presented at Third annual workshop on the weblogging ecosystem, Edinburgh, UK. Retrieved from <http://www.blogpulse.com/www2006-workshop/papers/www2006-blogcomments.pdf>
- Řehůřek, R. (2010). Software framework for topic modeling with large corpora. In *Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks* (pp. 45–50). Valletta: ELRA. Retrieved from <http://is.muni.cz/publication/884893/en>
- Russian Federal State Statistics Service. (2014a). *Chislennost naseleniya Rossiyskoy Federatsii po munitsipalnym obrazovanoyam na 1 yanvarya 2014 goda* [Population of the Russian Federation by municipalities as of January 1, 2014]. Russian Federal State Statistics Service. Retrieved from http://www.gks.ru/free_doc/doc_2014/bul_dr/mun_obr2014.rar
- Russian Federal State Statistics Service. (2014b). *Osnovnyye socialno-ekonomicheskiye pokazateli v 2014 g* [The main socio-economic indicators in 2014]. Retrieved from http://www.gks.ru/bgd/regl/B15_14p/lssWWW.exe/Stg/d01/01-01-1.doc
- Russian Federal State Statistics Service. (2018). *Srednedushevyye denezhnyye dohody po subyektam Rossiyskoy Federatsii 2013–2017 gg* [Per capita income in the regions of the Russian Federation 2013–2017]. Moscow: Russian Federal State Statistics Service Retrieved from http://www.gks.ru/free_doc/new_site/population/urov/2013-2017kv.xls
- Sayre, B., Bode, L., Shah, D., Wilcox, D., & Shah, C. (2010). Agenda setting in a digital age: Tracking attention to California Proposition 8 in social media, online news and conventional news. *Policy & Internet*, 2(2), 7–32. <https://doi.org/10.2202/1944-2866.1040>
- Scharkow, M. (2011). Thematic content analysis using supervised machine learning: An empirical evaluation using German online news. *Quality & Quantity*, 47(2), 761–773. <https://doi.org/10.1007/s11135-011-9545-7>
- Scheufele, D. A., & Tewksbury, D. (2006). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication*, 57(1), 9–20. <https://doi.org/10.1111/j.0021-9916.2007.00326.x>
- Shoemaker, P. J., Johnson, P. R., Seo, H., & Wang, X. (2010). Readers as gatekeepers of online news: Brazil, China, and the United States. *Brazilian Journalism Research*, 6(1). Retrieved from <https://bjr.sbpjor.org.br/bjr/article/view/226>
- Smith, S. G. (1911). *Social pathology*. New York, NY: Macmillan.
- Spector, M., & Kitsuse, J. (1977). *Constructing social problems*. Menlo Park, CA: Cummings.
- Sriteja, A., Pandey, P., & Pudi, V. (2017). Controversy detection using reactions on social media. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 884–889). Piscataway, NJ: Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/ICDMW.2017.121>
- Steinfeld, N., Samuel-Azran, T., & Lev-On, A. (2016). User comments and public opinion. *Computers in Human Behavior*, 61, 63–72. <https://doi.org/10.1016/j.chb.2016.03.004>
- Stroud, N. J., Van Duyn, E., & Peacock, C. (2016). *News commenters and news comment readers* (Engaging news project report). Retrieved from <https://mediaengagement.org/wp-content/uploads/2016/03/ENP-News-Commenters-and-Comment-Readers1.pdf>
- Tewksbury, D. (2003). What do Americans really want to know? Tracking the behavior of news readers on the internet. *Journal of Communication*, 53(4), 694–710. <https://doi.org/10.1111/j.1460-2466.2003.tb02918.x>
- Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in twitter events. *Journal of the Association for Information Science and Technology*, 62(2), 406–418. <https://doi.org/10.1002/asi.21462>
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544–2558. <https://doi.org/10.1002/asi.21416>
- Timonin S. A., Ryazancev, S. V., & Tikunova, I. N. (2011). Matematiko-kartograficheskaya otsenka etnicheskogo raznoobraziya v regionah Rosiyi [Mathematical-cartographic assessment of ethnic diversity in the regions of Russia]. In V. Tikunov (Ed.), *InterKarto/InterGIS 17: Ustoychivoye razitiye territoriy: teoriya GIS i prakticheskiy opyt* [InterCarto/InterGIS 17: Sustainable development of territories: GIS theory and practical experience] (pp. 15–21). Barnaul: Institute for Water and Environmental Problems. Retrieved from <http://www.iwep.ru/ru/bibl/books/matkonf/INTERCARTO17-1.pdf>
- Toepfl, F., & Litvinenko, A. (2018) Transferring control from the backend to the frontend: A comparison of the discourse architectures of comment sections on news websites across the post-Soviet world. *New Media & Society*, 20(8), 2844–2861.
- Ürper, D. Ç., & Çevikel, T. (2014). Reader comments on mainstream online newspapers in Turkey: Perceptions of web editors and moderators. *Communications*, 39(4), 483–503. <https://doi.org/10.1515/commun-2014-0022>
- Van Dijk, T. A. (1988). *News as discourse*. Hillsdale, NJ: Erlbaum.
- Wallsten, K. (2007). Agenda setting and the blogosphere: An analysis of the relationship between mainstream media and political blogs. *Review of Policy Research*, 24(6), 567–587. <https://doi.org/10.1111/>

[j.1541-1338.2007.00300.x](#)

Weber, P. (2014). Discussions in the comments section: Factors influencing participation and interactivity in online newspapers' reader comments. *New Media & Society*, 16(6), 941–957. <https://doi.org/10.1177/1461444813495165>

Wu, T.-Y., & Atkin, D. (2017). Online news discussions: Exploring the role of user personality and motivations

for posting comments on news. *Journalism & Mass Communication Quarterly*, 94(1), 61–80. <https://doi.org/10.1177/1077699016655754>

Zhou, Y., & Moy, P. (2007). Parsing framing processes: The interplay between online public opinion and media coverage. *Journal of Communication*, 57(1), 79–98. <https://doi.org/10.1111/j.0021-9916.2007.00330.x>

About the Authors



Olessia Koltsova is the Head of the Laboratory for Internet Studies at the National Research University Higher School of Economics, Russia. Holding a PhD in sociology, she advocates interdisciplinary problem-focused research based on rigorous analysis of big data. In the recent years, she has extensively published on new media's role for political participation, online communities and the methods of automated text analysis. She is also the author of *News Media and Power in Russia*, Routledge, 2006.



Oleg Nagorny graduated from the master program in Advanced Social Analysis at the National Research University Higher School of Economics (HSE). Currently, he is a data analyst at Qiwi Bank. Prior to that, he was a research intern at HSE Laboratory for Internet Studies and a PhD student in sociology at the same university.