Identifying the Topics of Russian Political Talk Shows

Alexander Petrov[a], Olga Proncheva[b]

[a]*Keldysh Institute of Applied Mathematics (Russian Academy of Sciences), Miusskaya sq., 4, Moscow, 125047, Russia*
[b]*Moscow Institute of Physics and Technology (National Research University), 9 Institutskiy per., Dolgoprudny, Moscow Region, 141701, Russian Federation*

Abstract

The paper studies the agenda by popular talk shows on TV in Russia over a more than three-years period in 2016-2019. Two major talk shows are considered, namely "Meeting Point" (NTV Channel) and "60 minutes" (Russia 1 Channel). Four long-run topics are considered, which are related to situation in Ukraine, Middle East (civil war in Syria, Russian-Turkey relations etc.), USA (including Russian-US relations), Europe (including EU-Russian relations, the migrant crisis and Brexit). Abstracts of each episode are taken from the broadcasters website as raw data for analysis, the Python code was developed in order to collect them and save in the txt format. Topic-related lists of key words are made to assign one or several topics to each episode. Basing on those lists, one or several topics were assigned to each episode. Making this for each episode of a given week, we obtain weekly scores for each topic; these scores being the numerical measures of the topics' saliency in the shows.

Keywords

media agenda, topic modeling, political talk show, long-run political topics, TV

1. Introduction

The agenda-setting theory by McCombs and Show [1] has been recognized as one the most important advances in political communication research. Its central idea is given in short form by Cohen [2] (quoted from [1]): the press "may not be successful much of the time in telling people what to think, but it is stunningly successful in telling its readers what to think about". Accordingly, agenda-setting research is in large part the study of what the mass media talking about.

Traditionally, the activity of the mass media is associated with telling the news, however, today we can see the rise of political talk shows on TV in Russia. This trend can be considered favorable for researchers simply because the whole idea of these shows is to focus on issues that the audience believes important.

In this paper we study empirically the content of two of the Russian political talk shows, namely "Meeting Point" and "60 minutes", which are generally known broadcasts with a large audience. For example, during the week 13 to 19 November 2017, according to Mediascope [3], their audiences were as follows:

- "Meeting Point": 2.3 rating points, 11.9 share points (14 November),

- "60 minutes": 4.5 rating points, 14.4 share points (14 November).

To our knowledge, there are no studies of agenda by talk shows.

2. Method

2.1. Method in Brief

The whole idea of our approach is as follows. We consider four topics that are widely recognized as main long-run topics of political discourse in Russia, and aim to estimate the proportion between them as a function of time.

A topic's saliency during a given week is characterized by its weekly score that is calculated, roughly speaking, as the number of episodes of all shows featuring the topic, adjusted to their duration and number of topics addressed in the episode. Technically, the central point in our work was the identification of the topic (or several topics) featured in a certain episode. In order to do that, we collected the abstracts of all episodes of two popular Russian Political Talk Shows, namely "Meeting Point" and "60 minutes" from more than three-years period. Four political topics often covered by political talk shows were in the focus. These topics are:

- UKR (Ukraine);

- MES (Middle East: civil war in Syria, Russian-Turkey relations etc.);

- USA (US politics and economy, Russian-US relations, NATO, Global security and relation between Russia and the West as a whole);

- EUR (European politics and economy, EU-Russian relations, The European migrant crisis, European Union, Brexit etc.).

There were, of course, episodes on other topics such as Far East (Russian-China and Rusian-Japan relations, economic performance of China etc.), Former Soviet Union states excluding Ukraine and EU members (their politics, economy, relations with Russia, USA and EU), Russian economy, Russian elections, conservative vs liberal values, historical legacy and so on. We did not consider these topics but took the corresponding episodes into account when counting the proportions of the weekly number of episodes.

For each of the four topics, the list of key words was made. Using these lists, each episode of the show was assigned one or several topics (for example, 0.5 UKR and 0.5 USA). After that the proportions of each topic in all episodes of the week were calculated.

Bearing in mind the perspective of studying the process of agenda-setting by these shows, we weighted different shows according to their typical duration thus getting the score of each topic in each show for each week. For instance, an episode of "Meeting Point" last somewhat longer that that of "60 Minutes", therefore their duration-related weights are 5 and 4, respectively. That is, the score of a topic in "Meeting Point" is its proportion multiplied by 5. The total score of a topic across all shows is the average of its show-specific scores. This total score reflects the proportion of the topic in media agenda.

The future plan is to compare dynamics of proportions with public attention to topics thus identifying the agenda-setting role of political talk shows.

2.2. Data Collection

Two generally known daily talk shows were considered, namely "60 minutes" (Russia 1 Channel) and "Meeting Point" (NTV Channel). Brief description of each episode of each show is available on the TV broadcaster's website. For both show, description of an episode typically has about 40 words (with nearly all descriptions falling between 25 and 60 words). An original Python code was developed in order to collect them and save in the txt format.

The collected data encompass the period of time from the first episode of each show (29 February 2016 for "Meeting Point" and 12 September 2016 for "60 minutes") to 18 April 2019.

These texts were used as raw data for the analysis.

2.3. Topic Scores

First, we assign each show a duration-related coefficient C_{show}. For example, C_{60min} = 4, $C_{MeetingPoint}$ = 5.

The topic score for each episode was assigned basing on the following algorithm. For each of the four topics specified above a list of keywords was developed. The English translation is given below. It should be noted that this English list should not be understood literally, because the correspondence between Russian and English words is not perfectly one-to-one. For example, the English word "Chinese" is translated into Russian differently in cases when it stands for a Chinese man and when it is an adjective (say, Chinese smartphone).

1. UKR - Ukraine, Maidan, Euromaidan, Kravchuk, Kuchma, Yushchenko, Yanukovych, Poroshenko, Turchinov, Yatsenyuk, Timoshenko, Zelensky, Paruby, Groisman, Akhmetov, Kolomoisky, Dobkin, Medvedchuk, Vakarchuk, Sentsov, Savchenko, Vyshinsky, Zakharchenko, Plotnitsky, Pushilin, Bolotova, Avakov, Shukhevych, Poltorak, Lyashko, Goncharuk, Bogdan, Filaret, Akhmetov, Firtash, Opposition platform, BYuT, Party of Regions, Regionals, Right Sector, Azov, Dobrobat, non-canonical, Kiev, Donetsk, Lugansk, Kharkov, Lviv, Vinnitsa, Ilovaysk, Debaltseve, Shakhtersk, Gorlovka, Novoazovsk, Mariupol, Zaporizhia, Chernobyl, Kherson, Dnepropetrovsk, Dnipro, Crimea, Donbass, Transcarpathia, Galicia, Southeast, DPR, LPR, certain areas, Kiev, Donetsk, Lugansk, Kharkov, Lviv, Vinnitsa, Ilovaisk, Debaltsevo, Kerch, Zaporizhzhya, Chernobyl (except the phrase "Chernobyl Way"), Svyatoshinsky, Kherson, Ukrainian, seiner Nord, Rada, Square, APU, ATO, Norman format, Norman format, Bandera, Russian aggression, Minsk agreements, Minsk process, Minsk-2, people's deputy, tomos, Steinmeier plan, Rosukrenergo, Ukrnafta, Naftogaz, Russian coal, Dneproges, Antonov concern, "Motor Sich", Ukrainian, Return + sailors.

2. MES - Turkey, Syria, Iran, Iraq, Golan Heights, Assad, Hamas, Hezbollah, Hezbollah, Defense Army, ISIS, Istanbul, Damascus, Aleppo, Deir ez-Zor, Palmyra, Idlib, Tehran, Baghdad, Basra.

3. USA - USA, NATO, Obama, Bush, Trump, Clinton, Cheney, Rumsfeld, McFaul, Huntsmann, Tillerson, Pompeo, Pence, Biden, Nuland, Miller, Washington, Pentagon, White

House, Department of State, Department of State, Secretary of State, Congress, Stratfor, Stoltenberg, START (note: the corresponding Russian abbreviation does not coincide with a common word), Missile Defense, inference + election, Skripal, neoconservatives, Assange, Venezuela, Congress, PMCs, Guantanamo, sanctions, transatlantic, disarmament, arms race.

4. EUR - EU, Britain, England, Germany, France, Hungary, Bulgaria, Austria, Spain, Italy, Belgium, Netherlands, Luxembourg, Denmark, Finland, Sweden, Greece, Cyprus, Ireland, Iceland, Poland, Czech Republic, Slovakia, Slovenia, Serbia, Croatia, Yugoslavia, Albania, Kosovo, Macedonia, Lithuania, Latvia, Estonia, Council of Europe, European Parliament, PACE, European Commission, Old World, European Commissioner, London, Salisbury, Scotland, England, Paris, Berlin, Edinburgh, Lisbon, Madrid, Brussels, Vienna, Rome, Warsaw, Barcelona, Catalonia, Wales, Prague, Amsterdam, Belgrade, Pristina, Stockholm, Helsinki, Copenhagen, Rompuy, Junker, Mogherini, European Commissioner, Nord Stream, Nord Stream, Tusk, Johnson, May (note: in Russian this surname does not coincide with a common word), Cameron, Faraj, Macron, Fillon, Hollande, Herro, Chancellor, Merkel, Tsipras, Orban, Kaczynski, Le Pen, Salvini, Steinmeier (except: Steinmeier plan), Grybauskaite, British, English, French, German, Lisbon, euro, migration crisis, Notre Dame, Bataclan, Charlie Hebdo.

Each abstract is processed using this list of keywords. The first stage is lemmatization (using morphological analyzer "pymorphy2" [4]), the we count the keywords from each of the topics' list, taking into account the following points.

First, it should be borne in mind that some of "keywords" are actually combinations of two consecutive words. For example, words "party" and "region" can appear in many contexts, but we count a keyword only if the appear in a row to form the phrase "Party of Regions", which clearly refers to the UKR topic.

Similarly, the words "Minsk", "agreement" and "process" are not keywords by themselves, however the phrases "Minsk Agreements" and "Minsk Process" also refer to the topic UKR.

If an abstract contains the words "Return + sailors" (not necessarily in a row, as in "return sailors home" or "return Russian citizens in exchange for Ukrainian sailors"), then we counted it in the topic UKR.

Having counted the keywords in the abstract, we can assign one or several topics to it. If all the keywords refer to a single topic, then we assign this topic to the abstract (that is, to the episode). If there are keywords from the several topics, we use the following algorithm.

Suppose $n > 1$ topics were addressed in the abstract. Denote by a_i, $i = 1, ..., n$ the number of keywords for the topic i.

If all $a_i \leq 1$, $i = 1, ..., n$, then put $b_i = a_i$ for $i = 1, ..., n$.

If there is at least one $a_i > 1$, then for each i, put $b_i = a_i$ if $a_i > 1$, otherwise $b_i = 0$.

The score S_i for topic i is calculated as $S_i = C_{show} b_i / \sum_{1}^{n} a_i$, $i = 1, ..., n$.

That is, if there is more than one keyword from one topic, and exactly one keyword from another topic, then the second topic's score is zero. If there are more than one word from several topics, then the score is divided between these topics proportionally to the number of words. If no topic has more than one keyword, and several topics have exactly one keyword, then the score is divided between then equally.

Table 1

Number of episodes by topic.

	Meeting Point	60 minutes
All Episodes	670	1100
UKR	309	576
MES	63	77
USA	257	378
EUR	198	217

Table 2

Number of episodes with positive score by topic

	Meeting Point	60 minutes
All Episodes	670	1100
UKR	273	540
MES	34	52
USA	195	282
EUR	142	158

Thus, each episode is assigned topic scores. Topic score for a week is the sum of topic scores for all episodes of the week.

3. Results

Some descriptive statistics are shown in Tables 1, 2. The difference between them is that the latter presents the number of episodes that added to the scores. For example, UKR topic was touched upon in 347 episodes of "Meeting Point", from which 290 added to the score of this topic. In other words, there are 57 abstracts with one keyword from UKR topic and more than one keywords from another topic.

It can be easily seen that the agenda was dominated by USA and UKR topics, and "Meeting Point" paid significant attention to EUR topic as well.

In Figures 1-4 we present the weekly scores for these two topics starting from 12 September 2016, when the first episode of "60 minutes" was broadcasted. From that day to 24 August 2017 this show had one episode in a day, and from 28 August 2017 there are two episodes. Accordingly, graphs for the two periods of time are presented separately.

We believe the findings to be applicable to studying agenda-setting function of political talk shows and estimating the influence of these shows on public attention to certain topics and, more broadly, on public opinion.

4. Discussion

The previous Section shows how proportion of each topic in the agenda can be estimated. The important application is the study of propaganda battles which have recently started to implement some ideas of the agenda-setting theory [5, 6].
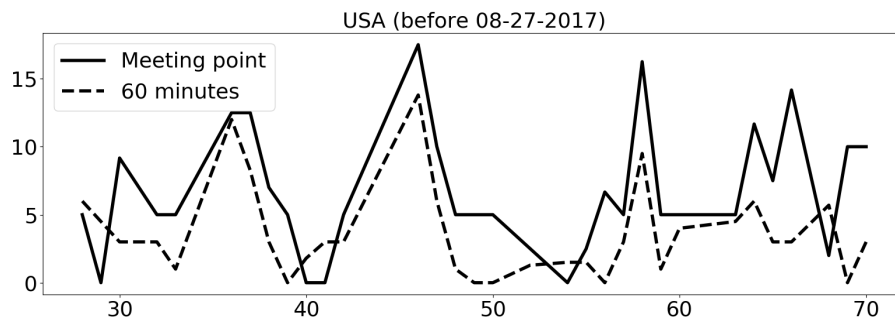
Figure 1: Scores for USA topic: 12 September 2016 to 24 August 2017 (abscissa: week, ordinate: score).
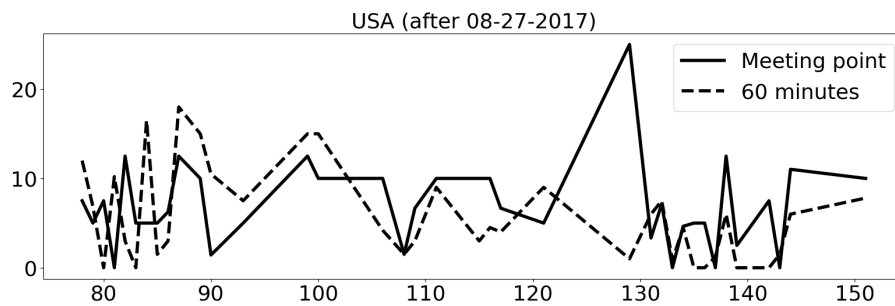

Figure 2: Scores for USA topic: 27 August 2017 to 18 April 2019 (abscissa: week, ordinate: score).
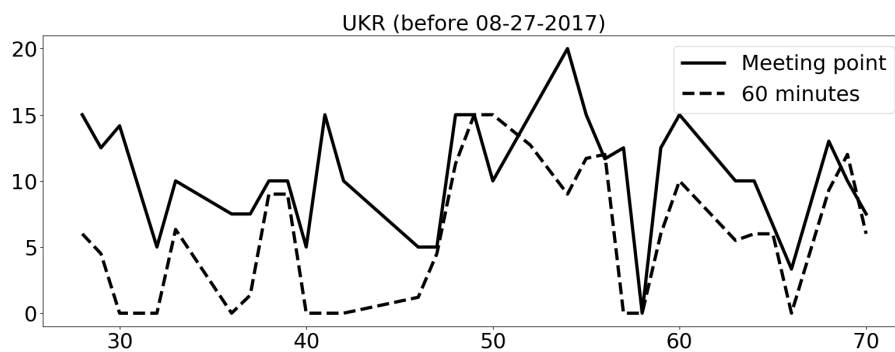

Figure 3: Scores for UKR topic: 12 September 2016 to 24 August 2017 (abscissa: week, ordinate: score).

Still, models of rumors, information influence and information warfare predominantly focus on other aspects, such as network effects [7] or opinion dynamics [8, 9, 10].

Some models emphasize the coaction of mass media and interpersonal communications in spread of information [11, 12, 13].

There is also a wide range of related empirical studies including analysis of search queries [14, 15, 16, 17].
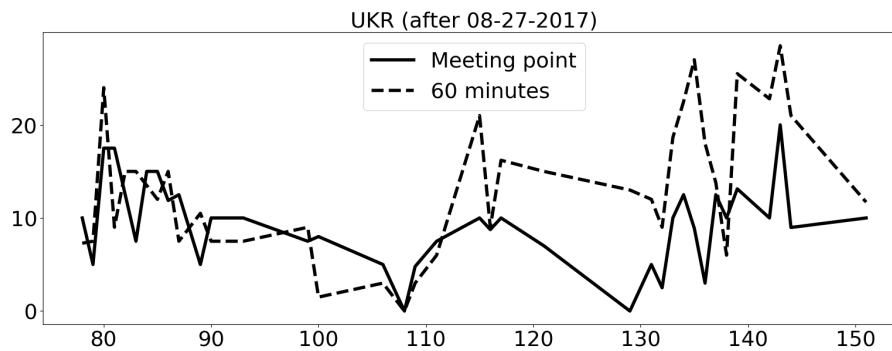
Figure 4: Scores for UKR topic: 27 August 2017 to 18 April 2019 (abscissa: week, ordinate: score).

The growing volume of literature, however, keeps dealing with a single-topic situation. Thus, the area of information battles with multicomponent agenda remains underresearched.

Acknowledgments

References

[1] M. McCombs, D. Shaw, The agenda-setting function of mass media, Public opinion quarterly 36 (1972) 176–187.

[2] B.C. Cohen, The Press and Foreign Policy, Princeton University Press, Princeton, 1963.

[3] Mediascope tv index, 2020. URL: https://mediascope.net/data.

[4] M. Korobov, Morphological analyzer and generator for russian and ukrainian languages, Analysis of Images, Social Networks and Texts (2015) 320–332.

[5] A. Petrov, O. Proncheva, Propaganda battle with two-component agenda, in: Proc. of the MACSPro Workshop 2019. Vienna, Austria, March 21-23, CEUR Workshop Proceedings, 2019, pp. 28–38.

[6] A.P. Petrov, O.G. Proncheva, Modeling position selection by individuals during informational warfare with a two-component agenda, Mathematical Models and Computer Simulations 12 (2020) 154–163.

[7] A.G. Chkhartishvili, D.A. Gubanov, D.A. Novikov, Social Networks: Models of information influence, control and confrontation, Springer International Publishing, Cham, Switzerland, 2019.

[8] D. Gubanov, I. Petrov, Multidimensional model of opinion polarization in social networks, in: 2019 Twelfth International Conference "Management of large-scale system development" (MLSD). Moscow, Russia, Moscow, Russia: IEEE, 2019, pp. 1–4.

[9] A.G. Chartishvili, I.V. Kozitsin, V.L. Goiko, E.R. Saifulin, On an approach to measure the level of polarization of individuals' opinions, in: 2019 Twelfth International Confer-

ence "Management of large-scale system development" (MLSD). Moscow, Russia, Moscow, Russia: IEEE, 2019, pp. 1–5.

[10] I.V. Kozitsin, A.M. Marchenko, V.L. Goiko, R.V. Palkin, Symmetric convex mechanism of opinion formation predicts directions of users' opinions trajectories, in: 2019 Twelfth International Conference "Management of large-scale system development" (MLSD). Moscow, Russia, Moscow, Russia: IEEE, 2019, pp. 1–5.

[11] A.A. Samarskii, A.P. Mikhailov, Principles of Mathematical Modelling: Ideas, Methods, Examples, Taylor and Francis Group, 2001.

[12] A.P. Mikhailov, N.A. Marevtseva, Models of information struggle, Mathematical Models and Computer Simulations 4 (2012) 251–259.

[13] A.P. Mikhailov, A.P. Petrov, O.G. Proncheva, A model of information warfare in a society with a piecewise constant function of the destabilizing impact, Mathematical Models and Computer Simulations 11 (2019) 190–197.

[14] A. Boldyreva, O. Sobolevskiy, M. Alexandrov, V. Danilova, Creating collections of descriptors of events and processes based on internet queries, in: Proc. of 14-th Mexican Intern. Conf. on Artif. Intell. (MICAI-2016), volume 10061(26), Springer Cham, 2016, pp. 303–314.

[15] A. Boldyreva, M. Alexandrov, O. Koshulko, O. Sobolevskiy, Queries to internet as a tool for analysis of the regional police work and forecast of the crimes in regions, in: Proc. of 14-th Mexican Intern. Conf. on Artif. Intell. (MICAI-2016), volume 10061(25), Springer Cham, 2016, pp. 290–302.

[16] L. Akhtyamova, M. Alexandrov, J. Cardiff, O. Koshulko, Queries to internet as a tool for analysis of the regional police work and forecast of the crimes in regions, in: Opinion Mining on Small and Noisy Samples of Health-related Texts. In: Advances in Intelligent Systems and Computing III (Proc. of CSIT-2018), volume 871, Springer, AISC, 2019, pp. 1–12.

[17] L. Akhtyamova, J. Cardiff, LM-Based Word Embeddings Improve Biomedical Named Entity Recognition: A Detailed Analysis, Springer,Cham, 2020.