Quantifying and comparing web news portals' article salience using the *VoxPopuli* tool

Bonacci, Duje[a, b]; Jelinić, Antonija[b]; Jurišić, Jelena[a] and Alujević-Vesnić, Lucija[a]

[a]Communicology Department, School of Croatian Studies, University of Zagreb, Croatia
[b]VoxPopuli project

Abstract

VoxPopuli *tool enables quantification of absolute and relative salience of news articles published on daily news web portals. Obtained numerical values for the two types of salience enable direct comparison of audience impact of different news articles in specified time period. Absolute salience of a news article in a specified time period is determined as the total number of distinct readers who commented on the story in that period. Hence, articles that appear on web portals with larger audiences will in general be (absolutely) more salient as there are more potential commentators to comment on them. On the other hand, relative salience of a particular article during a particular time period is calculated as the quotient of a number of distinct readers who comented on that particular story and the number of all readers who in the same period commented on any news story published on the same news portal. As such relative salience will always be a number between 0 and 1, irrespective of the popularity of particular news portal, the (relative) salience of news stories on different news portals can be compared.*

Keywords: *VoxPopuli; news article salience; agenda setting theory; daily news web portals; readers' comments analysis.*

1. Introduction

1.1. Concept of issue salience and its measurement

As elaborated by Wlezien (2005), the concept of issue salience emerged in political sciences referring to the ''importance'' an individuals placed on certain issues. It indicates the perceived importance and/or prominence that a person attaches to particular (most often political) issue. Hence it can be said that salience of an issue determines the "ranking" of particular item on that individual's "private agenda". The higher the salience of the item, the higher up the agenda it is positioned.

When taken over a certain population, the cummulative "private agenda" turns into "population agenda". Hence, the capability to determine the salience of particular issue opens up the possibility of measuring – in quantitative terms - the structure of "public agenda" of the specified "public".

In practical research terms, salience is usually measured using a survey, typically by asking respondents to indicate 'the most important problem' facing the nation (Wlezien (2005)).

New media platforms, such as daily news web portals, enable new approach to measurement of issue salience. Namely, as interaction of the audience with the news stories published on such portals can be precisely quantitatively and systematically monitored, and as that interaction is necessarily related to the subjective importance the portal visitors attach to the issue elaborated by the respective news story, the quantitative measure of salience of issues raised by various news stories can be constructed/calculated and compared. As such web portals' audience can be taken to represent to some extent some broader social group (e.g. inhabitants of a particular town or city for a small local web portal, or citizens of particular country for a broader national web portal), this opens up the possibility of systematic real-time monitoring of issue salience

1.2. Readers' comments as an indicator of the news article salience

Salience of a particular news article published on some daily news web portal can be quantified using several distinct parameters. First that comes to mind is the number of (page)views the story attracts, as this parameter indicates how many portal visitors opened the story in their browser. However, as will be argued below, the validity of this parameter as an indicator of news story salience is questionable. The second potential indicator – and we argue much more valid – is the number of distinct visitors who engaged in a commenting of a particular news story.

Comments are better indicator than pageviews in several respect. First, not all daily news web portals have publicly available number of pageviews, whereas all the readers' comments (albeit only for those portals that enable commenting – but in fact nowdays great majority of them do) are necessarily publicly available and visible. Certailny, from the publishers' perspective, one of the point of the readers comments is actually to attract more visitors pageviews and so to expand the opportunity for online advertising.

Second, pageviews can be much more easily artificially "boosted" by the interested parties using automated software scripts that send page requests for a particular news stories on a particular portal. The comments, however, cannot be so easily manipulated for two reasons. First, it is not such an easy taks to generate a large number of false readers' profiles, and without these it is impossible to post comments. Namely, great majority of commenting platforms (e.g. Facebook comments plugin - https://developers.facebook.com/docs/plugins/comments, Disqus - https://publishers.disqus.com/) require some form of user authentification in order to be able to post the comments. Second, as comments are very contextual and intertextual pieces of text, it is impossible to generate huge number of diverse random comments which will appear as posted by the real person in a particular comment discussion. Hence, artificial/fake commentators and all of their respective comments can easily be filtered out from the analysis.

Third, the number of pageviews can be significantly influenced by the editorial interventions such as physical positioning of the news story on the web portal frontpage and/or vesting it with the exaggerated or attractive but misleading headline. On the other hand, it is safe to assume that readers will make an effort to comment on an article they opened not based on its primary appearance (reflected in afore mentioned editorially controlled parameter), but on the content of the news story contained. In other words, whereas readers of the daily news web portals will open many articles published on the news portal (hence increasing respective articles' pageview number), they will comment only on those articles whose content they find substantially important, as judged by their own subjective personal standards – i.e. the ones that really are for some reason salient to them.

In this light, we argue that the number that best quantifies the salience of a particular news article published on some web portal is actually not the number of comments the article attracts, but the number of distinct commentators that engage in commenting the article. This number indicates how many readers found the issues behind the news article personally significant to such an extent that they had an urge to actively state their opinon regarding that issue. What is more, from the qualitative content analysis of the comments themselves, the deeper reasons and explanations for such readers attitutes can be glanced. The number of comments then additionally indicate how controversial the issue is, because

the more controversial the issue, the more discussion develops among the involved commentators, and the more comments are hence generated.

2. VoxPopuli tool

2.1. Data harvesting engine properties

VoxPopuli is a software system/tool that enables automatic and systemtic monitoring and comparison of salience of issues published on daily news web portals. It achieves this by analyzing in real time the number of visitors who commented any of the articles published at these portals in a specified period by monitoring .

Engine currently harvests data from 43 Croatian local, regional and national daily news web portals, but is fully universal and can be adapted to harvest data from any daily news portal in the world.

Harvesting of news articles is set to 10-15 minutes intervals whereas each article is checked for new comments at interval between 8 minutes and several hours, depending on the current intensity of commenting activity for particular article, which is automatically determined by the engine. All these parameters can be modified to greater or smaller values if required.

The system is build as an server Java™ application which stores the data in Oracle™ MySQL database. System also has a web front-end (http://voxpopuli.hr) built in PHP which enables users to search for the news stories published on monitored portals in various periods.

2.2. Data quantities and data interpretation

VoxPopuli is a big data tool. Table 1. presents the typical quantity of data harvested by the *VoxPopuli* harvesting engine during a day, week and month.

Table 1. Typical daily, weekly and monthly quantities of data harvested by the VoxPopuli engine.

	Day (18.2.2016)	Week (8.-14.2.2016)	Month (january 2016)
Articles published	1.757	9.768	45.745
Articles commented	1.222	4.460	23.211
Total comments	24.578	117.793	699.251
Distinct commenters	6.166	15.974	42.453

The number of articles published by the monitored portals can be taken as an estimate of the number of (public) issues raised by these portals in specified period. Certainly, as "big news stories" always get covered in multiple articles, the number of issues is actually smaller than the number of articles. But this number is certainly still much greater that the one that can be administered through any survey questionnaire. As can be seen from data in Table 1., daily number of issues raised is counted in hundreds (~1700 articles for the day analised).

The fact that the number of published articles is smaller than the number of articles commented (roughly half of all the published articles attract any comments) indicates that not all news stories are considered equally significant (i.e. salient) by the readers-commentators. What is more, the data presented in Figure 1. show that, even among the articles that did attract some comments, the overwhelming majority - about 70% - of them attracted less than 10 commentators, whereas only about 2% of them attracted more than 100 commentators. This clearly demonstrates that even though daily news web portals raise a huge number of issues through the articles they publish, the readers' interest in these issues is actually very focused.
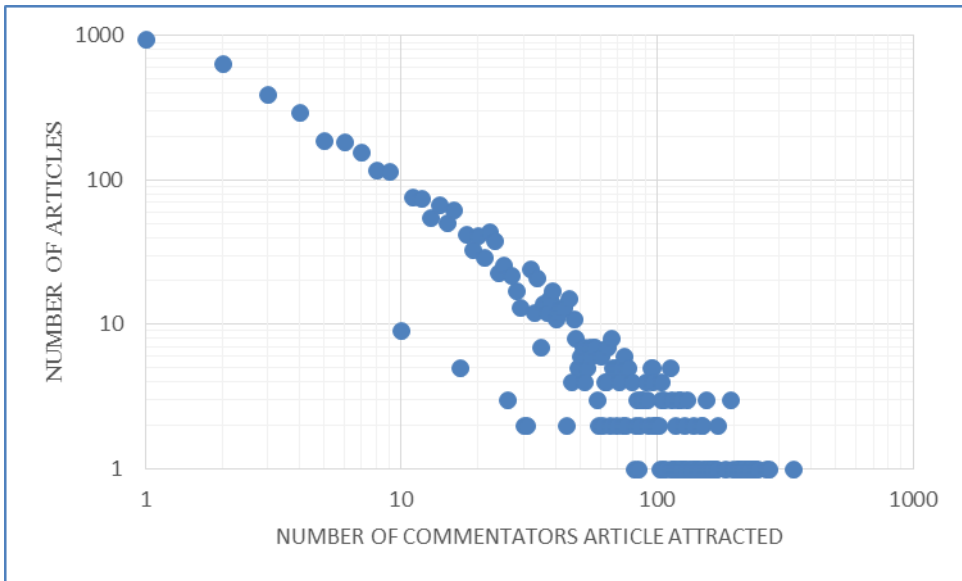

Figure 1. Distribution of commented articles with respect to the number of commentators who engaged in commenting them. Data correspond to the one week period between 8.-14. February 2016..

2.3. VoxPopuli *analysis*

Finally, Figure 2. presents the "*VoxPopuli* analysis" - comparative overview of all articles on all the daily news web portals currently monitored by the *VoxPopuli* system that

attracted some readers' comments, with respect to 3 distinct parameters of each article: absolute salience, relative salience and controversiality. Each bubble on the chart represents a single commented article.

Absolute salience of a news article in a specified time period is calculated as the total number of distinct readers who commented on the story in that period and is represented by the size of the bubble – the greater the bubble, the more commentators engaged in commenting the article. Articles that appear on web portals with larger audiences will in general be absolutely more salient as there are more potential commentators to comment on them and their respective bubbles will always stand out.

Further, **relative salience** of particular article during particular time period is calculated as the quotient of a number of distinct readers who comented on that particular article and the number of all readers who in the same period commented on any news article published on the same news portal. As such relative salience will always be a nulber between 0 and 1 (i.e. 0% and 100%), irrespective of the popularity of particular news portal, the relative salience of news articles on different news portals can be compared. This parameter is represented by the position of the article's circle on the vertical axis.

Finally, **controversiality** of the article is calculated as the percentage of commentators who posted more than one comment to the respective article. This parameter is represented on the horizontal axis of the graph. Namely, for any article analysed, great majority of commentators leave just one single comment. Multiple comments from a single author usually stem from the discussion in which that commentator engaged with other commentators. The more commentators engaged in discussion usually indicates that they "hotly debate" the issue presented by the article.
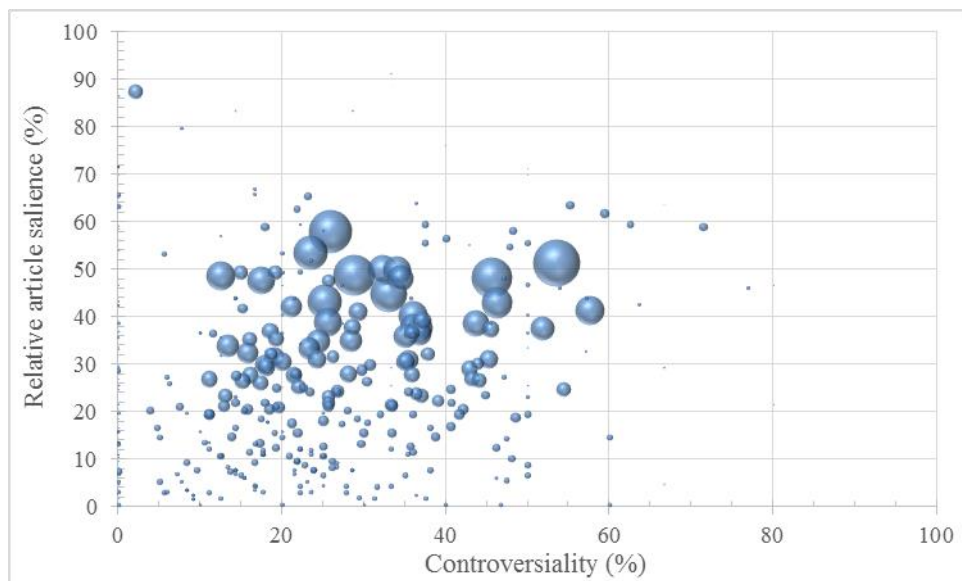
Figure 2. VoxPopuli *analysis graph. Data correspond to the one week period between 8.-14. February 2016.*

Using such chart, which can be (and in fact in current version of the VoxPopuli tool is) generated in real time, enables us to simultaneously monitor the salience of different news articles on various news portals and hence to quickly identify the "issue of the hour/day/week/etc." among all the issues raised by the monitored portals.

References

Wlezien, C. (2005). On the salience of political issues: The problem with 'most important problem'. Electoral Studies, 24, 555-579.