



Quantitative and Qualitative Analysis of Linking Patterns of Mainstream and Partisan Online News Media in Central Europe

Journal:	<i>Online Information Review</i>
Manuscript ID	OIR-10-2020-0441.R1
Manuscript Type:	Research Paper
Keywords:	Partisan online news media, Media profiling, Central Europe, Hyperlink network analysis, Content analysis, COVID-19

SCHOLARONE™
Manuscripts

Citation:

Hrckova, A., Moro, R., Srba, I. and Bielikova, M. (2021), "Quantitative and qualitative analysis of linking patterns of mainstream and partisan online news media in Central Europe", *Online Information Review*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/OIR-10-2020-0441>

Quantitative and Qualitative Analysis of Linking Patterns of Mainstream and Partisan Online News Media in Central Europe

Abstract

Purpose: Partisan news media, which often publish extremely biased, one-sided or even false news, are gaining popularity world-wide and represent a major societal issue. Due to a growing number of such media, a need for automatic detection approaches is of high demand. Automatic detection relies on various indicators (e.g., content characteristics) to identify new partisan media candidates and to predict their level of partisanship. The aim of our research is to investigate to a deeper extent whether it would be appropriate to rely on the hyperlinks as possible indicators for better automatic partisan news media detection.

Design/methodology: We utilized hyperlink network analysis to study the hyperlinks of partisan and mainstream media. The dataset involved the hyperlinks of eighteen mainstream media and fifteen partisan media in Slovakia and Czech Republic. More than 171 million domain pairs of inbound and outbound hyperlinks of selected online news media were collected with Ahrefs tool, analyzed, and visualized with Gephi software. Additionally, three hundred articles covering COVID-19 from both types of media were selected for content analysis of hyperlinks to verify the reliability of quantitative analysis and to provide more detailed analysis.

Findings: We conclude that hyperlinks are reliable indicators of media affinity and linking patterns could contribute to partisan news detection. We found out that especially the incoming links with dofollow attribute to news websites are reliable indicators for assessing the type of media, as partisan media rarely receive links with dofollow attribute from mainstream media. The outgoing links are not such reliable indicators as both mainstream and partisan media link to mainstream sources similarly.

Originality/value: In contrast to the extensive amount of research aiming at fake news detection within a piece of text or multimedia content (e.g., news articles, social media posts), we shift to characterization of the whole news media. In addition, we did a geographical shift from more researched U.S.-based media to so far under-researched European context, particularly Central Europe. Our results and conclusions can serve as a guide how to derive new features for an automatic detection of possibly partisan news media by means of AI.

Keywords: Partisan online news media, Media profiling, Central Europe, Hyperlink network analysis, Content analysis, COVID-19

Introduction to Partisan and Mainstream Online News Media

A plethora of efforts have been conducted by researchers, industry, or media professionals (journalists and fact-checkers) to tackle with biased or false information in the recent years. The most of these efforts focused on content at the level of claims (i.e., manual, or automatic fact-checking (Cazalens et al., 2018; Kotonya & Toni, 2020)) or larger pieces of texts (such as news articles/blogs or users' posts on social networks (Guo et al., 2020; Zhou & Zafarani, 2020)). Various characterization studies have been

conducted (e.g., (Vosoughi et al., 2018)) and (semi-)automated detection methods powered by machine learning, natural language processing or image processing have been proposed (Choraś et al., 2021).

Nevertheless, biased or false information may spread in an extremely fast fashion (Vosoughi et al., 2018) and even automatic solutions may not provide a timely response. In addition, new stories (e.g., false narratives related to COVID-19) may cause significant concept drifts and hamper models' performance. Recently, an alternative promising branch of solutions has been recognized by focusing on *sources* - online news media that publish the news articles (Baly et al., 2018; Nakov et al., 2021). Online news media profiling aims to *analyze or predict* different types of *biases or factuality of reporting* of entire news outlets with respect to various topics. The outcome can serve to improve detection of biased or false information or help fact-checkers to focus on problematic sources.

We particularly focus on profiling online news media that show extreme biases – commonly denominated as *hyper-partisan media* (or *partisan media* in short). Partisan bias refers to a strong leaning towards one opinion or party (e.g., political party). In general, online media (including mainstream ones) may contain manipulated or one-sided opinions, whether it is yellow journalism or biased news with clickbait articles or PR articles (Hrčková et al., 2019), however, their presence is significantly more common in partisan media. Partisan media may even go further beyond the border of truth with fabricated content (fake news), manipulated images, videos (deep fake), or images presented in different context as well as misquoted resources. These media also go beyond the borders of (journalism) ethics or even the law by publishing articles with aggressive personal attacks or defamations of minorities. They also do not tend to publish corrective reports.

Partisan media are sometimes marked as fake or conspiracy news outlets that gives them a negative connotation. These media rather call themselves as “alternative” to the pro-western mainstream (traditional) media, often referring to Edward S. Herman & Chomsky (1988). The online mainstream media are for the purpose of this article defined as media news portals that have a longer history, employ journalists doing fieldwork and respect some journalism ethics and law, whether it is tabloid or serious news.

In this article, we focus on the partisan and mainstream media operating within *Central European countries* with both communistic and fascistic past that is currently rediscovered by these media to attack the emotions of their readers with increasing success (see Appendix A for more contextual information). Slovak and Czech partisan media with their pro-Russian or conservative (on the border of neo-Nazi) attitude, are a suitable model case for this kind of study, as these tendencies can be observed also in other parts of the world.

Our objective is to analyze particularly the *external hyperlinks and linking patterns* between partisan and mainstream media and thus provide some indicators that can be further utilized to automatically detect partisan media. With this objective in mind, these research questions were set: Are the mainstream and partisan media interconnected with hyperlinks or co-linked? Are the linking patterns different in mainstream and partisan online news media? Who are the main actors in the networks? What kind of hyperlinks and sources can be found in online news media? Are they different in partisan and mainstream news media?

Our main contributions are as follows:

1. We present an in-depth study focused on news media themselves (contributing to the recent shift from content to sources). Within the task of media profiling, we particularly focus on hyperlink network analysis. Previous works, which analyzed hyperlinks contained in news articles (e.g.,

- (Dhoju et al., 2019)), mostly considered only quite simple metrics, such as a total number of references. In our work, we go one step further and analyze the surrounding hyperlink network. In addition to quantitative analysis, we also employed qualitative content analysis method.
2. While most of the existing works focus on USA (especially on spreading false information during presidential election in 2016 (Fourney et al., 2017)), only limited research addresses partisan media in other countries (e.g., during German federal election in 2017 (Morstatter et al., 2018)). Additional work on non-English corpora and different political systems has been recognized in the recent survey (Nakov et al., 2021). Our study is purposefully focused on the influence of partisan online news media in the European context. The young post-communist democratic countries of Central Europe (Czech and Slovak republic) were selected to fill the gaps in the research, as these countries are severely affected by the spread of biased as well as false information that has a secondary impact on the development of the whole region.
 3. Finally, we provide a discussion, how the achieved results can help in consequent automatic detection of partisan content and sources that create or spread such undesired content (such as (Baly et al., 2018)).

Theoretical background

Factuality and Bias of News Media

Prevalence of biased or false content on the web has already made it challenging to evaluate each claim or article, either manually or automatically. Recently many researchers have been shifting their attention to profile entire news outlets (Baly et al., 2018; Nakov et al., 2021). The problem of discovering news sources of false information (mainly fake news) remains largely unexplored (Chen & Freire, 2020).

Profiling typically aims to investigate either media’s factuality (source reliability) or biases (e.g., selection or presentation bias) - in practice, a low factuality and a high bias are commonly correlated (Nakov et al., 2021).

A part of existing research papers aims on automatic prediction of media factuality/bias. Chen & Freire (2020) proposed an automatic discovery system that proactively identifies and classifies domains with a high “fakeness”. In the first step, Twitter is used to uncover co-sharing structures to discover new domains; in the second step, topic-agnostic classifier is used to evaluate and rank the discovered domains. Recently, Bali (2020) proposed a classification model utilizing a wide spectrum of features: derived from the produced content (articles themselves, YouTube video channels, and Twitter accounts associated with news media), readers (at Twitter, YouTube, and Facebook) and what was written about the target medium (at Wikipedia).

Another part of the existing research consists of characterization studies that are needed to shed more light on the distinction between credible and uncredible sources. The valuable output from this type of studies serves also as a source of potential features for automated prediction methods. Bhatt et al. (2018) described the ecosystem of partisan websites from several perspectives: producers, consumers, traffic flows, and Facebook communities. Dhoju et al. (2019) similarly described the structural, topical, and semantic differences between reliable and unreliable health related news sources.

Hyperlinks play a critical role for partisan websites since they are used to direct users to other allied partisan media or to advertisement (revenue from ads is one of the major financial motivations for owners of partisan outlets). In addition, the presence or the absence of hyperlinks was proven to have a significant influence of users’ trust towards news media (Verma et al., 2017). Existing studies revealed some

preliminary differences in usage of hyperlinks and linking patterns between partisan (unreliable) and mainstream (reliable) news outlets. However, a deeper insight reflecting their importance is still missing. At the same time, hyperlinks/linking patterns have a high potential to be good indicators for prediction methods and in addition, they can also serve as good sources for identification of new partisan media candidates (i.e., co-linking can be used as an alternative to co-sharing structures employed in the work by (Chen & Freire, 2020)).

Social and hyperlink network analysis

Hyperlinks are technological capabilities that enable websites to link one with another, but also to exchange information, and maintain cooperation between actors with common background, interest, or projects (Park & Thelwall, 2008). The quantity and quality of external hyperlinks to the websites are set as the indicators of web content quality by search engines like Google (Makulová & Hřčková, 2011).

As a methodology to study hyperlinks among websites, Jackson (1997) first suggested that the social network analysis (SNA) methods are applicable. SNA is a research method for identifying structures or patterns between social entities in various social systems based on the relations among the system components (or nodes) rather than on the attributes of individual actors. Hyperlink network analysis (HNA) method is considered a type of SNA and might enable the analysis of cooperation on internet websites.

Hyperlinks have been studied in the context of news outlets for a long time. International hyperlink structure of online news media was analyzed by Himelboim (2010), utilizing the coding of hyperlinks typical for content analysis method. The author conceptualized hyperlinks as relations among countries in the world system and as such as indicators of information flow and influence among the countries. He identified the scarcity of hyperlinks use by journalists and a strong hierarchical information flow from countries of power and from traditional sources, dominated by US news. Szabo and Bene (2015) analyzed the position of radical right mass communication channels in the media sphere in the Hungarian media network. Pat et al. (2020) applied a gravity model to analyze relationships between pairs of 89 U.S.-based news outlets and produced a novel spatial network visualization.

Nevertheless, HNA, despite its long tradition, remains underutilized in the context of new phenomena like partisan media. SNA in social networks (e.g., Twitter) is a more common method to analyze data also in the field of biased news, fake news, or rumors detection, e.g., (Allcott & Gentzkow, 2017; Morstatter et al., 2018). With this article, we show that HNA still has a potential to reveal important information about the media relationships and linking patterns of media – both mainstream and partisan.

Research methodology

The study was conducted both by quantitative and qualitative research methods. For the implementation of the quantitative part, two software systems were utilized: Ahrefs¹ and Gephi². Ahrefs is a crawler, the original purpose of which is to serve search marketing professionals. It collects, processes, and stores data, consisting of hyperlinks, keywords, and user behavior. Using commercially available crawlers to discover hyperlinks between websites has been a common practice in webometrics (Björneborn & Ingwersen, 2004). AhrefsBot is also mentioned as the second most active crawler on the internet after

¹ <https://ahrefs.com>

² <https://gephi.org/>

Googlebot (Zeifman & Breslaw, 2017). Gephi is an open-source software for visualization and exploration of graphs and networks, applicable also for link analysis. Thanks to Gephi visualizations, the networks of directed hyperlinks from and to the online news media and the importance of some nodes, based on the automatic calculation of the sum of hyperlinks can be visible.

Data from the most popular Czech and Slovak online news media were obtained throughout July and August 2019. The outlets were selected with regards to their popularity and according to the existing lists (e.g., [konspiratori.sk](https://www.konspiratori.sk/)³), confronted with the above-mentioned definitions of partisan and mainstream media. The popularity of the mainstream media was identified by available statistics (e.g., IABmonitor by Gemius⁴) and the popularity of partisan media, where official statistics are often missing, was determined by the number of Facebook fans or the number of associated hyperlinks. We strived for equal representation of partisan and mainstream media – in Slovakia, we selected seven partisan and nine mainstream media. In Czech Republic, where the news media scene is richer, ten partisan and nine mainstream media were selected. The complete list of the analyzed media with their description can be found in the Appendix A.

The data that was of our particular interest was the number and domains of incoming and outgoing hyperlinks (associated with dofollow and nofollow attributes) to and from the selected media. Nofollow attribute is a setting on a web page hyperlink that directs search engines not to use the link for page ranking calculations. Hyperlinks with the nofollow attribute are also not officially followed by the search engines. Hyperlinks with the dofollow attribute (that is a default value) are commonly considered to pass “weight” and help a website or URL to rank better in the organic search results. These data were downloaded in the CSV format and manually processed for the needs of Gephi. Hyperlinks to social networks, Google tag manager and hyperlinks created by link shorteners (like bitly) were purposefully removed. The duplicates were merged and the sum of the hyperlinks in duplicates was calculated automatically by Gephi. To define the weight of edges and the modularity class of the nodes, the number of all hyperlinks (with dofollow and nofollow attributes) was selected. The number of dofollow links was set as an additional attribute to analyze the data, because many of the analyzed hyperlinks had a nofollow attribute.

Finally, we collected and analyzed more than 171 million domain pairs, linking from and to the selected media (more than 160 million of hyperlinks from and to the mainstream media and more than 11 million of hyperlinks from and to the partisan media). The number of hyperlinks to and from mainstream news media is noticeably higher than the number of links of partisan media as mainstream media are more established in the news landscape and one of their link building strategies is buying smaller and specialized portals with online history and hyperlinks. To visualize the data in the graph properly, Fruchterman Reingold layout (1991) was selected. To filter the data, degree range $n > 1$ and edge weight of $n > 0.01$ of the total number of hyperlinks (total edge weight) was invoked. The graphs in the article are created from both dofollow and nofollow attributes.

The qualitative analysis of news media hyperlinks was conducted in May 2020 by the content analysis method. A subset of thirty media, and within each of them a sample of ten representative articles related to COVID-19, were selected manually - most often by browsing the topic (tag) COVID-19 or “coronavirus” (resulting into 140 articles from partisan and 160 articles from mainstream media). To be

³ <https://www.konspiratori.sk/>

⁴ <https://rating.gemius.com/sk/tree/112>

representative, we included diverse types of articles from each medium (in similar ratio, where possible), such as medical advice, interviews with experts, commentaries, or reportage.

The hyperlinks and other non-hyperlink mentions of sources within the selected articles were manually identified and coded (in total 12 036 hyperlinks and mentions). Each link (hyperlink or non-hyperlink mention) was annotated by 4 different aspects: 1) its type (if the link targets to any kind of website/PDF document/multimedia/author's profile), 2) placement (if the link was placed directly in/alongside/underneath the article content), 3) function (with/without the functional/clickable hyperlink), and finally: 4) the nature of the external linked source (a recommended article/a category or tag from the directory/a credible source, such as a research paper/a less credible source, such as Wikipedia or a social media post/a public institution or NGO/a mainstream or partisan media).

While we deeply analyzed all annotated aspects, for the purpose of visualizations twelve aggregated categories of links were identified from the most common combinations of the listed aspects (other less common combinations were not representative or relevant to our aims); see Fig. 9-11 for the list of all resulting categories). The visualizations by means of pie charts and tree maps were done using the Infogram⁵ tool.

Results

Are the mainstream and partisan media interconnected with hyperlinks or co-linked?

The local news media landscape can be inferred from the graph of all hyperlinks from and to the most popular mainstream and partisan media of the researched country. The graph (Fig. 1) suggests that there are separated mainstream media clusters with more than one million hyperlinks. One popular partisan medium (hlavnespravdy.sk) can be spotted that is co-linked with a mainstream medium (pravda.sk) by a questionable domain (harekrsna.cz). When analyzing this mainstream medium (pravda.sk), a technical mistake was found in the HTML code that was caused by their content management system (CMS generated nofollow tags as "no follow", "n ofollow" etc. that the crawler did not recognize as nofollow). Ignoring this mistake, the clusters of mainstream media and their congenial media of the same ownership remain separated.

According to the analysis in Ahrefs tool, Slovak mainstream media rarely link to partisan media, except of tens of thousands of outgoing nofollow links from the mainstream media blogs. We assume that guest blogging is one of the popular link building practice of Slovak partisan media. Although these links are marked as nofollow for the search engines, they are visible to users and thus connecting the partisan media scene with the mainstream media.

Fig. 1 Network of directed hyperlinks from and to the most popular Slovak partisan and mainstream media

In Czech Republic, three big clusters of mainstream and partisan news media are visible (Fig. 2). The position of partisan media is stronger, half of one cluster on the right side belongs to the partisan news

⁵ <https://infogram.com/>

media (ceskoaktualne.cz, az24.cz, zvedavec.org and parlamentnilisty.cz), all linked with a questionable portal harekrsna.cz. This portal also links these news media to the mainstream media (idnes.cz and lidovky.cz), owned by a Czech politician and millionaire. Nevertheless, the links from Czech mainstream media to the partisan media are also not present.

Fig. 2 Network of directed hyperlinks from and to the most popular Czech mainstream and partisan media

Are the linking patterns different in mainstream and partisan media? Who are the main actors of the networks?

Different linking patterns can be found between partisan and mainstream media. The clusters of the hyperlinks from and to the mainstream media correlated with the ownership (publisher) of these media both in Slovak (Fig. 3) and Czech Republic (Fig. 4). There seems to be only one exception (hyperlinks between two media sme.sk and hnonline.sk of different ownership), but a deeper analysis of Ahrefs results revealed that this number of hyperlinks was created due to redirections of one subdomain in one media to another. This was caused probably by the transfer of ownership between these two subdomains. The other links were usually automatic interlinks between these portals. This result of individual clusters suggests independence and separateness of the researched mainstream media and/or unwillingness to link to the “competition”.

Fig. 3 Network of directed hyperlinks from and to the most popular Slovak mainstream media

Fig. 4 Network of directed hyperlinks from and to the most popular Czech mainstream media

The analyzed partisan media (both Slovak and Czech) create a denser and more interlinked network that confirms their interdependence in the network (Fig. 5). Six co-cited (co-linked) Slovak partisan media nodes can be spotted in contrast with mainstream Slovak media, where the network consists of usually separated clusters. The graph also shows partisan portals with the highest importance in the network: dolezite.sk, infovojna.sk, and hlavnespravy.sk. The first two link massively both to mainstream and partisan media that is a very different linking pattern from the mainstream media that link to partisans just from their blogs and discussions (with nofollow attribute). Slovak partisan media position themselves as parts of the online news media landscape and the hyperlinks to the established media might at first glance look like an indication of their credibility. Czech partisan media link to the mainstream media less often. In the Czech partisan online news media landscape, the actors with the highest importance (parlamentnilisty.cz and slobodnyvysilac.cz) are apparent similarly as in the Slovak partisan online news media landscape (Fig. 6).

Fig. 5 Network of directed hyperlinks from and to the Slovak partisan media

Fig. 6 Network of directed hyperlinks from and to the Czech partisan media

Besides the links to news media (and their web hosting companies), most of the links from partisan news media point to the portals about boundary topics as bitcoin, green architecture, and esotericism. Only the most popular partisan medium (hlavnespravdy.sk) is attracting a wider audience with the links from a music portal and a portal for infographics/memes. The links from and to the mainstream media are more business oriented or targeted to a mainstream lifestyle. One of the mainstream media (sme.sk) links regularly to its source that is a database of published court decisions.

When studying the cross-border hyperlink relations between the most popular mainstream media in Czech and Slovak Republic (represented by .cz and .sk domains), clusters of mainstream media remain separated except of one mainstream medium (pravda.sk) due to the above-mentioned technical problems with CMS (Fig. 7). A more significant hyperlink connection may be expected given that some of them have the same publisher (owner). The explanation may lie in a slight language barrier and different local news coverage of these media.

Fig. 7 Network of directed hyperlinks from and to the Slovak and the Czech mainstream media

On the other side, partisan news media of the two independent states are much more likely to be interlinked also internationally than mainstream media. This fact can be explained by transmitting the content from one medium to another that is a well-known strategy of the partisan news media proved by the hyperlinks (Fig. 8).

Fig. 8 Network of directed hyperlinks from and to the Slovak and the Czech partisan media

What kind of hyperlinks and sources do online news media include? Are they different in partisan and mainstream news media?

The previous quantitative analysis has been supplemented with additional qualitative analysis performed on three hundred news articles covering COVID-19 from both types of media. Some trends are visible from the results of our content analysis, where the aggregated categories of hyperlinks were identified, and their occurrence visualized as pie charts (Fig. 9) and tree maps of other links (excluding links to recommended articles and directories, i.e., topical categories or tags) in mainstream (Fig. 10) and in partisan (Fig. 11) media.

As we can see in the pie charts in Fig. 9, the vast majority of the links, both in mainstream and partisan news media is formed by the internal links to the pages on the same website. These links consist either of categories leading the user to read more about a particular topic or person or the recommended articles with the older (more or less) relevant articles in the same news media. Strengthening the internal (on page) linking structure is a common search engine optimization (SEO) strategy - obviously for both types of news media and usually has similar purpose - to keep the user on the same website as long as possible. Just a minority of all hyperlinks in and around news articles are formed by hyperlinks pointing to the external sources (almost 8 percent in mainstream and less than 12 percent in partisan online news media). Based on these results, we can state that hyperlinks in both types of media serve much more as partner

bonds (outgoing links) or as on page links, than as sources of the articles and this effect is even more apparent in the mainstream media.

Fig. 9 Ratio of links' main aggregated categories in mainstream (1) and partisan (2) media

Fig. 10 Ratio of links' other aggregated categories in Slovak and Czech mainstream media (excluding links to recommended articles and directories)

Fig. 11 Ratio of links' other aggregated categories in Slovak and Czech partisan media (excluding links to recommended articles and directories)

The tree maps (Fig. 10 and Fig. 11) indicate the specific aggregated categories of hyperlinks leading to the external sources/domains (labeled as “other” in the pie charts). As we can see, most of the external hyperlinks in the mainstream media lead to the congenial media or in other words media under the same publisher (domain), usually under different subdomain. The content analysis also confirmed the results of the quantitative analysis, indicating the rare, respectively nonexistent links from mainstream media to partisan media.

Partisan media are more generous in mentioning the external resources by hyperlinks than mainstream media - most of the external hyperlinks aim to diverse mainstream media (one partisan media even utilized 6 hyperlinks to mainstream media per article on average) as well as other partisan media. Mainstream media are even the third most common hyperlinks in partisan media overall, leaving behind also other partisan outlets. This finding can be explained in two ways. Firstly, it is to attack and criticize some (usually the most liberal) mainstream media. Secondly, it is also used to justify the claims of partisan media, if it suits the narrative of the article. Partisan media do also link to other partisan media quite often, but the hyperlinks are mostly located in the footer or under the articles (i.e., links to the cited resources or to the original content being adapted/translated). The hyperlinks of partner websites are generally located in the footer for basic link building purposes.

When writing about COVID-19, the primary external sources (e.g., research articles or various trusted public and NGO organizations, human experts) are rarely included in both types of media. Only a half of partisan media (7 out of 14) mentioned such sources at least once in ten articles. This pattern was just slightly better in mainstream media, where 10 out of 16 media relied on the primary sources at least once in the selected ten articles. There results therefore suggest that partisan media, which are usually considered as media that are not doing quality (journalistic) work, formally met this credibility criterion at the similar level as mainstream media.

The low ratio of primary external sources is what we may not expect in the case of mainstream media, as utilizing and citing quality resources is considered as one of the credibility indicators. However, this kind of contradiction with the journalistic values have already been observed in the past (De Maeyer, 2012). One of the reasons, why mainstream media do not back up their claims with hyperlinks to resources, might be business (saving the link juice) and tradition of not revealing their sources. Another reason is a different format of the articles in mainstream media; these are often reports, commentaries or interviews, whereas partisan media rather translate or combine the existing articles with the opinion of the author.

The difference between partisan and mainstream media, however, lies in the way of dealing with these primary sources. In the deeper insight, partisan media use linking to less credible Wikipedia articles (ten of them) or to unverified/unofficial PDF documents hosted on some free document servers (five of them), while such citation in mainstream media is almost impossible to observe. We recognized cherry-picking and fudging with facts and mixing them with non-representative news in some partisan media. In addition, the claims were backed by the researchers of the same opinion. This is a common strategy in the partisan media also in other medical topics (as e.g., in vaccination, the distortion of facts presented by an expert is often present, as well as presenting the opinion of an expert in a different field). Mentioning the authority alone is not a guarantee of presenting the facts correctly, but it can be misleading for some inexpert readers.

The content analysis also showed that articles in partisan media are in the large extent only compiled from the information that was previously published elsewhere. Ten partisan media out of fourteen relied on average on more than one clickable and dofollow resource mentioned at the end of the article, when writing about COVID-19. One partisan medium (zvedavec.org) even mentioned 13 sources per article on average. In contrast, there are just 3 out of 16 mainstream media that relied on average on more than one resource per article mentioned at the end of an article.

Citation of social networks' posts, which are usually perceived as less reliable sources of information, is similar in both types of media. The reason for inclusion of such posts is either to support or illustrate the presented facts. Nevertheless, the mainstream media mention official fan pages of public institutions more often than the citations of citizens.

Partisans back up their claims and content with videos quite often, sometimes mentioning video presence in the title of the article. Although the ratio of utilizing videos is similar between the two types of media, some partisan media used videos more often in the articles - five of fourteen partisan media used the audio-visual material in at least one third of the analyzed articles. This might also contribute to the perceived credibility of the partisan articles, although the videos are often just video records of a monologue or a dialogue.

Finally, we analyzed the presence of article authors. Only few partisan media increase their perceived credibility by mentioning the authors of articles, part of them providing also a hyperlink to other articles of the author. Ten of fourteen partisan media did not mention the name of the author at all or rarely. Surprisingly, although to a lesser extent, this is also a problem of some mainstream media – mostly tabloids, as well as one opinion forming medium. Five out of sixteen mainstream media did not mention the authors of the articles and two mentioned the authors without a hyperlink.

Conclusions and Research Limitations

Different linking patterns and link building strategies of mainstream and partisan news media have been demonstrated in this article. The hyperlink network analysis of mainstream and partisan online news media in two countries showed that hyperlinks are reliable indicators of affinity of these media and that linking patterns could contribute to the partisan news detection. In addition, hyperlinks no longer serve just as citations between the portals but may be also (mis-)used for other purposes as well. In addition, the pure presence of links to sources may not be an indicator of the quality of the media, since sources can be misinterpreted or placed into different context.

The main findings can be summarized as follows:

- Mainstream online news media create hyperlink clusters within their own publisher. More context would be provided to the reader if these news media of different publishers were interlinked.

However, it is not the case of selected mainstream media, and the consequence of this fact is that the ownership of these news media can be reliably inferred from the hyperlink structure. Another important conclusion is that mainstream online news media rarely link to the mainstream online news media of another state, even if they share the same publisher.

- Partisan news media are much more interlinked between the clusters within one state, but also within two independent states. This fact might indicate either their interdependence or partnership in the “partisan fight” against mainstream media or even the same ownership as it is commonly the case in mainstream media.
- Most of hyperlinks in both mainstream and partisan online news media are formed by internal links on directories (categories/tags) or recommended articles from the same domain. Other external resources are mentioned in the form of hyperlinks to a much lesser extent in both types of media.
- The outgoing links are not reliable indicators of partisan news as both partisan and mainstream media link to other mainstream media as well. On the other side, the incoming links with dofollow attribute to the news websites seem to be reliable indicators for assessing the type of the media, as partisan media almost never receive links with dofollow attribute from mainstream media.

Based on these conclusions, we see two interesting options how hyperlinks may serve as input for automatic detection of partisan media. Following the two-step approach proposed by (Chen & Freire, 2020), at first hyperlinks may be utilized to identify partisan media candidates (given a seed of partisan websites), i.e., co-linking can reveal other allied partisan media that constantly appear and hyperlinks are used to direct traffic to them from existing established sources. Secondly, deeper analysis of hyperlink patterns may serve as a source for additional valuable features that may be used (together with other features already used in the existing works such as in (Baly et al., 2020)) by machine learning models to predict partisanship class or score. Such output can be further used in many ways, for example to help fact-checkers identify new partisan media outlets or to automatically detect fake news (due to a correlation between a high bias and a low factuality reporting, partisanship of the source may be a good predictor of content falseness).

Two limitations of this research were identified. The first limitation is the focus on the online news media landscape of two states: Czech Republic and Slovakia. However, this limitation is necessary, as the situational awareness of the news media landscape was needed. Also, the content analysis of hyperlinks would be hard to conduct without knowing the language. Nonetheless, the selected countries represent European post-communist countries, affected nowadays by partisan media. Still, repeating this research with the same research methods, but in different countries is suggested.

The limitation of the conducted quantitative analysis is the restriction to the links that are crawlable by robots. Some mainstream media are under paywall and their links remain not crawled. Nonetheless, the hyperlinks to the affiliated media surely are crawlable and crawled by robots as hyperlinks with dofollow attribute serve also for the search engine optimization of these portals. Moreover, part of the sources, especially in partisan news media remain listed as pure text and part of sources are not stated at all. Therefore, a hyperlink analysis does not serve as citation analysis of the sources, but rather as an analysis of cooperation between the selected actors. These problems were addressed to some extent with the qualitative content analysis, the results of which do not contradict, rather complement the conclusions of the quantitative hyperlink network analysis.

The obtained results outlined the possibilities of applying a hyperlink analysis in the credibility research of news media. We presume, the situation might be different in different domains or in different social

and political context. Therefore, it would be worth to further explore the information flow in online news media through hyperlinks on the international level.

Disclosure and Conflicts of Interest

This article is an extended version of the conference paper presented at DIONE, held online on April 17, 2020.

Acknowledgments

We would like to thank Daniel Duris from Basta Digital for providing access to Ahrefs tool for analytical purposes.

Appendix A - Overview of online news media analyzed in the article

Although there are some similarities in media landscapes in the two countries because of their shared past, they developed more or less independently since Czechoslovakia split into Czech Republic and Slovakia in 1993. A recent study on media plurality in Slovakia noted a high risk regarding the market plurality due to lack of transparency of media ownership, news media concentration, and risks of owner influence over editorial content (Sampor, 2020), mainly due to lacking regulations and standards. The media landscape (and the whole country) was influenced by the murder of an investigative journalist Ján Kuciak and his fiancée in 2018, which shed light at the level of corruption in police, judiciary, and among politicians. Although many news outlets in the country are owned by local oligarchs, there are mainstream media that managed to defend the integrity of their newsroom (e.g., SME – sme.sk, Aktuality – aktuality.sk) and in some cases even their financial independence (Denník N – dennikn.sk). As to the partisan news media, their names often refer to their supposed independence or to them being an alternative to the mainstream. For example, Slobodný vysielateľ - slobodnyvysielac.sk (Free broadcaster) took the name of a Slovak anti-fascist resistance radio from World War II, although it provides a platform to many activists and politicians of the extreme right. In other example, Infovojna - infovojna.sk (Infowar) took the name of a well-known extreme right US website.

The Czech media landscape has undergone large changes in the past two decades as well, especially in the media ownership. The recent study on media plurality noted that most of the Czech media market is divided among a few local oligarchs and businessmen (Stetka & Hajek, 2020). There are high risks of political influence, since some of the biggest and most influential media are owned or influenced by politicians. Most notably, the current Czech Prime Minister Andrej Babiš is an owner of a media company MAFRA that owns online news portal idnes.cz (connected to the printed newspaper Mladá fronta DNES also owned by MAFRA) as well as lidovky.cz (connected to Lidové noviny published by MAFRA). In partisan news media, we see some direct connections to Slovakia (e.g., svobodny-vysilac.cz vs. slobodnyvysielac.sk, parlamentnilisty.cz vs. parlamentnelisty.sk). There is also a strong presence of Russian-backed news, mainly the Czech version of Sputnik News (cz.sputniknews.com) or Aeronet (aeronet.cz) which often disseminate Russian narratives and propaganda.

Slovak partisan news media

Name	Description
dolezite.sk	News hub
hlavnespravdy.sk	Politics, army, interesting facts from technologies and pop culture. Conservative targeting
infovojna.sk	Politics, army
parlamentnelisty.sk	Politics
protiprud.sk	Politics, health, spiritual, conspiracies
slobodnyvysielac.sk	Politics, conspiracies, culture
zemavek.sk	Health, spiritual, “science”, society

Slovak mainstream news media

Name	Description
dennikn.sk	Politics, science, culture, sport, environmental issues
hnonline.sk	Finance, politics
pluska.sk	Showbusiness, politics, sport. Tabloid
pravda.sk	Politics, finance, sport, culture, auto-moto, health, and cooking
sme.sk	Politics, economy, sport, culture, technologies, health, environmental issues
topky.sk	Politics, showbusiness, leisure. Tabloid
webnoviny.sk	Politics, finance, sport
aktuality.sk	Politics
cas.sk	Showbusiness, politics, sport. Tabloid

Czech partisan news media

Name	Description
ac24.cz	Politics, army, “science”
aeronet.cz	Politics, technologies
ceskoaktualne.cz	Politics, health, culture, finance, army, conspiracies, and factoids
cz.sputniknews.com	Politics
parlamentnilisty.cz	Politics
pravyprostor.cz	Politics, society
protiproud.cz	Politics, health, spiritual, conspiracies
svobodnenoviny.eu	Anti-migrant, conspiracies, mysteries, health
svobodny-vysilac.cz	Politics, spiritual, conspiracies, culture
zvedavec.org	Politics, army, society, anti-LGBTI and migrants

Czech mainstream news media

Name	Description
ahaonline.cz	Leisure, showbusiness. Tabloid
blesk.cz	Leisure, showbusiness, politics. Tabloid
ct24.ceskatelevize.cz	Politics, finance, culture, science, sport
idnes.cz	Politics, finance, culture. Tabloid
ihned.cz	Politics, finance, culture, technologies
lidovky.cz	Politics, finance, culture, leisure
pravo.cz	Politics, leisure, finance
reflex.cz	Politics, interesting facts, culture
respekt.cz	Politics, culture, society

Bibliography

1. Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
2. Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., & Nakov, P. (2018). Predicting Factuality of Reporting and Bias of News Media Sources. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3528–3539. <https://doi.org/10.18653/v1/D18-1389>
3. Baly, R., Karadzhov, G., An, J., Kwak, H., Dinkov, Y., Ali, A., Glass, J., & Nakov, P. (2020). What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 3364–3374. <https://doi.org/10.18653/v1/2020.acl-main.308>
4. Bhatt, S., Joglekar, S., Bano, S., & Sastry, N. (2018). Illuminating an Ecosystem of Partisan Websites. *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, 4, 545–554. <https://doi.org/10.1145/3184558.3188725>
5. Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), 1216–1227. <https://doi.org/10.1002/asi.20077>
6. Cazalens, S., Lamarre, P., Leblay, J., Manolescu, I., & Tannier, X. (2018). A Content Management Perspective on Fact-Checking. *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, 565–574. <https://doi.org/10.1145/3184558.3188727>
7. Chen, Z., & Freire, J. (2020). Proactive Discovery of Fake News Domains from Real-Time Social Media Feeds. *The Web Conference 2020 - Companion of the World Wide Web Conference, WWW 2020*, 584–592. <https://doi.org/10.1145/3366424.3385772>
8. Choraś, M., Demestichas, K., Giełczyk, A., Herrero, Á., Ksieniewicz, P., Remoundou, K., Urda, D., & Woźniak, M. (2021). Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study. *Applied Soft Computing*, 101, 107050. <https://doi.org/10.1016/j.asoc.2020.107050>
9. De Maeyer, J. (2012). The journalistic hyperlink: Prescriptive discourses about linking in online news. *Journalism Practice*, 6(5–6), 692–701. <https://doi.org/10.1080/17512786.2012.667273>
10. Dhoju, S., Main Uddin Rony, M., Ashad Kabir, M., Hassan, N., Kabir, M. A., Rony, M. M. U., Hassan, N., Main Uddin Rony, M., Ashad Kabir, M., & Hassan, N. (2019). Differences in Health News from Reliable and Unreliable Media. *The Web Conference 2019 - Companion of the World Wide Web Conference, WWW 2019*, 981–987. <https://doi.org/10.1145/3308560.3316741>
11. Edward S. Herman, & Chomsky, N. (1988). *Manufacturing Consent: The Political Economy of the Mass Media*. Pantheon Books.
12. Fourney, A., Racz, M. Z., Ranade, G., Mobius, M., & Horvitz, E. (2017). Geographic and Temporal Trends in Fake News Consumption During the 2016 US Presidential Election. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management - CIKM '17*, 2071–2074. <https://doi.org/10.1145/3132847.3133147>
13. Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11), 1129–1164. <https://doi.org/10.1002/spe.4380211102>
14. Guo, B., Ding, Y., Yao, L., Liang, Y., & Yu, Z. (2020). The Future of False Information Detection on Social Media: New Perspectives and Trends. *ACM Computing Surveys*, 53(4), 1–36. <https://doi.org/10.1145/3393880>
15. Himelboim, I. (2010). The international network structure of news media: An analysis of hyperlinks usage in news web sites. *Journal of Broadcasting and Electronic Media*, 54(3), 373–390. <https://doi.org/10.1080/08838151.2010.499050>
16. Hřčková, A., Srba, I., Móro, R., Blaho, R., Šimko, J., Návrát, P., & Bielíková, M. (2019). Unravelling the basic concepts and intents of misbehavior in post-truth society. *Bibliotecas, Anales de Investigacion*, 15(3), 421–428.
17. Jackson, M. H. (1997). Assessing the structure of communication on the World Wide Web. In *Journal of Computer-Mediated Communication* (Vol. 3, Issue 1, pp. 0–0). Wiley Blackwell. <https://doi.org/10.1111/j.1083-6101.1997.tb00063.x>
18. Kotonya, N., & Toni, F. (2020). Explainable Automated Fact-Checking: A Survey. *Proceedings of the 28th International Conference on Computational Linguistics*, 5430–5443. <https://doi.org/10.18653/v1/2020.coling-main.474>

19. Makulová, S., & Hrčková, A. (2011). *Optimalizácia webových stránok pre vyhľadávacie nástroje*. Elet.

20. Morstatter, F., Shao, Y., Galstyan, A., & Karunasekera, S. (2018). From Alt-Right to Alt-Rechts: Twitter Analysis of the 2017 German Federal Election. *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, 8, 621–628. <https://doi.org/10.1145/3184558.3188733>

21. Nakov, P., Sencar, H. T., An, J., & Kwak, H. (2021). *A Survey on Predicting the Factuality and the Bias of News Media*. <http://arxiv.org/abs/2103.12506>

22. Pak, C., Cotter, K., & DeCook, J. (2020). Intermedia Reliance and Sustainability of Emergent Media: A Large-Scale Analysis of American News Outlets' External Linking Behaviors. *International Journal of Communication*, 14, 3546–3568.

23. Park, H. W., & Thelwall, M. (2008). Link analysis: Hyperlink patterns and social structure on politicians' Web sites in South Korea. *Quality and Quantity*, 42(5), 687–697. <https://doi.org/10.1007/s11135-007-9109-z>

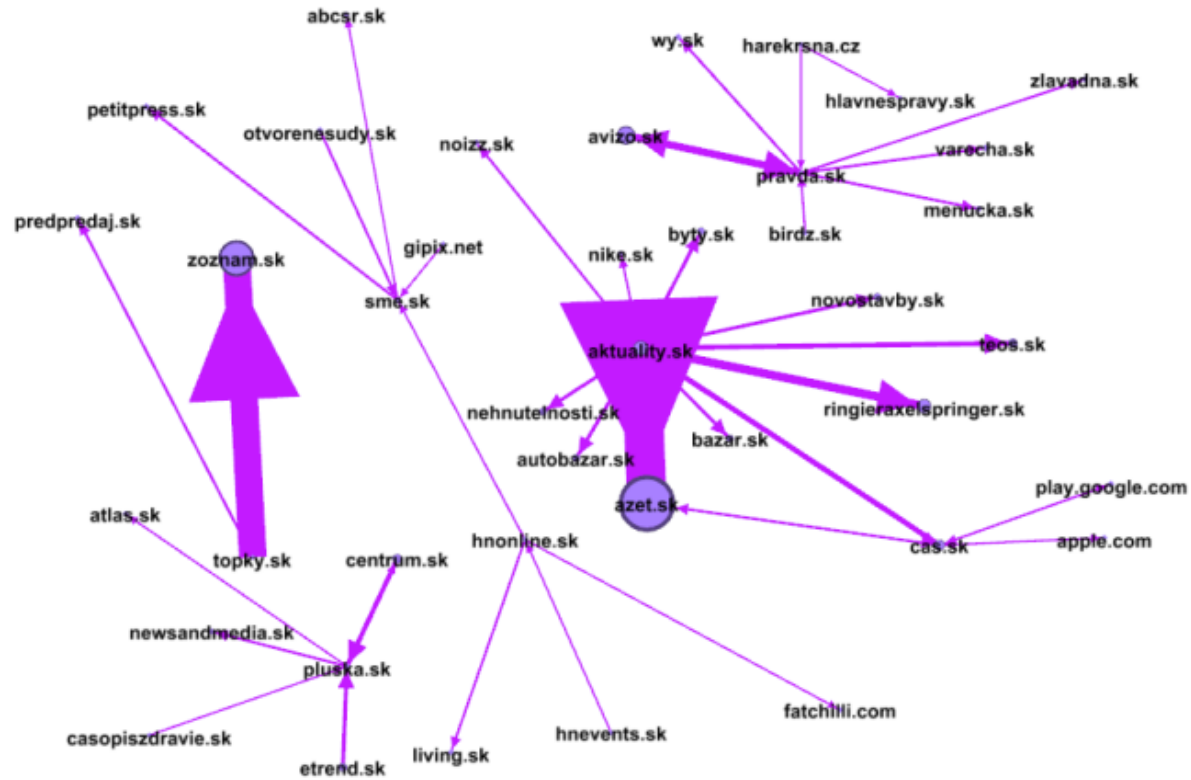
24. Szabó, G., & Bene, M. (2015). Mainstream or an alternate universe? Locating and analysing the radical right media products in the Hungarian media network. *Intersections East European Journal of Society and Politics*, 1(1), 122–146. <https://doi.org/10.17356/ieejsp.v1i1.30>

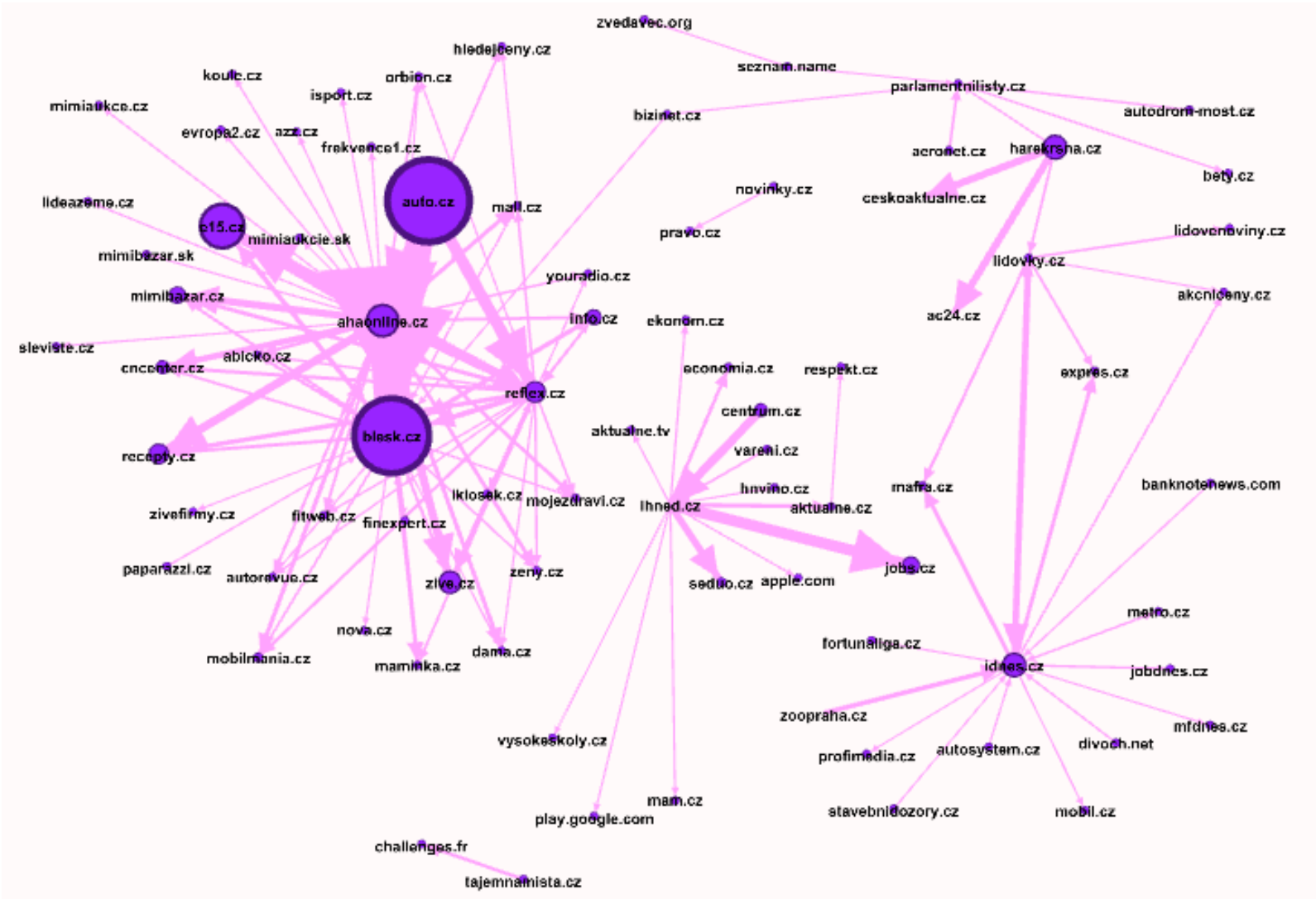
25. Verma, N., Fleischmann, K. R., & Koltai, K. S. (2017). Human values and trust in scientific journals, the mainstream media and fake news. *Proceedings of the Association for Information Science and Technology*, 54(1), 426–435. <https://doi.org/10.1002/pa2.2017.14505401046>

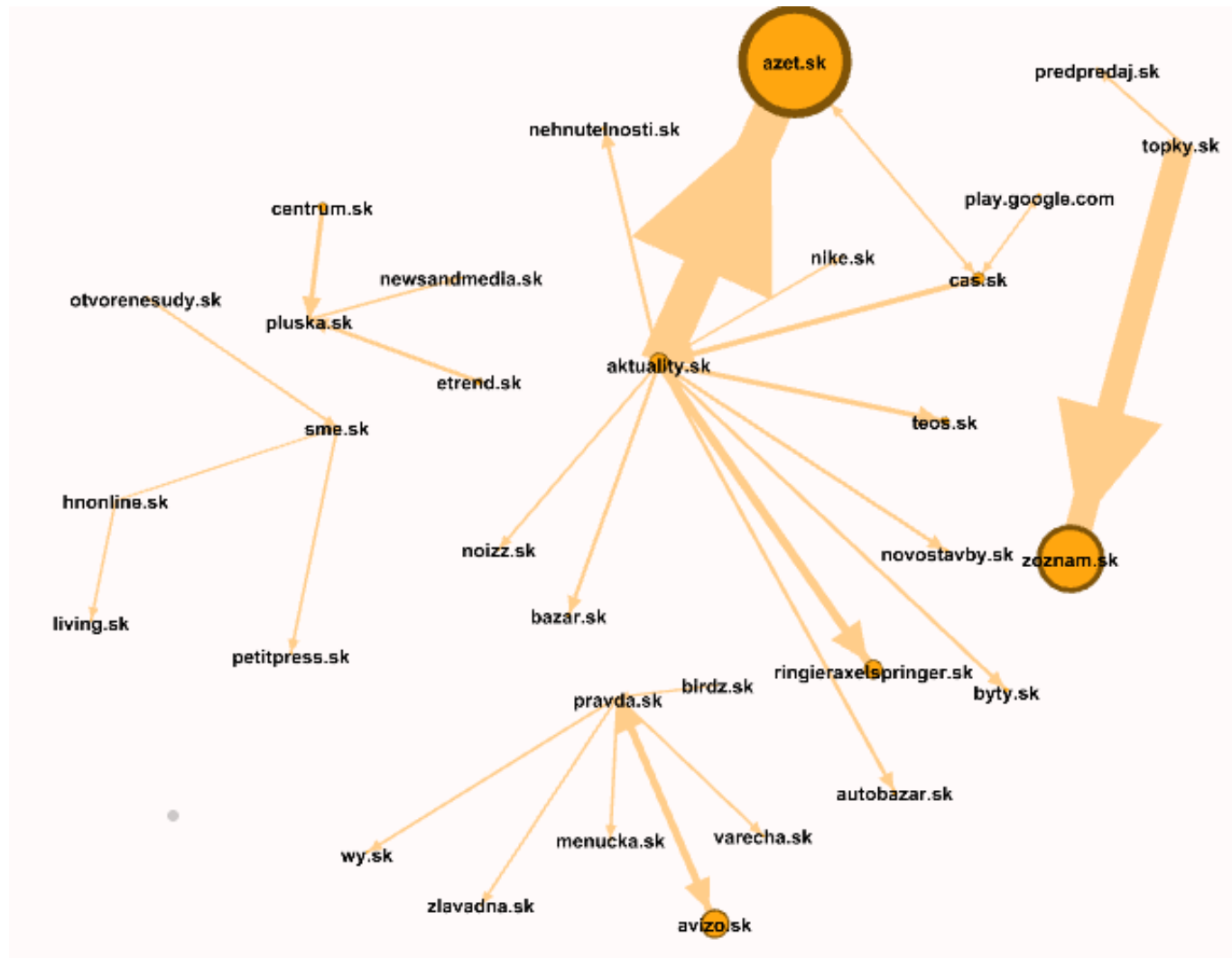
26. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>

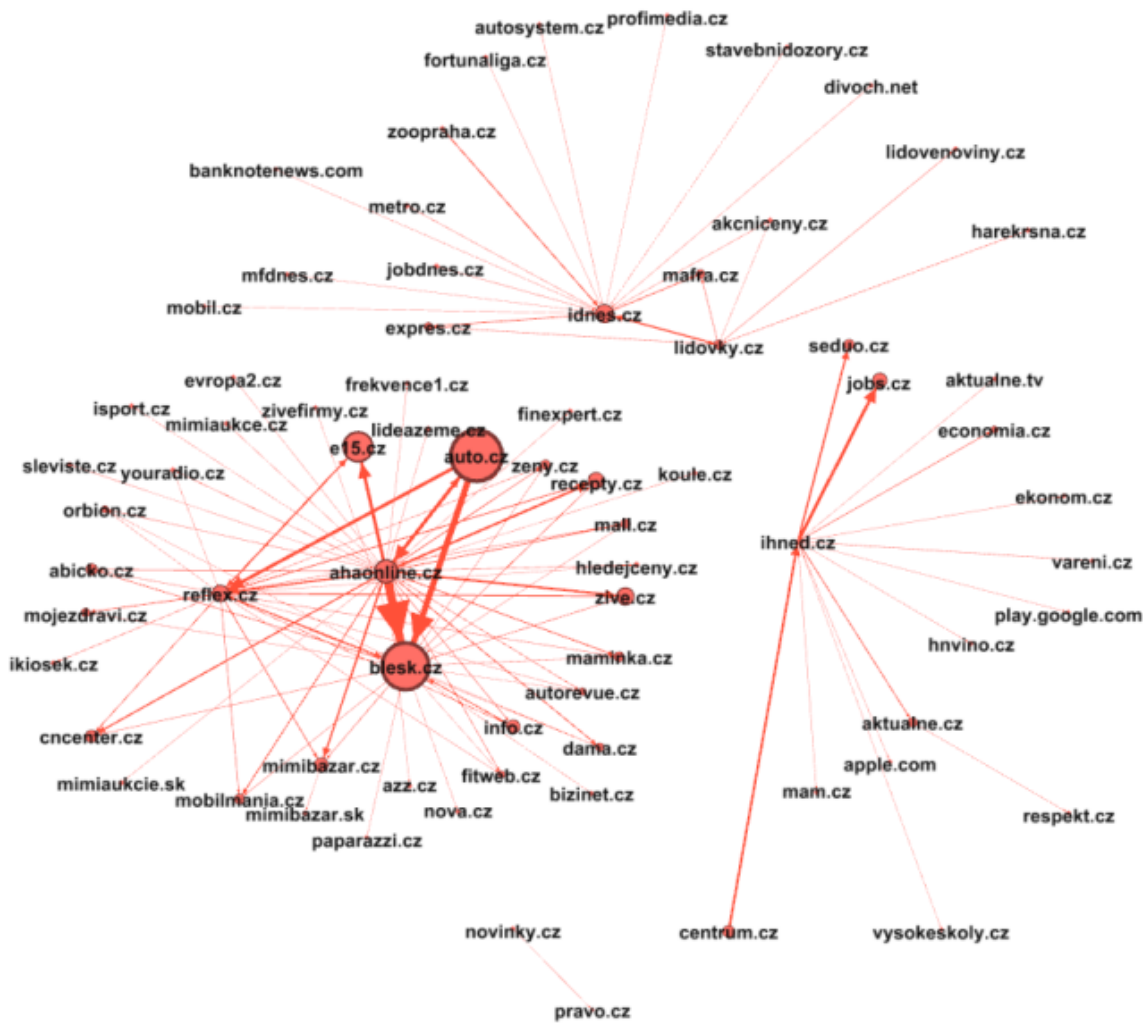
27. Zeifman, I., & Breslaw, D. (2017). *A Closer Look at the Most Active Good Bots*. <https://www.imperva.com/blog/most-active-good-bots/>

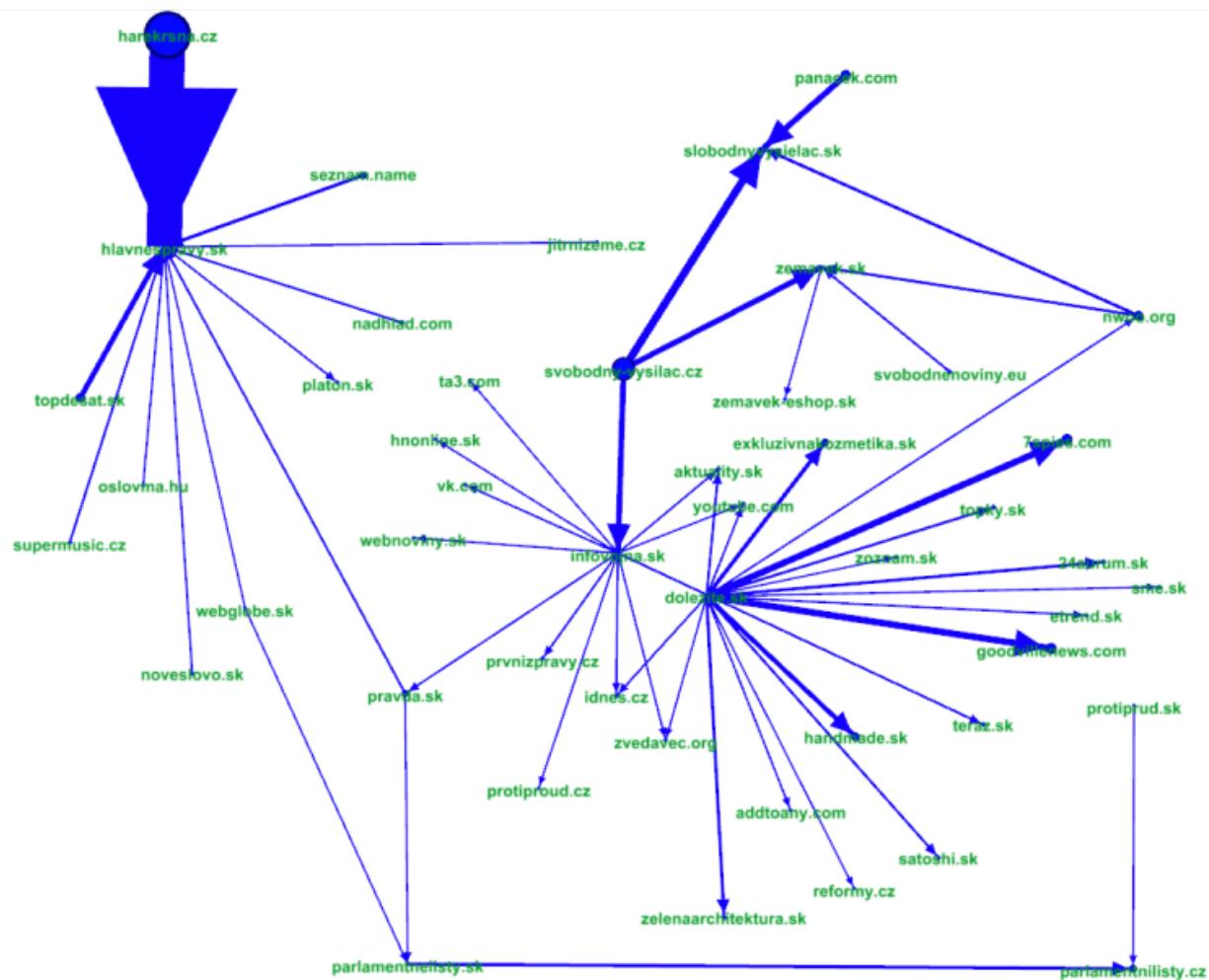
28. Zhou, X., & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys*, 53(5), 1–40. <https://doi.org/10.1145/3395046>

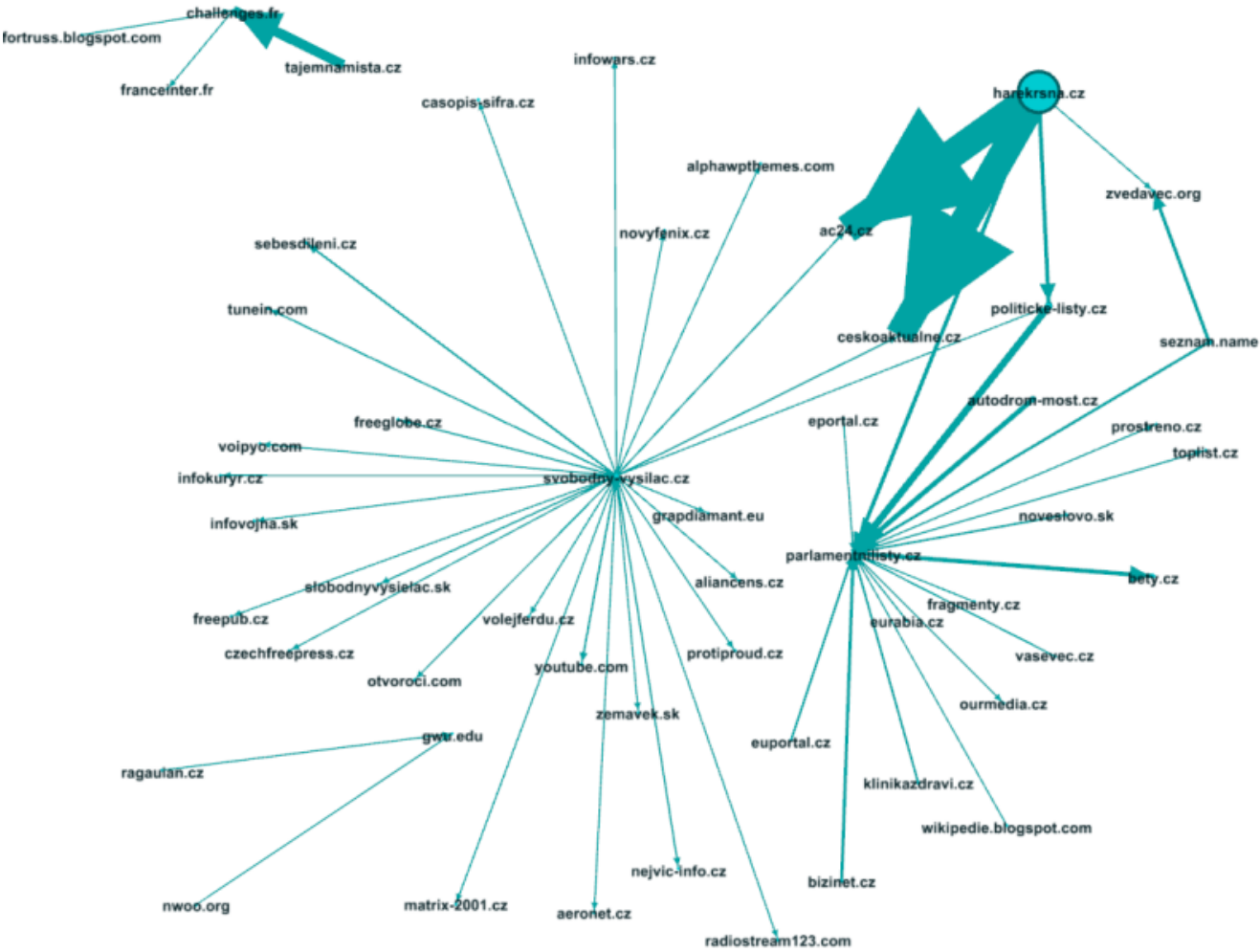


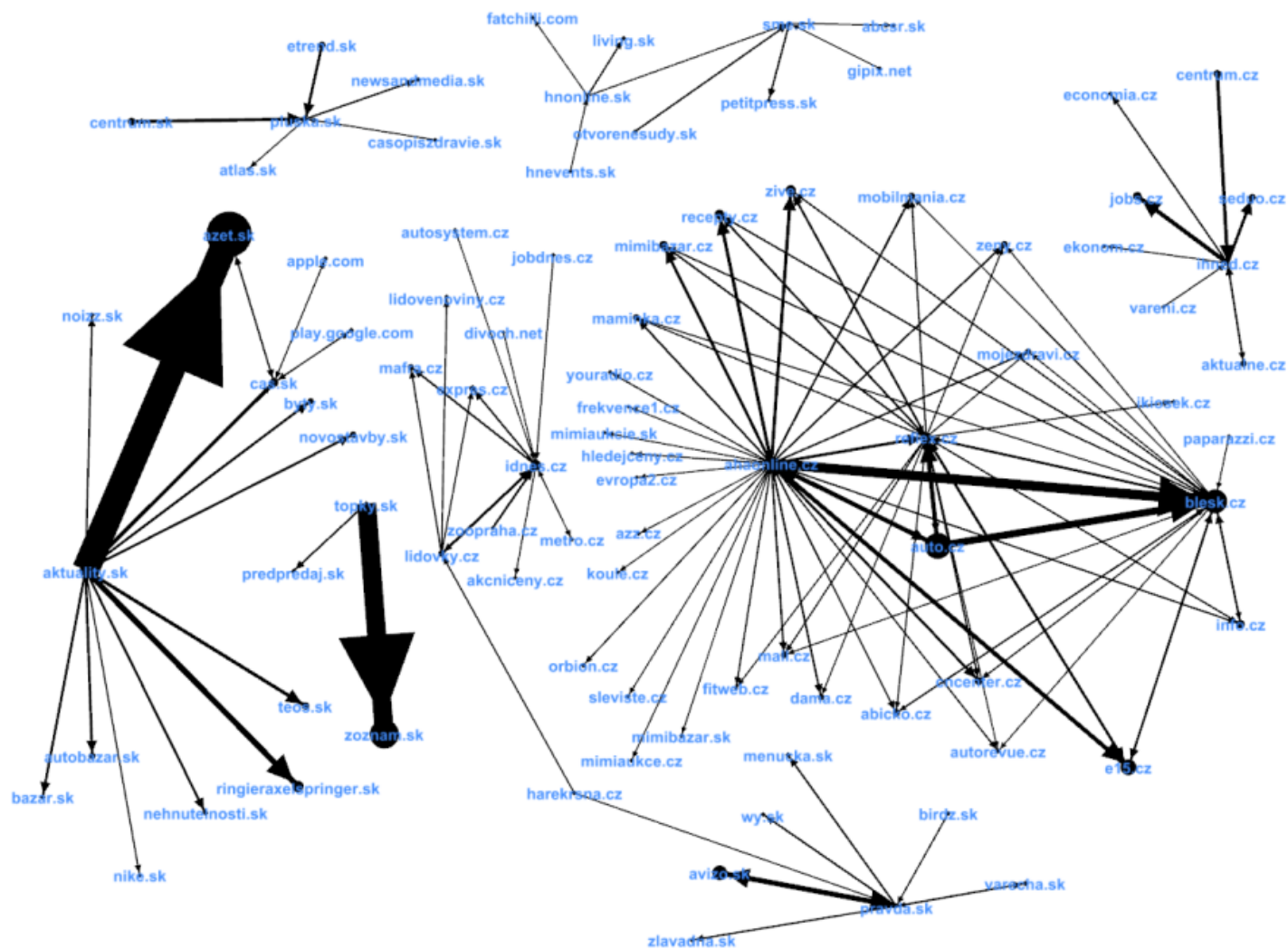


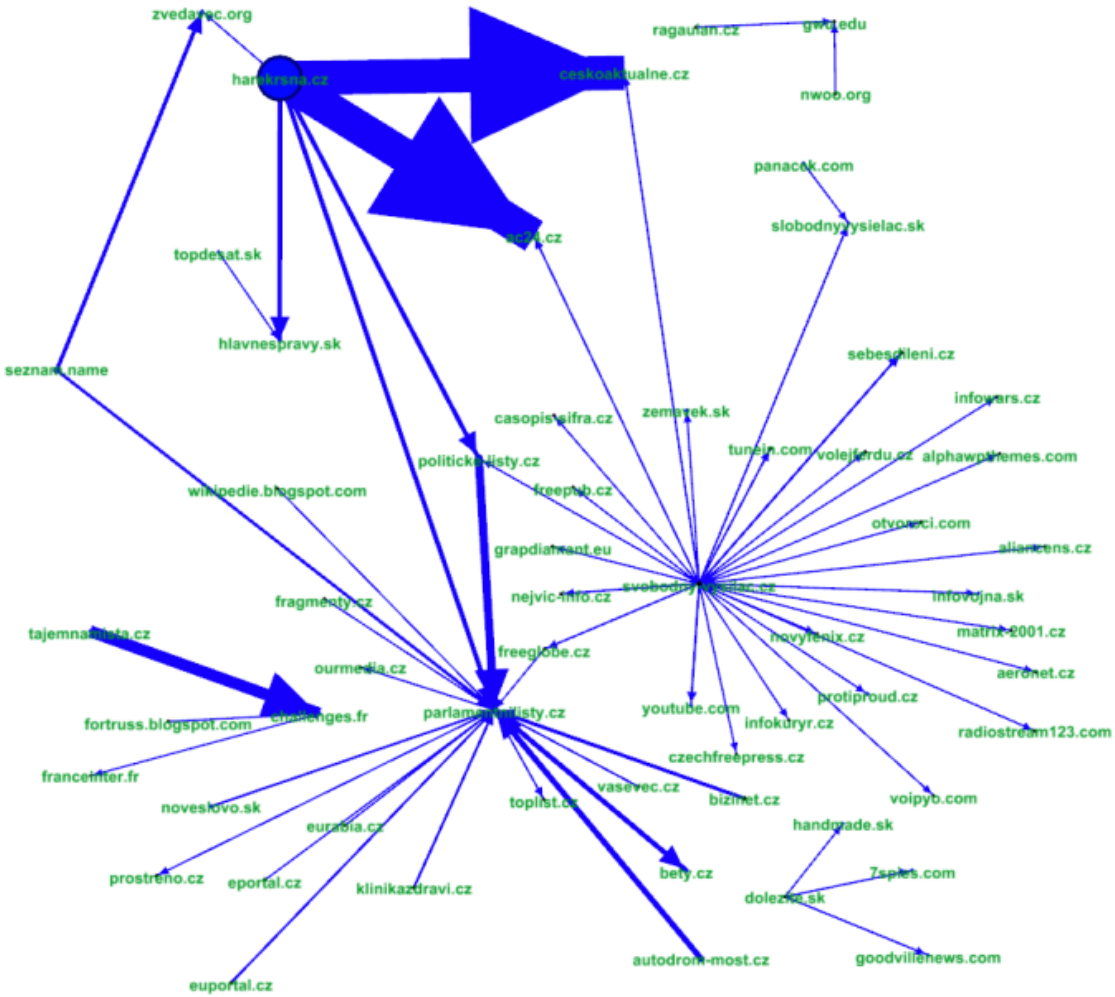


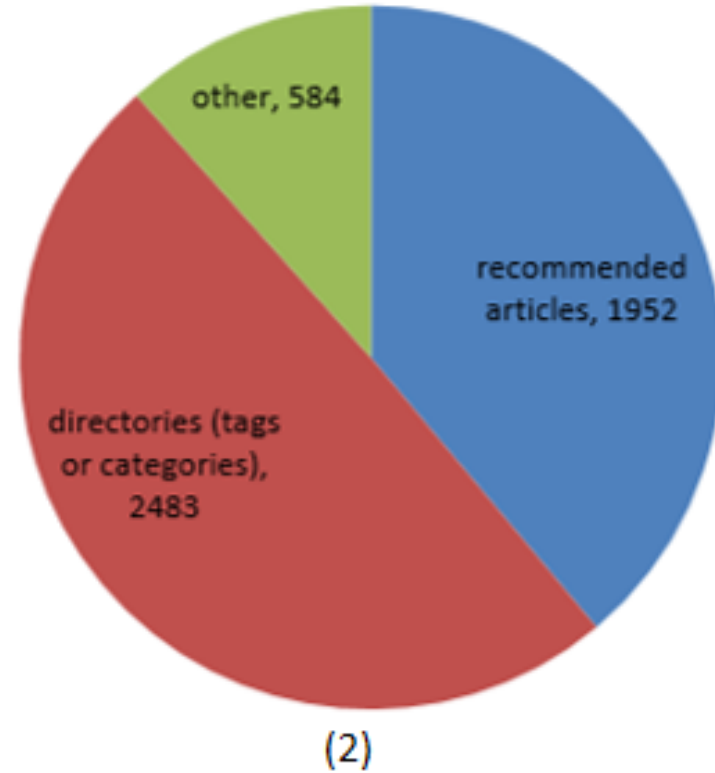












1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41



