



Research article

News consumption patterns on Twitter: fragmentation study on the online news media network

Ford Lumban Gaol^{a,*}, Ardian Maulana^b, Tokuro Matsuo^c^a Computer Science Department, BINUS Graduate Program – Doctor of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia^b Information Systems Management Department, Binus Graduate Program, Bina Nusantara University, Jakarta 11480, Indonesia^c Advanced Institute of Industrial Technology, Tokyo 140-0011, Japan

ARTICLE INFO

Keywords:

Computer science
Network analysis
Media fragmentation
Online news media
Twitter
News consumption

ABSTRACT

The development of news media and social media has radically changed the way the public consumes information. This study explores the structure of online news media networks in three countries, namely Indonesia, Malaysia and Singapore, to investigate the phenomenon of fragmentation in the news consumption pattern on social media. Based on the results of the three network indicators used in this study, it can be concluded that the structure of online news media networks in Indonesia and Malaysia shows a tendency of fragmentation. In contrast, this study did not find sufficient evidence that the phenomenon of fragmentation was occurring in the Singapore media network. In-depth analysis on each formed media cluster shows that online news media in Indonesia and Malaysia tend to group based on similarity in market segments, regions or political alignments.

1. Introduction

The rapid growth of online news media and the presence of social media as a platform for public interaction in the digital space have changed the pattern of news consumption in society [1]. This change raises discourse about fragmentation in the news media landscape as well as in the way people consume information. Fragmentation in media landscape, or often expressed through the metaphor “echo-chambers” and “filter bubbles” [2, 3, 4], refers to situations where audience cluster around the news sources they access [5]. When this is intertwined with an individual tendency for confirmation bias on specific narratives, for example preferences over certain political issues, then the fragmentation of the media landscape can lead to extreme polarization of public opinion [6].

There are many studies that examine the phenomenon of fragmentation in the news media landscape [5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. However, it remains an empirical question to determine whether this is really a characteristic of current news consumption patterns, or whether the extent of fragmentation differs across countries and media platforms. Studies conducted by Mukerjee et al. [5] and Majó-Vázquez et al. [13] using news site traffic data shows that there are massive audience overlap between news outlets, which indicates that the tendency for fragmentation is not as extreme as imagined. Meanwhile, a study of news consumption patterns on Facebook by Schmidt et al. [15] suggests there is a

strong tendency for fragmentation in the news media landscape on Facebook.

In this study we will use network analysis to explore the anatomy of media networks. on Twitter, one of the most popular social media platform. The study goal is threefold: (i) to investigate fragmentation in news media landscape on social media environment. In contrast to the majority of previous studies, we examine the fragmentation tendency in media network using three indicators at once, namely density [8, 17], centrality [5, 7, 8, 13] and network modularity [13]; (ii) to explore the characteristics of media clusters in fragmented network structures. Different with previous research, we conducted an in-depth analysis of each media cluster, and we expected to find relatively homogeneous news outlet characteristics within each cluster; (iii) to compare the characteristic of news media landscape in three countries, namely Indonesia, Malaysia and Singapore. As suggested repeatedly in the literature, comparative analysis between countries with different media environments is needed to avoid making conclusions about various media markets using one single case study [13]. In this study, the three selected countries have different media regulations and journalistic practices [18], and we expected to find substantive and significant differences in the way audiences in these countries consume news online.

In general, this research contributes both methodologies and empirical facts for understanding whether news consumption trends in digital

* Corresponding author.

E-mail address: fgaol@binus.edu (F.L. Gaol).

Table 1. Related works.

No	Ref	Type	Indicator	Result
1	[7]	Single case study	Centralization score	No fragmentation
2	[11]	Single case study	Degree distribution	No fragmentation
3	[8]	Single case study	- Density - Centralization score	Fragmentation
4	[13]	Comparative study	- Centrality - Centralization score - Modularity	No fragmentation
5	[5]	Comparative study	- Centrality	No fragmentation

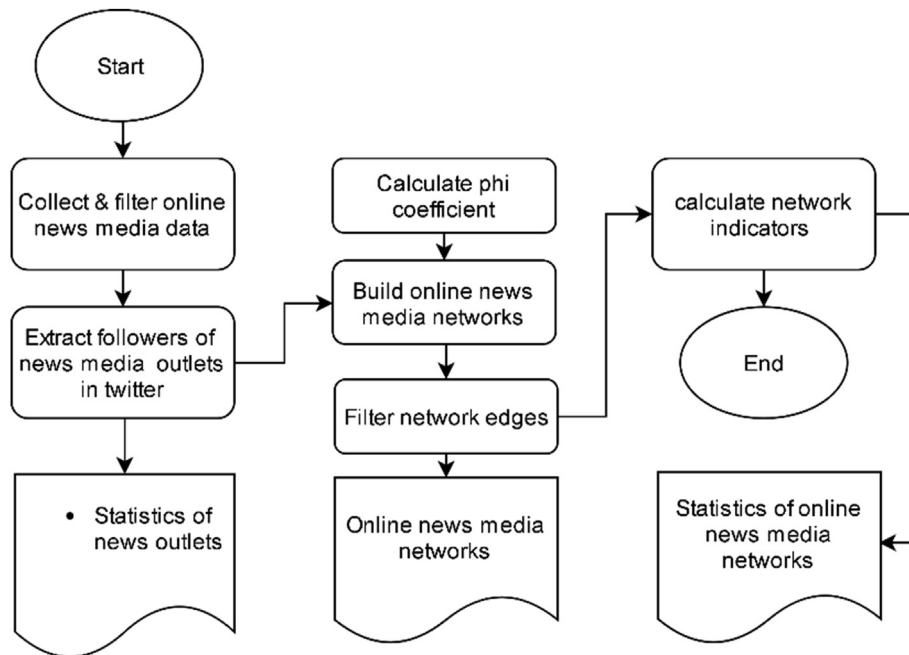


Figure 1. Research methodology.

media environment are accompanied by fragmentation of news sources. Furthermore, the findings in this study present an alternative portrait of the phenomenon of balkanization or echo-chamber, as feared by some [2, 3, 4], which is usually investigated from the perspective of news consumers [3, 19].

The rest of the paper is organized as follows: section 2 discuss some other works that related to this study, section 3 introduces the steps for implementing a network approach to news media follower overlap data. Section 4 presents results of the fragmentation analysis. Some conclusions are drawn in Section 5.

2. Related works

The concept of audience overlap has a relatively long history in media studies, but the potential of overlap data to build network of media was only lately uncovered [5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. Audience overlap can be interpreted as a measure of similarity: the more audience any two news outlets share, the closer those outlets are in terms of their audience base. A collection of interconnected news outlets based on audience similarity collectively forms a media network that represents the news media landscape. At this level, the media network not only informs which news outlets are the central nodes in the media network, but further how these media are grouped into specific clusters. Thus, in contrast to the traditional approach which merely evaluates the increasing number of news sources, the phenomenon of fragmentation in

this analytic framework is investigated based on the behavior of audience in navigating various news sources available online.

The model of media network based on audience overlap data was first proposed by Ksiazek (2011), where the nodes represents media outlets and the level of duplication between their audiences are represented by ties or edges. More recent research has proposed some methodological improvements to the original approach [5, 13]. Different with [7], Mukerjee et al. (2018) proposed phi coefficient as a metric to measure the strength of audience overlap, and applied t-test to filter out insignificant edges between media outlets. Majó-Vázquez et al. (2018) made a methodological contribution by using the disparity filter algorithm [19] to identify the most significant overlap in audience networks. As opposed to a static value of t-test used in [5], this algorithm operates the null model to accept or reject edges based on the distribution of edge weights at the node level. Following these works, Maulana et al. [17] use the phi coefficient and the disparity filter method to build media network, as well as proposing a frameworks, as we will use in this study, for applying a network approach to news media follower duplication data on Twitter.

Network indicators that commonly used in fragmentation studies are density [8], degree distribution [11], centrality [5, 13] and centralization score [5, 7, 13], and network modularity [13]. Table 1 shows that centralization score is the most frequently used to investigate audience fragmentation in media network. In contrast to network density which only provides information about the size of audience duplication, centrality distribution and centralization score considers the configuration of

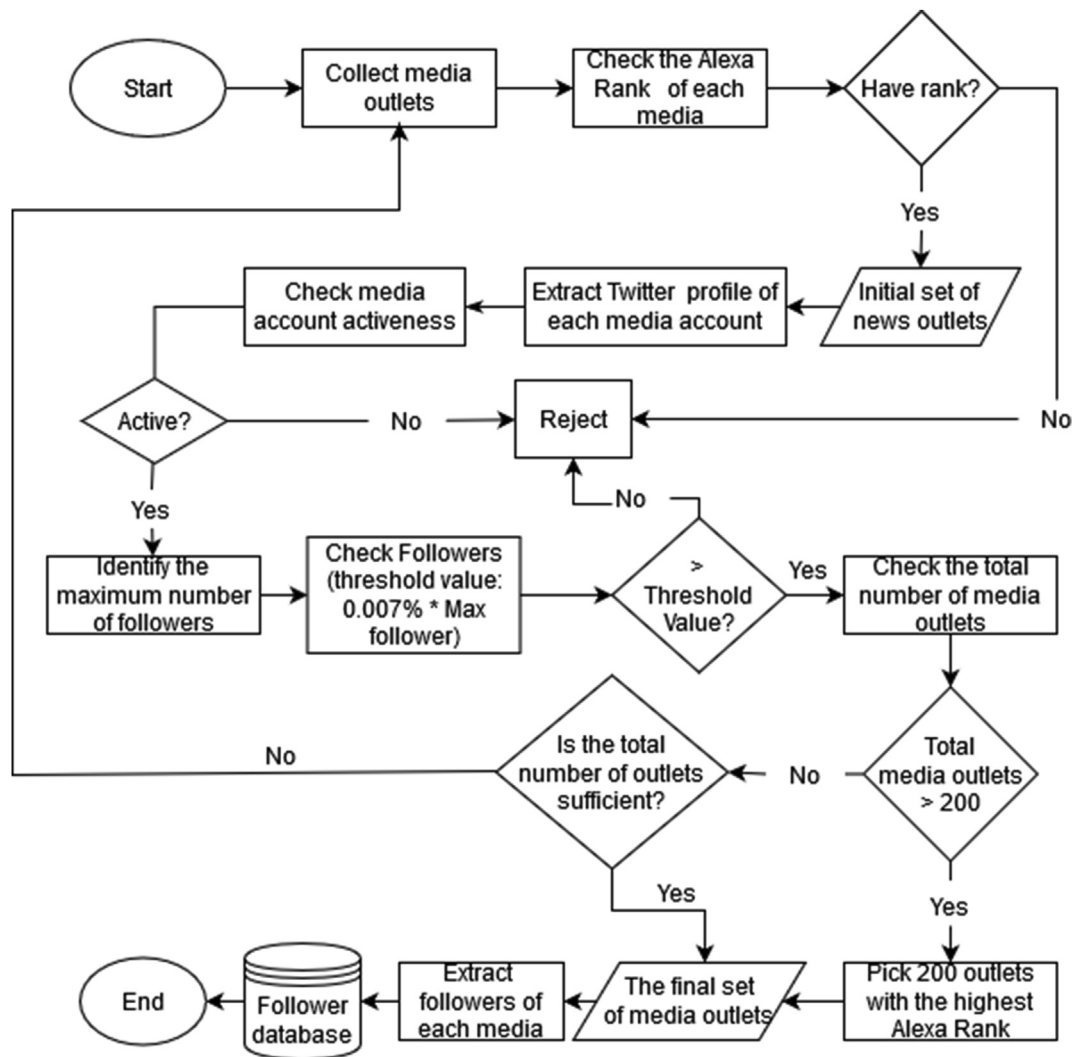


Figure 2. Data extraction flow chart.

Table 2. Statistics of the final list of news outlets.

Country	Number of news outlets	Maximum number of followers in Twitter	Total followers
Indonesia	165	15,070,895	70,480,756
Malaysia	94	1,771,234	13,641,437
Singapore	42	1,008,205	4,590,358

Table 3. Statistics of online news media networks before and after edge filtering. The density value is in the interval [0,1]. Density is the fraction of the number of edges to the total possible edge.

Network Statistics	Edge Filtering					
	Indonesia		Malaysia		Singapore	
	Before	After	Before	After	Before	After
Number of node	165	162	94	86	42	30
Number of edge	12346	754	3668	227	798	46
Density	0.912	0.0578	0.839	0.0621	0.927	0.1057
Minimum Degree	1	1	10	1	19	1
Maximum Degree	163	36	91	15	41	7
Minimum Weight	0.000003	0.0047	0.000008	0.0169	0.000004	0.0151
Maximum Weight	0.5540	0.5540	0.504945	0.5049	0.7014	0.7014

Table 4. The density of online news media networks in Indonesia, Malaysia and Singapore, with value range [0,1].

Indicator	Indonesia	Malaysia	Singapore
Density	0.057818	0.062107	0.105747

Table 5. Comparison of centralization score between countries. The value of network centralization is in the interval [0,1]. The random network is constructed using the edge reshuffling method [30] with the same number of nodes and edges as the observed network ($n_{\text{random}} = 1,000, n_{\text{observation}} = 1000$).

Country	Degree based centralization			z-score
	Observation	Random		
		Mean	St. dev	
Indonesia	0.1679	0.0554	0.0097	11.559
Malaysia	0.1171	0.0842	0.0178	1.848
Singapore	0.145	0.2225	0.0623	-1.239

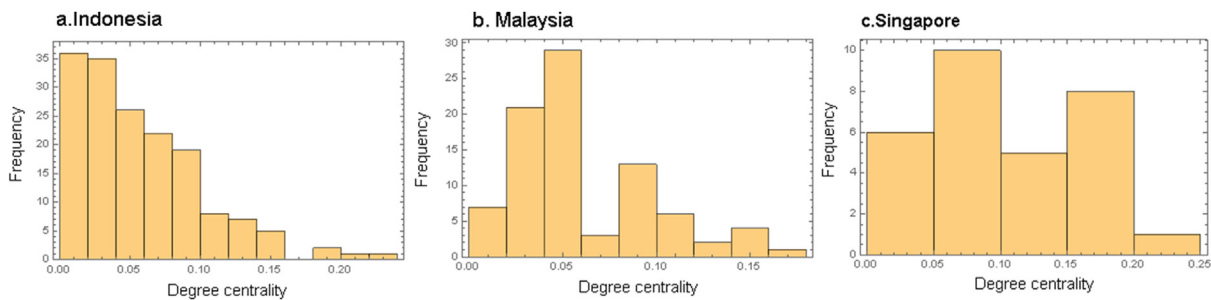


Figure 3. Histogram of media frequency based on news media centrality in 3 countries. The centrality value is normalized into the interval [0,1].

Table 6. Modularity values of Indonesian, Malaysian and Singapore media networks. Modularity optimization using the Louvain method [29] is implemented in the largest connected component network ($V_{\text{indonesia}} = 162, E_{\text{indonesia}} = 754; V_{\text{malaysia}} = 73, E_{\text{malaysia}} = 202; V_{\text{singapore}} = 16, E_{\text{singapore}} = 27$). Random networks are constructed using edge reshuffling method [30] with the number of nodes, edges, densities, degree sequences same as the observed networks ($n_{\text{random}} = 1,000, n_{\text{observasi}} = 1000$).

Country	Modularity Value			Z-score
	Observation	Expectation		
		Mean	St. dev	
Indonesia	$0.53 \pm 3.49E-3$	0.364	0.008	20.51
Malaysia	$0.57 \pm 2.48E-3$	0.461	0.015	7.55
Singapore	$0.209 \pm 4.1E-4$	0.380	0.048	-3.57

Table 7. Summary of measurement results for the three network indicators.

Country	Indicator	Centralization z-score			Modularity z-score	Result
		Density	Degree	Eigenvector		
			Degree	Eigenvector		
Indonesia	0.057818	11.559	8.9517	20.51	Fragmentation	
Malaysia	0.062107	1.848	4.5904	7.55	Fragmentation	
Singapura	0.105747	-1.239	3.1996	-3.57	No fragmentation	

relations between news outlets, and quantifies the tendency of audiences to duplicate disproportionately with a small number of outlets. Modularity is a network indicator that directly measures the fragmented structure due to the presence of node clusters called communities. Majó-Vázquez et al. (2018) use this indicator to investigate fragmentation tendency in US and British media networks.

Most of the studies presented in Table 1 found no evidence of audience fragmentation in the analyzed media network. However, Taneja (2017) who explored network of popular web domains found that the

network was fragmented based on geo-linguistic lines. Mixed results from these studies are closely related to diverse media environments (e.g. regulatory frameworks, journalistic practices etc.), which affect how audiences navigate media landscapes. Thus, to have a better understanding of the structure of online news consumption, fragmentation analysis needs to be done within comparative studies framework (across countries, across demographic groups, and over time) [13].

In contrast to the previous studies, Del Vivario (2017) uses the bipartite network model to observe the spontaneous emergence of well

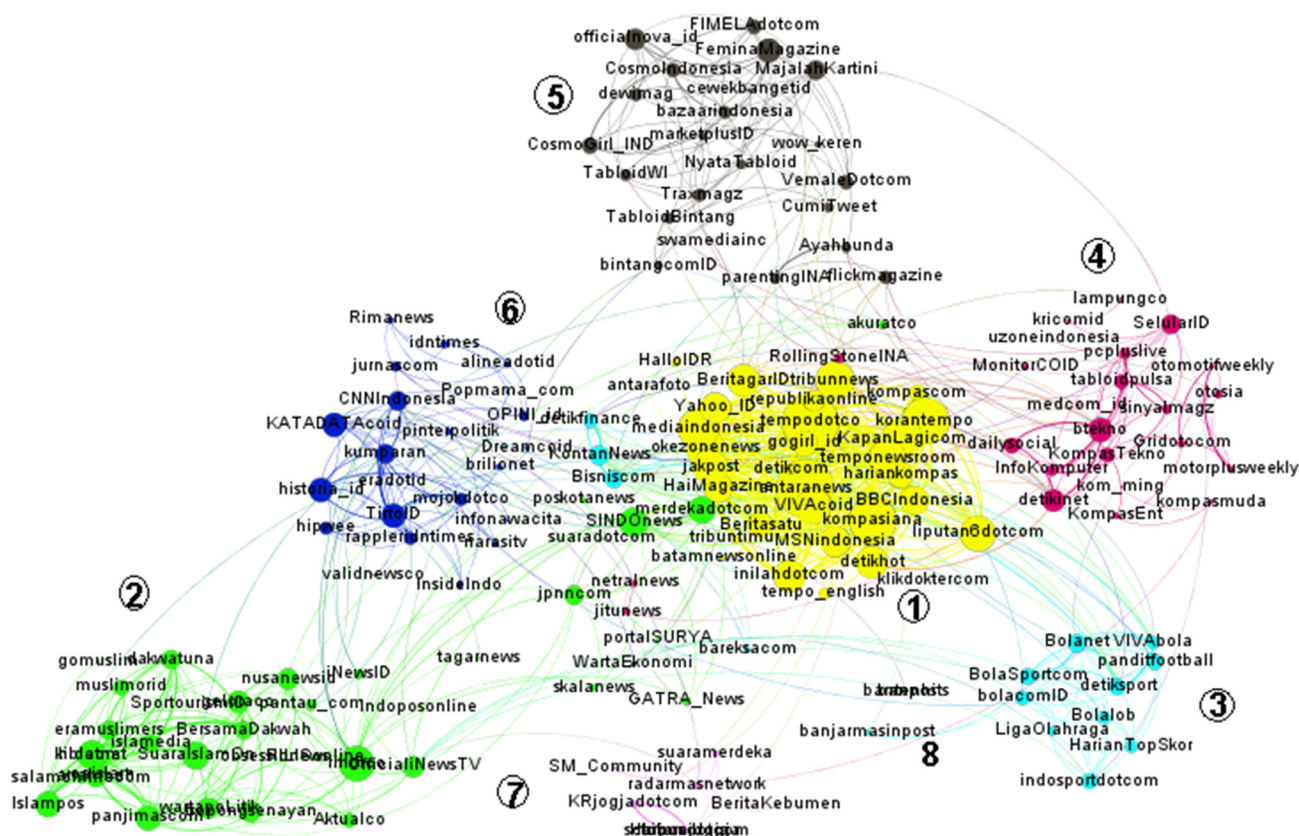


Figure 4. Node Clusters in the Indonesian news media network ($V_{Indonesia} = 162$, $E_{Indonesia} = 754$). Nodes are online news media and edges are overlapping follower of news media on Twitter, and edges weight show the strength of relationships between outlets. The size of the node is proportional to the degree centrality of the media outlet. The colors indicate different cluster of node: cluster 1 (yellow), cluster 2 (green), cluster 3 (cyan), cluster 4 (red), cluster 5 (black), cluster 6 (blue), cluster 7 (pink), cluster 8 (purple). Visualization using Gephi 0.91.

separated communities of news sources, on both Twitter and Facebook, during Referendum event in Italy. By exploring the dynamics behind the discussion, they also find that users tend to restrict their attention to a specific set of Facebook pages/Twitter accounts and thus create a distinct community structure within these news outlets [19]. Combination between fragmented media landscapes and confirmation biases to certain narratives, e.g. preferences over certain political issues/candidate, can lead to extreme polarization of public opinion [6]. Marozzo et al. (2018) proposes a methodology aimed at analyzing the polarization of social network users and news sites during political campaigns characterized by the rivalry of different factions. The methodology permits to analyze what news sites can be considered in favor, against or neutral to a given faction [20]. Recent works [21] provide empirical evidence of the pivotal role of confirmation bias and selective exposure in online social dynamics. This finding challenges the conclusions of Morgan et al. (2013) that users tend to share news without bias toward or against the perceived ideology of news outlets. Referring to the study [25], while selective exposure might be a strong influence at some level, but it does not seem to strongly influence the choice of news outlets.

3. Materials and methods

The research methodology consists of three stages, namely data extraction, modeling & simulation and analysis, with reference to the framework proposed in [17]. As shown in Figure 1, the collection and filtering of media data in the first stage will result in a set of news media and their statistics, which will be analyzed at a later stage. In stage two we identify the relation between news outlets, filter out insignificant relationships and represent them as a news media network. In stage three we calculate a number of network indicators that are relevant to the

purpose of this study. Based on this methodology, we analyze online news media data in Indonesia, Malaysia and Singapore to investigate the tendency of fragmentation in online news media landscape in these three countries.

3.1. Data extraction

Figure 2 shows the extraction process flowchart. The first step is to build a list of news media outlets from Indonesia, Malaysia and Singapore. The initial list was compiled from various sources on internet, where we used Alexa rank [22] to eliminate news outlet whose audience numbers were insignificant. The initial list consisted of 215 Indonesian news media, 111 Malaysian news media and 58 Singapore news media. Then the media statistics (number of posts, number of followers etc.) are extracted from each media's official Twitter account. News media that does not have an active Twitter account and do not have a significant number of followers (follower threshold = 0.007 percent of the largest number of followers in each country) are eliminated from the initial list of news outlets. The follower threshold in each country is as follows: Indonesia = 1054; Malaysia = 123; Singapura = 76. As shown in Table 2, the final list of news outlet contain 165 Indonesian online news media, 94 Malaysian online news media and 42 Singapore online news media. The number of news media in Indonesia is far more than the other two countries. This is not surprising considering the freedom of the press in Indonesia is relatively better than Malaysia and Singapore [18].

Then, based on the final list, the process of extracting followers is carried out to retrieve followers of these news outlets on Twitter. The follower extraction process uses Twarc, the Python package for Twitter data archiving [23]. Based on these data we then build media network using procedure that will be explained next. However, in this study we

Table 8. Groups of news outlets in Indonesian online news media network.

Cluster	Online news media	Segemntation	Percentages (%)
1	tempo, antaranews, Kompas, tribunnews, Kompasiana, antara foto, koran tempo, msnindonesia, beritagar, kapanlagi, viva, okezone, tribun timur, hallo indonesia, batamnews, inilah, detik, media indonesia, harian Kompas, liputan6, hai, yahoo indonesia, tempo english, detikhot, tempo news room, gogirl, the jakarta post, beritasatu, republika, bbc news indonesia, klikdokter	cross-segment of national news media	22.84
2	eramuslim, inews, islampos, gelora news, arrahmah, gomuslim, sportourism, wartapolitik, rmo1, rilis, dakwatuna, obsessionnews, indopos, inews, merdeka, salam-online, harian surya, muslim,voa-islam, hidayatullah, sindonesis, islamedia, poskota news, jpmn, nusaneWS, gatra, skalaneWS, bersama dakwah, suara, pantau, teropongsenayan, suara-islam, tagar news, panjimas, aktual, kiblat, akurat	religious news outlets	19.14
3	bola, bolaSport, bolalob, panditfootball, bisnis, detikfinance, banjarmasin post, bola, indosport, bareksa, vivabola, topskor, konitan news, liga olahraga, detiksport	sports	9.26
4	kompas muda, Kompas minggu, tabloid pulsa, selular, kompastekno, jituNews, dailysocial, medcom, gridoto, motor plus, tabloid otomotif, berita teknologi, Kompas entertainment, rolling stone, monitor, detikinet, sinyal magz, pcpplus live, netralnews, lampung, uzone, majalah infokomputer, kricom, otosia	technology and hobby	14.81
5	Cewek Banget, femina indonesia, tabloid nyata, swa media, majalah kartini, bintang, wanita indonesia, marketplus, warta ekonomi, wowkeren, nova, parenting indonesia, dewi magazine, tabloid bintang, cumicumi, majalahtrax, harper's bazaar, flick magazine, ayahbunda, cosmogirl, fimela, vemale, cosmopolitan	women and entertainment	14.2
6	Kata Data, idn Times tirta, validnews, popmama, era indonesia, historia, pinterpolitik, infonawacita, narasi, dream, brilio, inside indonesia, alinea, opini, Mojok, rapper, crn indonesia, kumparan, hipwee, jurnas, rimaneWS	cross-segment national news media	13.58
7	berita kebumen, suaramerdeka, suara merdeka, harian jogja, kedaulatan rakyat, tribun jogja, solopos, radar banyumas	regional-based media outlets	4.94
8	bantenpos, banten hits	regional-based media outlets	1.23

ignored the fact that the follower data contained a number of bot accounts [24], or some of these follower accounts are no longer active. This becomes a limitation that must be considered in interpreting the results of this study given that the actual size of the media audience will be smaller than the number of followers of the media.

3.2. Network of news media outlet

The number of followers does not give a complete picture of online news media landscape because it does not consider how news outlets are connected to each other. In this study, the relation between two media outlets is determined by the overlap between the two media followers on Twitter [17]. Followers overlap can be interpreted as a measure of similarity: the more followers any two news outlets share, the closer those outlets are in terms of their audience base. The similarity between two media is quantified using the phi coefficient [5, 13], which is an association index between 2 sets of binary variables, for example following or not following a news account on Twitter. The phi coefficient value can be calculated as follows:

$$\varphi_{xy} = \frac{m_{11} - n_{1.}n_{.1}}{\sqrt{n_{1.}n_{.1}(n - n_{1.})(n - n_{.1})}} \tag{1}$$

The value of the phi coefficient is in the interval $[-1, 1]$ where the negative phi value indicates a negative correlation between the two media, a positive value indicates a positive correlation, and a zero value means there is no correlation. The greater the overlap of the two media followers, the more positive the value φ_{xy} means the more similar the reader base of the two media.

The collection of relations between media outlets is then represented as a weighted undirected network $G(V,E,W)$ where outlet i and outlet j ($v_i \in V; i = \{1, \dots, N\}$) will be connected by edge e_{ij} ($e_{ij} \in E$) if the phi coefficient of the two outlets φ_{ij} is positive. In this study, phi coefficient is used as an indicator of the strength of relations between any two news outlets w_{ij} ($w_{ij} \in W$) that is connected by edge e_{ij} . As a correlation network, the news media network will have a dense structure in which not all edges have a statistically significant weight. Therefore we need to carry out the process of sparsification to bring out the actual structure so that the media network can be analyzed reliably.

A weighted graph can be easily reduced to a sub-graph in which any of the edges' weight is larger than a given threshold. This global weight threshold technique has been applied in [5] using the t-test and in [16] using simple strategy: choose the strongest one. The short come of this method is that it overpass the nodes with small strength. In this study we use the Disparity Filter [26] method to eliminate relationships whose weights are smaller than what is expected under random conditions. The final network will have a smaller number of edges but have the same multi-scale structural properties as the original network [26]. Table 3 shows statistics of online news media networks in the three countries analyzed. On average, about 94% of all original edges were successfully removed from the three networks analyzed, i.e., 11,592 edges on the Indonesian media network, 3441 edge on the Malaysian media network, and 752 edges on the Singapore media network. As a result, as shown in Table 3, the density values of the three analyzed networks are very small, indicating that the final networks have a sparse structure.

3.3. Network indicators

In this study we will use three network indicators to evaluate the phenomenon of fragmentation in media networks, namely:

- Density

Network density is the ratio between the number of edges that exist in the network with the total possible edge, as follows [27]:

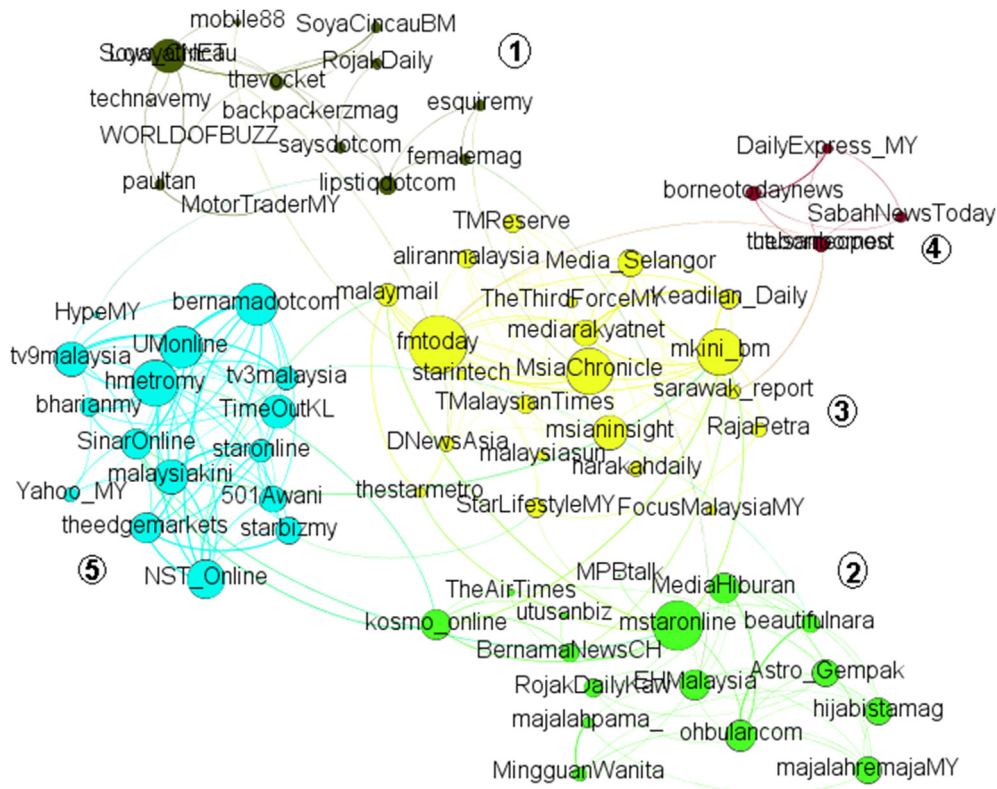


Figure 5. Node clusters in Malaysian online news media network ($V_{\text{malaysia}} = 73$, $E_{\text{malaysia}} = 202$). Nodes are online news media, edges are overlapping media followers on Twitter, and edges weight show the strength of relationships. The size of the node is proportional to the degree centrality of the node. The colors indicate different node cluster: cluster 1 (black), cluster 2 (green), cluster 3 (yellow), cluster 4 (red), and cluster 5 (cyan). Visualization using Gephi 0.91.

$$D_G = \frac{2|E|}{(|V| - 1)} \tag{2}$$

where $|E|$ is the number of edges and $|V|$ is the number of nodes in the network. The density value is in the interval $[0, 1]$, where the higher the value, the denser the structure of a network and the more massive the followers overlap between media.

• Centralization

Centralization score measure how unequal the distribution of node centrality in the network [28]. We use degree based centrality, that is the number of nodes connected to node i , to measure centrality of the position of a media within media network. Formally the centralization indicator is defined as follows:

$$C_G = \frac{\sum_{i=1}^N C_x(v^*) - C_x(v_i)}{\max \sum_{i=1}^N C_x(v^*) - C_x(v_i)} \tag{3}$$

$$C_x(v_i) = \sum_{j=1}^n e_{ij} \tag{4}$$

In which $C_x(v_i)$ is the centrality value of node i , i.e. the number of nodes connected to node i , $C_x(v^*)$ is the largest centrality value in the network. The value of C_G is in interval 0 and 1, where $C_G = 0$ when all nodes have the same centrality value, and $C_G = 1$ when one or a number of nodes are very dominant relative to other nodes.

• Modularity Value

In network analysis, the indicator commonly used to identify the presence of clusters of node is modularity [27]. Modularity value indicates how separate the node clusters are from each other, by comparing the number of edges in a cluster node (actual number of edges) with the expected value at the randomly distributed edge (expected number of edges). Formally, modularity is defined as follows:

$$Q = \sum_i (e_{ii} - a_i^2) \tag{5}$$

where $a_i = \sum_j e_{ij}$ j is the fraction of edges incorporated in cluster i .

Modularity values are in the interval $[0, 1]$. The greater the modularity value means the greater the difference between the actual value and the expected value of edges number in each cluster, the clearer the boundaries between the clusters, which means the more fragmented the network structure. High modularity networks have high interconnection between nodes in the same cluster and relatively do not have strong connectivity between nodes in different clusters. In this study we use the Louvain method to find optimal modularity values [29] from the largest connected component of the online news media network.

In this study, the modularity indicator is implemented in weighted graphs of the news media network, while density and centrality/centralization indicators are calculated from unweighted media networks. Centralization score and modularity value are very dependent on the pattern of relationships in the network. As a result, comparisons of

Table 9. Groups of news outlets in Malaysian online news media network.

Clusters	Online news media	Segment	Percentage (%)
1	Motor Trader MY, soyacinciau (bm), soyacinciau (bn), soyacinciau (bn), world of buzz, latest on says, mobile88, vocket, lipstick, female malaysia, backpackerz, rojak daily, esquire malaysia, paul tan, lowyat	technology and hobby	20.55
2	media hiburan, mstar online, mingguan wanita, MPBtalk, obbulan, kosmo online, astro gempak, berrama news channel (bnc), rojak daily kawi, eh! malaysia, air times news network, remaja, beautifulnara, utusanbiz, pa&ma, hijabista magazine	entertainment	21.92
3	malaysia chronicle, malaysia kini (bm), in.tech, raja petra kamarudin, the malaysian reserve, free malaysia today, star life style, the malaysian insight, the malaysian times, the third force, malaysia sun, media selangor, digital news asia, media rakyat, aliran, keadilan daily, the star metro, sarawak report, harakahdaily, malay mail, focus malaysia	cross-segment national news media	28.77
4	Utusanbomeo, dailyExpress_MY,sabahNewsToday, theborneopost,borneotodaynews	regional-based media outlets	6.85
5	hmetro, tv9 malaysia, Yahoo Malaysia, HypeMY, tv3malaysia, malaysiaikini, nst, thestar, utusanonline, bernamadotcom, SinarOnline, TimeOutKU, 501Awani, theedgemarkets, bharian, staronline	cross-segment national news media	21.92

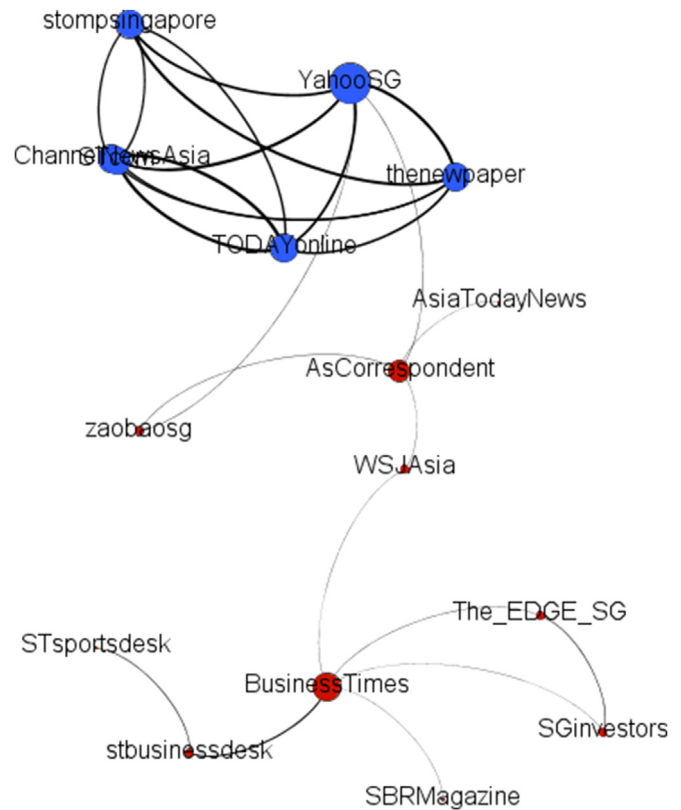


Figure 6. Node Clusters in Singapore online news media network ($V_{Singapore} = 16$, $E_{Singapore} = 27$). Nodes are online news media, edges are overlapping media followers on Twitter, and edge weight show the strength of relationships. The size of the node is proportional to the degree centrality of the node. To clarify the boundaries between clusters the following colors are used: cluster 1 (blue), cluster 2 (red). Visualization uses Gephi 0.9.

those two indicator values from different empirical networks with different sizes cannot be carried out directly. This study uses the z-score to evaluate the significance of the measured values, namely how far the observational values differ from the average value on random networks with same size. We construct random networks using the edge reshuffling method of the observed network [30].

4. Results

4.1. Fragmentation analysis

From the side of information demand or from the audience side, fragmentation in media networks is related to readers' efforts to expand their reach to digitally available information sources. The more number of news outlets that audience follow in Twitter, the greater the number of overlapping followers among news media outlets, and that means the smaller the tendency for fragmentation in the network. This happens because overlapping followers will produce a dense network structure where news outlets are connected to each other.

Table 4 shows the density of media networks in the three countries analyzed having relatively small values. This means that in general media networks in these three countries have a sparse structure, in which there is no massive overlap between media audiences. Furthermore, based on comparison of density values between countries, it can be seen that the media networks in Indonesia and Malaysia have a structure that is more sparse than the Singapore media network. This shows that the attention of media audience in both countries, does not spread widely to many news sources available on Twitter.

Low density networks have a great chance of having fragmented structures [8, 13]. But it really depends on the pattern of relationships on

Table 10. Groups of news outlets in Singapore media network.

Cluster	Online news media	Segment	Percentage (%)
1	Yahoo Singapura, The new paper, Channel News Asia, Straits Times Online Mobile (stomp) singapore, today online, Strait Times	Cross-segment news outlets	37.5
2	Business Times, Strait Times sportsdesk, the edge singapura, zao bao, AsCorrespondent, SBR Magazine, SG investors, Asia Today News, straits times businessdesk, WSJ Asia	Financial news outlets	62.5

the network. A network structure in which the pattern of relations between nodes is evenly distributed, i.e. nodes have a relatively equal number of relations, tend not to experience fragmentation compared to networks with the distribution of relations that are concentrated in a small number of nodes. The number of relations, or degree of nodes, is an indicator commonly used to measure how central the position of a node in a network is. In the context of media networks, the greater the degree value of a media node means that the more numbers of other outlets have common followers with the media, which means the more connected and more strategic the position of the media in the network.

The centralization score is a value that summarizes all information related to the centrality of the nodes in the network. Table 5 shows the z-score of centralization for the three countries analyzed. From this table, it is known that the tendency for centralization is only found in Indonesian and Malaysian media networks. While the Singapore media network shows no centralized tendency at all: the centralization score in the observed network is lower than the centralization score on the random network.

The centralization of the Indonesian media network is far greater than the other two countries. This indicates a relatively massive concentration of relations in a small number of media nodes in Indonesia. This can be seen clearly from the distribution pattern of centrality values in Figure 3. For Indonesia and Malaysia, the histogram of media frequency based on the number of relations tend to be skewed right with long tail distribution. This means that the majority of media has a low centrality value where the nodes are only connected with a handful of other media nodes. Conversely, a small amount of media has a very large number of relations, which means it has a heterogeneous reader base. Compared to the two countries, it is clear that the distribution of relations in the Singapore media network is almost evenly distributed.

The structural characteristic of a network that directly indicates fragmentation is the formation of node clusters in the network structure. The main characteristic of group of nodes is the existence of a strong connection between nodes in the same group (intra-cluster density) compared to relations between nodes of different groups (inter-cluster density) [27]. This structural pattern can be identified using modularity indicators. Fragmented networks will have a modular structure where there is high interconnection between nodes in the same cluster and relatively lack of strong connectivity between nodes in different clusters.

Table 6 shows the modularity and z-score values of the three countries analyzed. From the Table 6 we know that Indonesian and Malaysian media networks have a high modularity value. This indicates that media outlets in media networks in the two countries are grouped in a number of clusters with clear boundaries. The modularity value the two countries is also significantly different from the random conditions, which are 20.51 (Indonesia) and 7.55 (Malaysia) times the standard deviation above the average value of random network modularity with the same structure. Thus it can be said that the news media landscape in Indonesia has a very strong tendency of fragmentation, where media outlets are grouped based on the similarity of the reader base.

Table 6 also shows that the z-score value of Singapore media network modularity has a negative value, which is 3.57 times the standard deviation lower than the average value of random network modularity. This means that random network structures are far more modular than observed network structures. This finding reinforces what is indicated by indicators of centrality and network density that the phenomenon of

fragmentation does not occur in the online media landscape in Singapore.

Based on the values of the three indicators above we can say that the news media landscape in Indonesia and Malaysia is fragmented with the following characteristics: sparse network (low density), relation is concentrated in a small number of news sources and has a modular and fragmented structure. By observing the z-score value in Table 7 we can conclude that the tendency for fragmentation in the media landscape in Indonesia is far greater than the other two countries. Meanwhile, despite having a sparse structure (low density), low centralization and modularity value indicate that fragmentation phenomenon does not occur in Singapore news media network. A negative z-score for both indicators shows that random network structures are far more fragmented than the observed network. The difference shown by the Singapore media network of course related to the characteristics of Singapore media environment which is tightly controlled by government regulations. In addition, Singaporeans tend to be audience of global news outlets that are not included in the scope of this study.

4.2. News media cluster

Figures 4, 5, and 6 shows a visual representation of node groups formed in the three country media networks. The list of media within each clusters is shown in Tables 8, 9, and 10. From this figure, it can be seen that the boundaries between clusters are quite firm, where the density of relations within each cluster is relatively denser than relations between clusters. With such structural conditions it can be understood why the network modularity of the two countries is relatively high.

Various studies have shown that cluster nodes are networks sub-structure that often related to certain functions or categories. For example, node clusters in the World Wide Web network refer to a collection of Web pages on the same topic [31] while node clusters in biological networks are related to certain biological functions [32]. In this study we find that node clusters in media networks generally have homogeneous characteristics, referring to media with the same market segments, regions or political alignments. The online news media landscape in Indonesia, as shown in Figure 4 and Table 8 consists of 8 clusters in which the 2 biggest clusters, namely cluster 1 (22.84% of total media) are dominated by cross-segment national news outlets, and cluster 2 (19.14%) are dominated by news outlets with religious segmentation. In addition to cluster 1, cross-segment national news outlets are also in cluster 6 with a media percentage of 13.58 percent of the total media. The main difference between cluster 1 and cluster 6 is on the dominant type of media, where cluster 1 contains a lot of conventional national media (legacy news media), while cluster 6 is generally digital-born news media. News media with entertainment and lifestyle segmentation form 3 groups, namely cluster 3 (sports segment), cluster 4 (technology segment), and cluster 5 (female segment). The other two clusters are regional-based media groups, cluster 7 (Central Java region) and cluster 8 (Banten region).

Figure 5 shows the structure of the Malaysian online news media network which is composed of 5 media clusters. The list of news outlets within each clusters is shown in Table 9. As in Indonesia, Malaysian news media with national and cross-segment categories are grouped into two clusters, namely cluster 3 (28.77% of total media) and cluster 5 (21% of total media). Both clusters can be characterized by the political

associations of these media outlets. News media in cluster 5 such as *Harian Metro* (hmetro), *News Straits Times* (NST), *The Star* (thestar and staronline), *Utusan Malaysia* (utusanonline), and *Berita Harian* (bharian) are known to have political affiliation with the *Barisan Nasional* political alliance, namely political coalition of current government opposition parties (Wikipedia, 2019). On the other hand, the media which are included in cluster 3 are dominated by media that are politically opposed to the *Barisan Nasional* coalition such as *Keadilan Daily* (Malaysian Islamic Party media), *Harakah Daily* (Keadilan Rakyat Party's media), and a number of independent media. Other media clusters in the Malaysia news media news landscape are news outlets with technology segment (cluster 1: 20.55% of total media), entertainment segment (cluster 2: 21.92% of total media), and regional news media segment (cluster 4: 6.85% of the total media).

Figure 6 visualizes media clusters formed in the Singapore media network. In general, cluster 1 consists of major media outlets in Singapore and can be categorized as cross-segment news media. While in cluster 2 there are a number of news media with economic segments such as the *Bussines Times*, *World Street Journal Asia*, SG investors and a number of business and sports rubrics from the *Straits Times*.

5. Conclusions

This research explores the structure of online news media networks in three countries, namely Indonesia, Malaysia and Singapore, to investigate the phenomenon of fragmentation in news consumption patterns on social media. For this purpose, we model relationship between news outlets based on follower overlaps as news media networks, and implement network indicator such as density, centralization and modularity values to examine the fragmentation tendency in these networks.

Density analysis shows that media networks in these three countries have sparse structures. Low density values indicate that the audience of each news outlet on Twitter does not overlap massively with each other. Based on the indicators of centralization it is known that the media network in Indonesia has the strongest tendency to centralize, and then followed by the Malaysian media network. Furthermore, the clustering analysis shows that the Indonesian and Malaysian media networks have a modular structure with a positive z-score modularity. This indicates that news outlets in the two countries grouping in specific media clusters with clear boundaries between clusters. The opposite happened in the Singapore media network. The structure of the Singapore media network shows no centralized tendency at all, where the number of relations each media has is relatively even. The z-score of the network modularity is also negative, indicating that there is no grouping of media outlets in the Singapore media network. Based on the results of the three network indicators used in this study, it can be concluded that the structure of online news media networks in Indonesia and Malaysia shows a tendency of fragmentation. In contrast, this study did not find sufficient evidence that the phenomenon of fragmentation was occurring in the Singapore media network.

Based on a more in-depth analysis on each formed media cluster, this study identifies the characteristics of each media cluster, which is composed of relatively homogeneous news outlets. Online news media in Indonesia and Malaysia tend to group based on similarity in market segments, regions or political alignments.

In general the study of news consumption patterns on Facebook by Schmidt et.al (2018) also shows the same thing as found in this study, that is news consumption patterns on social media are fragmented based on the sources of information that are followed. Apart from variations in regulations and various media environments between countries, social media environmental factors certainly play a role in strengthening the phenomenon of fragmentation in news consumption patterns. As we know that social media platforms encourage users

with the same preferences and opinions to meet and network with each other.

Declarations

Author contribution statement

F. L. Gaol: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

A. Maulana: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

T. Matsuo: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

This research is supported by Department of Information Systems Management, Bina Nusantara University, Doctor of Computer Science, Bina Nusantara University and Advanced Institute of Industrial Technology, Tokyo, Japan.

References

- [1] N. Newman, R. Fletcher, A. Kalogeropoulos, D. Levy, R.K. Nielsen, Reuters institute Digital News Report 2017, Reuters Institute for the Study of Journalism, 2017. Available online: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf. (Accessed 27 June 2019).
- [2] E. Pariser, *The Filter Bubble: How the New Personalized Web Is Changing what We Read and How We Think*, Penguin Books, London, UK, 2011.
- [3] C.R. Sunstein, *Republic.com 2.0*, Princeton University Press, Princeton, USA, 2009.
- [4] J. Turow, *Breaking up America: Advertisers and the New media World*, University of Chicago Press, Chicago, USA, 1998.
- [5] S. Mukerjee, S. Majó-Vázquez, S. González-Bailón, Networks of audience overlap in the consumption of digital news, *J. Commun.* 68 (2018) 26–50.
- [6] A. Bessi, F. Zollo, M. Del Vicario, A. Scala, G. Caldarelli, W. Quattrociocchi, Trend of narratives in the age of misinformation, *PLoS One* 10 (8) (2015), e0134641.
- [7] T.B. Ksiazek, A network analytic approach to understanding cross-platform audience behavior, *J. Media Econ.* 24 (2011) 237–251.
- [8] H. Taneja, Mapping an audience-centric world wide web: a departure from hyperlink analysis, *N. Media Soc.* 9 (2017) 1331–1348.
- [9] H. Taneja, J.G. Webster, How do global audiences take shape? The role of institutions and culture in patterns of web use, *J. Commun.* 66 (2016) 161–182.
- [10] H. Taneja, A.X. Wu, Does the Great Firewall really isolate the Chinese? Integrating access blockage with cultural factors to explain web user behavior, *Inf. Soc.* 30 (2014) 297–309.
- [11] J.G. Webster, T.B. Ksiazek, The dynamics of audience fragmentation: public attention in an age of digital media, *J. Commun.* 62 (2014) 39–56.
- [12] S. Majó-Vázquez, A.S. Cardenal, S. González-Bailón, Digital news consumption and copyright intervention: evidence from Spain before and after the 2015 “Link Tax”, *J. Computer-Mediated Commun.* 22 (2017) 284–301.
- [13] S. Majó-Vázquez, K.R. Nielsen, S. González-Bailón, The backbone structure of audience networks: a new approach to comparing online news consumption across countries, *Polit. Commun.* 36 (2018) 227–240.
- [14] S. Majó-Vázquez, A network analysis of online audience behavior: towards a better comprehension of the agenda setting process, *IDP: Revista d'Internet, Dret i Política* (2015).
- [15] A.L. Schmidt, F. Zollo, M. Del Vicario, A. Bessi, A. Scala, G. Caldarelli, H.E. Stanley, W. Quattrociocchi, Anatomy of news consumption on Facebook, *Proc. Natl. Acad. Sci.* 114 (2017) 3035–3039.

- [16] J. An, M. Cha, K. Gummadi, J. Crowcroft, Media Landscape in Twitter: a world of new conventions and political diversity, in: Proceedings of the 8 International AAAI Conference on Weblogs and Social Media, ICWSM, Barcelona, 2011.
- [17] F.L. Gaol, T. Matsuo, A. Maulana, Network model for online news media landscape in Twitter, *Information* 10 (9) (2019) 277.
- [18] Reporters Without Borders, Press freedom Ranking 2018, 2018. https://rsf.org/en/ranking_table.
- [19] M.D. Vicario, S. Gaito, W. Quattrociocchi, M. Zignani, F. Zollo, News consumption during the Italian Referendum: a cross-platform analysis on Facebook and Twitter, in: IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2017.
- [20] F. Marozzo, A. Bessi, Analyzing polarization of social media users and news sites during political campaigns, *Soc. Netw. Anal. Min.* 8 (2017) 1.
- [21] W. Quattrociocchi, A. Scala, C. Sunstein, Echo Chambers on Facebook, SSRN, 2016. Scholarly Paper No. ID 2795110.
- [22] <https://www.alexa.com>.
- [23] E. Summers, Twarc, Available online: <https://github.com/DocNow/twarc>.
- [24] Onur. Varol, E. Ferrara, C.A. Davis, F. Menczer, A. Flammini, Online human-bot interactions: detection, estimation, and characterization. *Dalam Proc. International AAAI Conf. On web and social media (ICWSM)*, Diambil Februari 9 (2017), 2019, dari, <https://arxiv.org/abs/1703.03107>.
- [25] J.S. Morgan, C. Lampe, M.Z. Shafiq, Is news sharing on Twitter ideologically biased?, in: Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW '13, 2013.
- [26] M.A. Serrano, M. Boguñá, A. Vespignani, Extracting the multiscale backbone of complex weighted networks, *Proc. Natl. Acad. Sci.* 106 (2009) 6483–6488.
- [27] Newman, M. E. J. *Networks: an Introduction*. 2010 Oxford, UK: Oxford University Press.
- [28] L.C. Freeman, Centrality in social networks: conceptual clarification, *Soc. Netw.* 1 (1978).
- [29] V.D. Blondel, J.L. Guillaume, R. Lambiotte, Multilevel local optimization of modularity, in: *Graph Partitioning*, John Wiley and Sons, 2013, pp. 315–345.
- [30] S. Maslov, Specificity and stability in topology of protein networks, *Science* 296 (5569) (2002) 910–913.
- [31] G.W. Flake, S. Lawrence, C.L. Giles, F.M. Coetzee, Self-organization and identification of Web communities, *Computer* 35 (3) (2004) 66–70.
- [32] E. Ravasz, Hierarchical organization of modularity in metabolic networks, *Science* 297 (5586) (2002) 1551–1555.