Proceedings of the 9th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'23)
Brunel University, London, United Kingdom - August 03-05, 2023
Paper No. CIST 131
DOI: 10.11159/cist23.131

Can We Avoid Filter Bubbles or Only Burst Them? A Natural Experiment Investigating Filter Bubbles in Non-Personalised Content Feeds

Krishnamoorthy[1], Möhlmann[2], Henfridsson[3], Yaraghi[3]
[1]Warwick Business School, University of Warwick
Scarman Rd, Coventry CV4 7AL, United Kingdom
c.krishnamoorthy@warwick.ac.uk
[2]Bentley University
175 Forest St, Waltham, MA, U.S.A
mmoehlmann@bentley.edu
[3]Miami Herbert Business School, University of Miami
5250 University Dr, Coral Gables, FL, U.S.A
ohenfridsson@miami.edu; niamyaraghi@miami.edu

Abstract **-** While the amount of information available has exponentially increased, our cognitive abilities to process information have not improved. Hence, social media platforms employ algorithmic filtering to keep information load at manageable levels. Algorithmic filtering leads to algorithmic distortion, creating phenomena such as filter bubbles and echo chambers, resulting in information blindness. This paper investigates whether filter bubbles exist even before personalisation due to users' actions to combat information overload. Through a natural experiment and a quantitative study on clickstream data from an information curation app, we investigate the presence of filter bubbles in a non-personalised content feed. We have built five multiple regression models using a within-person approach to answer our hypotheses. We find that the position of content shown has a statistically significant effect on consumption, thereby creating a filter bubble effect. We develop further nuances of such filter bubbles in polarised topics and by end-user interests.

Keywords**:** Filter Bubbles, Societal Impact of IS, Information Diversity, Polarising Topics, Online Platforms

1. Introduction

According to World Economic Forum, as of 2020, there are more than 40 zettabytes of data in the digital universe [1]. While the amount of information available has exponentially increased, our cognitive abilities to process information have not improved [2, 3]. Social media platforms such as Facebook and Twitter employ algorithmic filtering in order to keep information load at manageable levels. They expose users to personalised information that reflects their personal preferences [4-7].

Yet, such practices create substantial challenges, and there is increasing public concern about their societal effects [2, 8]. Algorithmic filtering leads to algorithmic distortion, creating phenomena such as filter bubbles [8, 9]. Critics argue that social media content can alter users' behaviour and even influence political debates and elections [2, 10, 11]. Filter bubbles form when personalisation algorithms filter content based on an individual user's profile, resulting in limited exposure to attitude-challenging points of view and, thus, information blindness as the full range of information is not visible to the users [5, 8, 9]. This persistent exposure to like-minded opinions on online platforms leads to polarisation and fragmentation [12]. Prevailing IS studies predominantly focus on large-scale platforms with billions of users that heavily use algorithmic personalisation [8, 11], such as search engines [5] and social media platforms [8, 11]. Therefore, the filter bubble effects on non-personalised online information consumption is an underdeveloped research direction. In this paper, by isolating the effects of personalisation, we investigate whether user behaviour to combat information overload in itself leads to filter bubbles. By doing so, we address an important research question: In non-personalised content feeds, how does the position of content shown impact its consumption, and does it create a filter bubble effect? Further, by analysing the user reading habits over various topics, we answer the secondary question: How does the filter bubble effect vary in polarised topics and by topics of interest?

To approach these research questions empirically, we conducted a natural experiment. We analysed the reading habits of around 180K app users who read about 3K infographics and generated more than 570K raw data points on a live information curation platform. We then built five multiple regression models on this data using a within-person approach to answer our hypotheses. We find that the position of content shown has a statistically significant effect on consumption, thereby creating a filter bubble effect. Further, we find mixed results for the impact of polarised topics and topics of interest on the filter bubbles. Our findings contribute to previous research in algorithmic filtering, personalisation and filter bubbles in social media platforms [2, 8, 10, 11]. While previous research has created an important foundation to understand better the challenges associated with this phenomenon, we investigate whether filter bubbles can be avoided so that efforts can be made to burst them instead. Furthermore, we explore the nuances around topics of interest and polarised topics so that we can focus on shifting opinions on polarising topics to help restore balance in online discussions. In the following, we introduce previous literature, share details about our empirical study, and introduce and share our findings.

2. Conceptual background

There is plenty of IS literature on how personalisation leads to filter bubbles, their characteristics and the consequences. From a user experience perspective in recommender systems, the common assumption is that personalisation implies using better algorithms to generate better recommendations that improve the overall user experience [13]. Personalisation is an essential component in designing the end-user experience, ranging from simply displaying names to presenting personalised choices based on detailed modelling of users' past behaviour [14]. This personalisation process using demographic details, historical interactions, and social contacts [2] creates an independent filter bubble around each user [15], constricting the complete flow of information. Such distortion occurs, for example, when platforms such as Google and Facebook remove contradictory views from users without their consent [16]. Hence, personalisation can show entirely different sets of facts to different users when algorithms decide what information is relevant to each user, resulting in what is known as a "unique informational universe" [5].

As platforms suggest friends to add based on mutual connections [17] to increase the overall user base [15], it increases homophily within a social group [18]. Moreover, the first interactions in such homophilic groups determine how the final network is structured [19]. They also find that the sequence of interactions in such groups is vital in determining the final outcome, even though the factors evolve over time. Hence, we argue that using data from an early-stage platform is beneficial rather than a fully formed platform with complex structures wherein access to initial datasets can prove cumbersome. Moreover, if such a platform has non-personalised content feeds, we can isolate the effect of personalisation and study whether filter bubbles can truly be avoided.

2.1. Theoretical Model Development

We conceptually build our research ground up by starting with the root cause of filter bubbles: information overload. The fundamental motivation behind the research is to understand whether information overload by itself can explain the theoretical constructs behind filter bubbles. We analyse from two different perspectives to juxtapose the formation of filter bubbles. On the one hand, platforms use personalisation algorithms to filter content based on collaborative and content filtering to show users a narrow list of content that they might find helpful. While looking from the user perspective, users automatically filter content by themselves due to limitations in their processing capabilities. Looking mainly from the perspective of content consumed, current literature has developed multiple constructs: separation, quantity, dispersion, variety, and parity [9, 20]. Looking at the quantity construct in detail, "Information Source Quantity" is one such measure of information diversity [8], wherein exposure to different information sources, such as websites with divergent viewpoints [7], is used to study filter bubbles. These measures are based on personalisation from the source perspective and do not reflect how filter bubbles stem from user behaviour alone and present a gap in the conceptual theories addressing this phenomenon. Agreeing with the current literature, we consider quantity as the integral theory to conceptually model filter bubbles and propose the number of times content is viewed when randomly placed in multiple positions in a non-personalised content feed as a measure of this quantity construct.

Given the biological limitations of human minds [3], users resort to multiple ways to filter the content they consume, such as restricting how many items they read from a ranked list to tackle information overload. This restriction, based on the on the content's position, leads to a significant drop in readership counts between the cut-off point N and beyond [15] when when using personalisation algorithms. We posit that this results in information blindness as users do not consume a wide range of information beyond the initial content shown in the list. Synthesising the two theories, we seek to investigate this drop in views in a non-personalised setting to study filter bubbles in this context. Next, we focus our research on how the filter bubble effect varies concerning polarised topics. Climate change is one of the most important issues we face in the 21st century. Climate change is also a topic of intense debate online, heavily influenced by homophilic networks and exacerbated by filter bubbles, echo chambers, bots and fake news [21]. Confirmation bias plays a key part in such polarised topics [22], wherein users seek out content reinforcing their beliefs. This motivates us to choose climate change for studying the impact of polarising topics on filter bubbles. Similarly, end-user interests play a significant role in content filtering in personalisation algorithms [23], and topic-based content filtering leads to great gains in user views [24]. Generalising these theories in our research context, we are motivated to investigate filter bubble effects by topics of interest. In summary, based on our theoretical model, we propose the following hypotheses:

H1: *Even in non-personalised content feeds, the position of the content shown has a negative effect on the consumption, whereby content placed higher up the list benefits from a higher number of views resulting in a filter bubble effect*

H2: *When content is non-personalised, polarising topics have a positive effect on filter bubbles as users are also shown content not confirming with their bias and thereby do not read content beyond what's initially shown*

H3: *When content is non-personalised, the filter bubble effect varies by the topic of content consumed as users seek content beyond what's initially shown based on their topics of interest*

3. Data and Methods

3.1 Research Setting

Given that our study is seeking to analyse pureplay user behaviour before any potential influence of personalisation by the platforms, it is only appropriate to choose a small-scale platform with a low user base where the content shown is not influenced by the use of algorithms to filter based on user profile and behaviour. To investigate user self-created filter bubbles, we obtained data from an information consumption platform, a smartphone application currently live on Android Play and iOS App stores. Henceforth, we refer to the application as app in this paper for brevity purposes. The app lets users explore content in a visual format (images) from third-party sources, acting as an information aggregator. The app classifies content shown into high-level topics (such as Business, Environment, and Health). In total, the app displayed 2,962 images across various topics to over 180K users in a newsfeed format similar to the Facebook homepage and Twitter feeds. The content images are shown in an ordered list for each topic, the page feeds are not personalised, and the images are randomly shown to each user. The app captures detailed audit data for interactions such as opening content for reading, bookmarking for future reference, sharing content on social media, and providing feedback such as liking and unliking. For ethical reasons, the app developers have chosen not to capture detailed personal information and also anonymous the audit data with a unique non-personally identifying alphanumeric code. The audit data also captures the position of each content in the list displayed from where users click and read. We attained unfettered access to this detailed anonymous user-level clickstream data. Climate change and the environment are deeply polarising topics online, and all the content on this topic presented in the app used for our experiment is related to showing that climate change exists. Other than for environment, the content shown for all other topics in the app is of neutral nature and does not incline towards one view or another.

3.2 Measures and Analysis

The crux of our research model is the concept of quantity concerning information consumption and filter bubbles and how the effects vary by the topic of content consumed. The quantity construct is related to user behaviour without any intervention by personalisation and forms the basis for measuring filter bubbles in a non-personalised environment. Please refer to Table 1 for the definition of the variables used in our Position Effect experiment to test hypotheses on the filter bubble effect (H1), polarised topics (H2) and topics of interest (H3).

Our first independent variable is the position of the content shown in a list of items. The other variables are the topic the content shown, whether the content belongs to a polarised topic, the type of content shown, the language of the the size of the image shown, the time taken to read the content and whether the position in the list when read is less than We include independent variables to control for short and long formats and effects on factual and opinion reporting [25] and top N cut-off effect [15]. We extended the current theoretical construct in filter bubbles and developed our dependent variable, "Number of Views", to measure quantity, which measures the number of times users read content in the non-personalised feed in the app. This variable is similar to the measure Distinct Number of Sites Viewed used in the research model by [8] to study filter bubbles. We intuitively argue that if the content is read more in early positions in a list, it leads to filter bubble effects. The independent and dependent variables are shown in Table 1.

Table 1 –Variables for the Position Effect Experiment

<table>
<tr><td colspan="4">Independent Variables</td></tr>
<tr><td>Variable</td><td colspan="3">Description</td></tr>
<tr><td>Position</td><td colspan="3">This variable identifies the position where the content was placed when the users clicked on the image to study the infographic in detail.</td></tr>
<tr><td>Topic</td><td colspan="3">The topic of the infographic being read such as Business, Environment and Health.</td></tr>
<tr><td>Polarising Topic</td><td colspan="3">We classified "Environment" as a polarising topic. This Yes/No variable indicates whether the content being read is a polarizing topic or not.</td></tr>
<tr><td>Top Topic</td><td colspan="3">The top three topics, Business, Health and Education, for which the app had the most number of images shown to the users and views. This Yes/No variable indicates whether the content being read is from one of these three topics.</td></tr>
<tr><td>Positions 1 to 10</td><td colspan="3">This Yes/No variable indicates whether the position in the list is from one to ten.</td></tr>
<tr><td>Language - English</td><td colspan="3">Based on the language of the content (Mostly English and some Spanish infographics). We coded this variable as a Yes/No flag based on whether the language is English.</td></tr>
<tr><td>Size</td><td colspan="3">Whether the content read is small, medium or large. We classified images less than 500 DPI (Dots Per Inch) as small as they take 1/3 of the smartphone screen, 500 to 1000 DPI as medium as they take up to 2/3 of the screen and more than 1000 DPI as large.</td></tr>
<tr><td>Content-Type</td><td colspan="3">Whether the content read is Factual, Opinion, Information, News or Advertisement</td></tr>
<tr><td>Time to read</td><td colspan="3">Average time to read the content in minutes based on how much text is in the image and the size of the image</td></tr>
<tr><td colspan="4">Dependent Variable</td></tr>
<tr><td>Dimension</td><td>Construct</td><td>Measure</td><td>Description</td></tr>
<tr><td>Filter Bubbles</td><td>Quantity</td><td>Number of Views</td><td>The number of times content is viewed</td></tr>
</table>

Though the dependent variable "Number of Views" does not follow a normal distribution, we use linear regression for our analysis given the large data set. Adjusted R Squared (Adj R2) is a widely used goodness of fit measure for standard linear regression models. Given that we want to study the impact of multiple independent variables on the dependent variable, including control variables, we have used multiple linear regression modelling for analysis.

4. Results

We developed five multiple regression models to investigate the factors that affect filter bubbles in non-personalised content feeds. Estimation results for the multiple regression models (coefficients, standard errors and significance levels) are are presented in Table 2. All five models are statistically significant at 0.1% levels, and the adjusted R2 values imply that that the regression models are meaningful.

Table 2 – Regression Results

Impact of Position (H1)		Polarising Topics Effect (H2)			Topics of Interest (H3)
Model	Model 1 (All Topics)	Model 2 (Environment Only)	Model 3 (Non-Environment Only)	Model 4 (Including Interaction term)	Model 5 (By Topics)
Position	-0.133*** (0.010)	-0.130*** (0.027)	-0.134*** (0.010)	-0.135*** (0.010)	-0.140*** (0.010)
Time To Read	-0.525 (0.566)	-1.033 (1.014)	-0.466 (0.615)	-0.534 (0.566)	-0.482 (0.566)
Top Topics	23.167*** (1.167)	Omitted	23.812*** (1.227)	23.846*** (1.194)	30.158*** (2.853)
Positions 1 to 10	70.470*** (1.530)	58.202*** (3.293)	71.636*** (1.632)	70.595*** (1.531)	70.981*** (1.529)
<u>Size</u>					
Large	6.331*** (1.237)	3.869*** (2.644)	6.582*** (1.320)	6.336*** (1.237)	6.018*** (1.237)
Small	-14.743*** (1.581)	-7.250*** (3.156)	-15.712*** (1.697)	-14.848*** (1.581)	-14.888*** (1.582)
<u>Content-Type</u>					
Factual	12.645*** (1.687)	16.823*** (3.855)	12.302*** (1.791)	12.589*** (1.687)	13.318*** (1.697)
Information	25.287*** (1.614)	23.252*** (3.661)	25.399*** (1.715)	25.167*** (1.614)	25.505*** (1.618)
News	-0.848 (1.872)	1.010 (4.061)	-1.181 (1.997)	-0.959 (1.872)	0.062 (1.881)
Promoted Content	1.499 (16.551)		0.863 (17.007)	1.298 (16.550)	0.022 (16.577)
English	11.689*** (2.422)	23.125*** (6.029)	11.147*** (2.561)	11.730*** (2.422)	13.496*** (2.443)
Polarising Topic for Model 4				3.574 (2.928)	
Interaction Term – Position x PolarisingTopic for Model 4				-0.040 (0.432)	
<u>Topics of Interest for Model 5</u> (Baselined against Business topic, a smaller subset of topics is shown here for brevity purposes)					
Science and Technology					11.556*** (3.089)
Education					1.353 (2.306)
Social					9.684*** (3.071)

Politics					-0.683 (3.223)
Intercept	-15.908*** (3.365)	-20.002*** (7.497)	-16.214*** (3.600)	-40.250*** (3.788)	-22.570*** (4.149)
Observations	28,068	2,145	25,923	28,068	28,068
Adjusted R2	0.112	0.195	0.110	0.112	0.115
Prob > F	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***

Robust standard errors are reported in parentheses. *, **, *** denote significance at the 5, 1 and 0.1% levels respectively.

To test H1, we used a multiple regression model to estimate the association between position and the number of views. The model controls for content size, language, top 10 positions, time to read and content type. For the measure of quantity construct, the Number of Views, we found that the position was associated with a decrease in consumption, and the model is statistically significant. The difference between medium and small or large content is statistically significant, as is the difference between content types, justifying the inclusion of the control variables. Looking at the regression results, we can conclude that there is a filter bubble effect of users constricting their own consumption leading to information blindness even in non-personalised content feeds. To test H2, we split the data into two groups, one including only the environment topic as a proxy for a polarised topic and another group including all other topics to build models two and three. In model four, we introduced an interaction term between position and polarising topic to validate further whether polarising topic has an effect on filter bubbles. We found mixed results for these models. While model two explains more of the variance than model three, and there is a marginal difference between the coefficients, but the difference is not statistically significant. Similarly, the interaction term for model four is not statistically significant. To test H3, we introduced topics of interest into the multiple regression and built model five. With the topic of business as the baseline, the difference between topics gives mixed results, but the overall model five is statistically significant.

In summary, considering the quantity construct for filter bubbles, we conclude that there is consistent support for a negative effect of position on content consumption. We theorise that this association, in turn, corresponds to the filter bubble effect in non-personalised settings. The adjusted R2 value for our model two, which uses only data from the environment topic, was much higher than other models, including all the topics (19% vs 11%). So, there is much less unexplained variation in the environment-only model which suggests a better goodness-of-fit. But, the difference in coefficients for the position variable is very marginal and not statistically significant. By adding an interaction term between Polarised Topics and Position, we test, when all things being equal, whether polarised topics have any effect on the filter bubble effect. While the beta coefficient suggests that polarised topics indeed have an effect, the effect is not statistically significant ($b = -0.040$, $p > .05$). In model five, while the beta coefficients for some topics vary against the baseline topic "Business", the coefficients are a mixture of significant and insignificant and both positive and negative. The summary of our hypotheses' results is illustrated in Table 3.

Table 3 – Summary of Relationships

Dimension	Construct	Hypothesis	Model	Relationship
Information Overload	Quantity	H1	1	Supported – Filtering Effect
Filter Bubbles	Polarising Topics	H2	2, 3 and 4	Partially Supported – Strengthening Effect
	Topics of Interest	H3	5	Partially Supported – Diverging Effect

5. Conclusion

5.1 Implications for Theorising Filter Bubbles

The impact of filter bubbles on how we consume information is an important area of study in the information systems systems domain, and we identify several contributions our study makes to the prior research addressing filter bubbles. First, First, beyond merely discussing filter bubbles in a personalisation theoretical framework [3, 5, 11], our study proposes a a more nuanced approach to investigating filter bubbles in a non-personalised context, thereby investigating implications of of pureplay information overload [2, 3] and extends the critical lens for a holistic theoretical underpinning. We argue that given the information overload, users create filter bubbles around themselves even before any personalisation of content by the platforms. Using an empirical study, we demonstrate that we cannot avoid this formation of filter bubbles which can only be burst. Second, our study extends the current literature on the quantity construct and introduces new measures to evaluate the filter bubble effects. By following our approach to measuring the filter bubble effect, platforms can identify the true scale of the distortion problem with their respective user base, albeit from a personalised algorithm-driven perspective. Third, the impact of filter bubbles on polarised topics is well understood [3, 25] as the filter bubbles enhance polarisation, but we flip the angle and provide a critical lens to study how polarised topics affect the filter bubble effects. Similarly, we distinguish how the filter bubble effect varies by topics of interest and add further constructs to studying filter bubbles. Finally, we make multiple contributions from the methodological and conceptual research angle to the current literature on filter bubbles. We extend the body of literature on filter bubbles from heavily qualitative-focused research to more quantitative experiment-focused research through the use of a natural experiment and multiple regression analysis and lend a balance.

Our findings have significant implications for a practical understanding of filter bubbles. Lawmakers worldwide are actively addressing the impact of technology platforms on the polarisation of opinions and the misinformation burden in large-scale platforms such as Google, Facebook and Twitter. The policymakers are seeking solutions on how to regulate the likes of Facebook and Twitter for greater transparency of algorithmic feeds. We also argue that as platform scale grows, there is the complexity of commercial pressures around proprietary algorithms. However, little is known about small-scale real-life platforms in this context, and few studies have used data from such sources where information blindness is still a concern. In fact, there are hundreds of thousands of such small-scale sources of information, such as blogs and web forums. Hence, our study addresses this research gap and expands current theories on filter bubbles to include a variety of platforms from this scale dimension. Our findings on filter bubbles are also crucial for moderators of online forums and creators of social media groups. As responsible netizens who wield great powers in information dissemination, they can take a leaf out of our study and investigate the formation of filter bubbles in the micro-ecosystem they develop and how to burst them.

5.2 Limitations and Future Research

The extended version of this research paper includes a detailed analysis of the data used for our research, including descriptive statistics on the infographics shown in the content feed and profile of the platform users. It also explains the theoretical constructs used to develop our research model and hypothesis in much detail. We acknowledge that our study has multiple limitations related to the research focus, method, and subjects. One limitation of the study is that our data came from one source displaying information as images, and also, the content was uploaded only by the platform developers. Most of the content uploaded by the developers was in the English language, while the app users themselves were spread across the globe, with a high proportion of Spanish language users. Future research could use data from multiple platforms, including numerous content formats such as video, news articles, information in various languages and also include user-generated content. Also, studies can include data from other polarising topics such as politics, vaccine denial and immigration. Similarly, our study centred on the other extreme of the personalisation angle using non-personalised feeds. So, future researchers can focus on the nuances of filter bubbles through quantitative models by varying the depth or level of personalisation between full personalisation and non-personalisation.

References

[1] World Economic Forum (2019, April 17). How much data is generated each day? [Online] Available: https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

[2] S. Aral and D. Walker, "Identifying influential and susceptible members of social networks," *Science,* vol. 337, p. 337–341, 2012.

[3] E. Bozdag, "Bias in algorithmic filtering and personalization," *Ethics and information technology,* vol. 15, p. 209–227, 2013.

[4] J. Chen and J. Stallaert, "An economic analysis of online advertising using behavioral targeting," *Mis Quarterly,* vol. 38, p. 429–A7, 2014.

[5] E. Pariser, The filter bubble: What the Internet is hiding from you, penguin UK, 2011.

[6] S. Y. X. Komiak and I. Benbasat, "The effects of personalization and familiarity on trust and adoption of recommendation agents," *MIS quarterly,* p. 941–960, 2006.

[7] J. Shore, J. Baek and C. Dellarocas, "Network structure and patterns of information diversity on Twitter," *Mis Quarterly,* vol. 42, p. 849–872, 2018.

[8] B. Kitchens, S. L. Johnson and P. Gray, "Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption.," *MIS quarterly,* vol. 44, 2020.

[9] M. Haim, A. Graefe and H.-B. Brosius, "Burst of the filter bubble? Effects of personalization on the diversity of Google News," *Digital journalism,* vol. 6, p. 330–343, 2018.

[10] M. Tarafdar and D. Kajal Ray, "Role of social media in social protest cycles: A sociomaterial examination," *Information Systems Research,* vol. 32, p. 1066–1090, 2021.

[11] S. Venkatesan, R. Valecha, N. Yaraghi, O. Oh and H. R. Rao, "Influence in Social Media: An Investigation of Tweets Spanning the 2011 Egyptian Revolution.," *MIS Quarterly,* vol. 45, 2021.

[12] J. K. Lee, J. Choi, C. Kim and Y. Kim, "Social media, network heterogeneity, and opinion polarization," *Journal of communication,* vol. 64, p. 702–722, 2014.

[13] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu and C. Newell, "Explaining the user experience of recommender systems," *User modeling and user-adapted interaction,* vol. 22, p. 441–504, 2012.

[14] J. Kramer, S. Noronha and J. Vergo, "A user-centered design approach to personalization," *Communications of the ACM,* vol. 43, p. 44–48, 2000.

[15] S. Prawesh and B. Padmanabhan, "The "most popular news" recommender: Count amplification and manipulation resistance," *Information Systems Research,* vol. 25, p. 569–589, 2014.

[16] P. M. Napoli, "What if more speech is no longer the solution: First Amendment theory meets fake news and the filter bubble," *Fed. Comm. LJ,* vol. 70, p. 55, 2018.

[17] L. He, L. T. Millar and J. C. Kahn, "Cluster-based friend suggestion aggregator," U.S. Patent 8 055 720, November 8, 2011.

[18] X. Fang and P. J. Hu, "Top persuader prediction for social networks," *MIS Quarterly,* vol. 42, p. 63–A14, 2018

[19] F. Bagnoli, G. de Bonfioli Cavalcabo', B. Casu and A. Guazzini, "Community Formation as a Byproduct of a Recommendation System: A Simulation Model for Bubble Formation in Social Media," *Future Internet,* vol. 13, p. 296, 2021.

[20] D. A. Harrison and K. J. Klein, "What's the difference? Diversity constructs as separation, variety, or disparity in organizations," *Academy of management review,* vol. 32, p. 1199–1228, 2007.

[21] K. M. d. Treen, H. T. P. Williams and S. J. O'Neill, "Online misinformation about climate change," *Wiley Interdisciplinary Reviews: Climate Change,* vol. 11, p. e665, 2020.

[22] S. Modgil, R. K. Singh, S. Gupta and D. Dennehy, "A confirmation bias view on social media induced polarisation during Covid-19," *Information Systems Frontiers,* p. 1–25, 2021.

[23] S. Papneja, K. Sharma and N. Khilwani, "Content Recommendation Based on Topic Modeling," in *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 1*, 2021.

[24] Z. Lu, Z. Dou, J. Lian, X. Xie and Q. Yang, "Content-based collaborative filtering for news topic recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2015.
[25] J. Möller, "Filter bubbles and digital echo chambers 1," in *The routledge companion to media disinformation and populism*, Routledge, 2021, p. 92–100