News Aggregation

Laurence Dierickx, University of Bergen, Université Libre de Bruxelles,
Laurence.Dierickx@uib.no
Carl-Gustav Lindén, University of Bergen, Carl-Gustav.Linden@uib.no

This paper explores the multifaceted landscape of news aggregation, highlighting its diverse practices and the resulting implications for information dissemination, narrative construction and public engagement. News aggregation encompasses manual, semi-automated and fully automated processes undertaken by traditional news media or digital platforms such as search engines or social media. While the credibility of the process is a corollary to assessing the quality of aggregated content, algorithmically driven processes pose several challenges in this regard, not only due to their lack of transparency but also due to their limited ability to detect and prevent the spread of disinformation, especially in the era of generative artificial intelligence. This paper also highlights the dual nature of news aggregation for democratic discourse, either by reinforcing existing media biases or by contributing to the polarisation of public debates. At the same time, they hold the promise of expanding accessibility and fostering greater audience engagement.

News Media; Platforms; Algorithms; Media Biases; Information Quality; Democracy

News aggregation involves collecting published news stories from various sources, redesigning them and republishing them in a concise format in a single place. The repurposed content can appear in different forms, such as specialised email newsletters, roundups of trending topics, and web or mobile applications (Molyneux & Coddington, 2020). Aggregation encompasses different practices involving a range of news products, from traditional news pieces to algorithmically distributed content on platforms.

This paper explores the multifaceted landscape of news aggregation, highlighting its impact on information dissemination, narrative construction and public engagement. The first part provides information on the logics and practices of news aggregation, whether manual or automated. In the second part, the paper unfolds the discussions on the role and responsibility of news aggregation, considering its potential to reproduce media biases or contribute to the polarisation of public debates. The final part of the paper focuses on the democratic challenges associated with news aggregation. While examining the new threats posed by the dynamics of the expanding disinformation landscape, it also underlines the need to promote more balanced and transparent approaches to assess the quality and credibility of the primary source.

From humans to automation: The mechanics and practices of news aggregation

The term 'aggregation' primarily refers to automating web searching for collecting content and returning relevant articles related to specific topics. On the other hand, 'news aggregators' refer to actors collecting content based on ranking criteria such as importance, popularity and newsworthiness. Strategies encompass manual, semi-automated and automated mechanisms. Manual aggregation involves assessing the importance of individual news stories, establishing connections and rewriting content into a cohesive narrative. Semi-automated news aggregation integrates human judgement within predefined procedures, while automated news aggregation rather refers to platforms such as Google News, Apple News or MSN that employ algorithmic processes to present relevant content to their users (Anderson, 2013; Lee & Chyi, 2015). Although there are similarities between aggregation and curation, these two concepts distinguish themselves in that curation requires human intervention. It involves adding value by assessing newsworthiness and incorporating additional context or analysis. This process includes experts refining and contextualising information, providing audiences with more than a simple compilation of data. Nevertheless, curation activities also rely on digital tools, with specific services like Scoop.it designed to assist human curators in their work. While human news curation and aggregation are commonly associated with political blogs and social media-

oriented websites, some news organisations maintain dedicated aggregation teams to republish news. At the same time, aggregators may also employ journalists to provide original content, such as The Huffington Post, which is based in the US and has several international editions.

For news media organisations, news aggregation can be viewed as a valuable and cost-effective tool that catalyses the rapid production of content on trending topics. Simultaneously, it raises the issue of prioritising clickbait to maximise online traffic — a logic that favours quantity over quality. Balancing the enhancement of previously published content and the reuse of content to generate more content, news aggregation involves navigating the benefits and risks associated with managing fast and diverse news flows (Molyneux & Coddington, 2020). It also prompts questions about the journalistic legitimacy and quality of news aggregation and the potential to blur the lines between news aggregation and original reporting. In this context, facts, quotes, documents and links are assembled to construct narrative-driven stories (Anderson, 2013).

News aggregation extends beyond the exclusive domain of news media as it contends with the platforms of major American tech companies. News aggregators, defined as websites that consolidate information from different sources into a unified platform, exhibit versatility across various approaches. Aggregators organise material from different websites into feeds, categorised by source, topic, or story. Specialised or niche aggregators concentrate on specific topics or geographical locations, including hyperlocal websites. User-curated aggregators present user-submitted links and snippets of text from various sources, often encompassing blog posts, multimedia content such as video and links to traditional news media. These approaches involve different collection strategies, including aggregation through news or blog feeds and user-curated aggregation, which pertains to the sharing of headlines and articles by the public (Wang & Keith, 2021).

The challenges of assessing information quality

Alongside social media and search engines, digital news aggregators have emerged as primary access points to news information, reshaping the dynamics of news distribution and discovery as they take on the gatekeeping role traditionally reserved for the news media (Wang & Keith, 2021). However, should these algorithmically driven news aggregators be seen as competitors in a landscape where news media companies are grappling with economic challenges? Although little research has focused on news aggregator consumption, two studies conducted in the US suggested that news aggregators do not steal readers from news publishers. Instead,

they drive more readers to online news media, which benefit from the distribution of their content on third-party platforms thanks to brand loyalty. Aggregators were mainly used by younger and more educated audiences who were not looking for opinion content, suggesting a complementary rather than substitutive logic compared to traditional news media (Chen & Pain, 2021).

Platforms can also be viewed as a response to the lack of specialised news media, as seen in the US, where hyperlocal news aggregators have played a role in compensating for the decline of local news (Kavanaugh et al., 2012). Additionally, news aggregators have been regularly accused of stealing content without permission or compensation. Although most national laws recognise exceptions for the reuse of protected content, some publishers have resisted and taken legal action, leading Google to enter into licensing agreements in several countries (Wang & Keith, 2021). Nonetheless, news aggregators tend to prioritise free content, neglecting value-added content such as investigative journalism that may be behind paywalls. This oversight reduces the diversity and depth of information available to users.

Another challenge in news aggregation is the delegation of news gathering and evaluation responsibility to opaque algorithmic processes which are, by nature, defined by humans for human uses. The determination of newsworthiness by human gatekeepers implies an intertwining of objective and subjective criteria, which are also grounded in socio-professional and organisational factors such as the editorial line of the news media. News aggregators' algorithms tend to perpetuate or amplify these pre-existing media biases, as they result from a nuanced interplay between journalistic norms, platform-specific cues and algorithmic relevance (Engelmann, Luebke & Kessler, 2021). However, the lack of transparency on how these algorithms operate remains a critical issue insofar as it fails to provide guarantees either on the source credibility or on the fairness of the selection process

In news aggregation platforms, the assessment of information quality is intricately connected to the automated processes governing their functionality. These processes are driven by filtering mechanisms that include content-based filtering, which considers the news piece and its relevance to the user's interest; collaborative filtering, based on the news consumption of users with similar preferences; knowledge-based filtering, relying on the data collected from the user. These approaches differ from human assessments of source credibility, where the recipient assesses the expertise and trustworthiness of the sender through a mix of objective and subjective criteria, thereby shaping the overall perception of credibility. Consequently, the perceived credibility of algorithmic news content aggregation remains closely linked to the perceived credibility of its sources, regardless of the degree of automation or aggregation (Diel

& Roberts, 2021). News aggregators lack a direct assessment of the information quality they provide. Instead, their algorithmic processes incorporate the evaluation of various factors, including the source's authority, the content's keywords, the user's geographical location and past searches, and clicks (Chyi, Lewis & Zheng, 2016). This personalised agency of information underscores that individual preferences and choices have never been more decisive in navigating media environments. However, users are not always aware (or informed) that what they see on their screens is directly related to explicit or implicit choices, underscoring the need for greater transparency in this area.

Algorithmic duality and the challenges of disinformation

In the realm of political information consumption, the study of personalised news aggregation has drawn attention to its ability to shape filter bubbles and echo chambers. These phenomena are likely to foster an environment where individuals are predominantly exposed to information that aligns with their existing beliefs, personal preferences or socio-cultural references. This selective exposure has the potential to fortify biases, restrict exposure to diverse viewpoints, and diminish overall informational diversity. Filter bubbles and echo chambers also refer to adversarial consequences, considering that they are likely to lead to an exacerbation of segregation and polarisation of public opinion due to the biases they convey (Bozdag & Van Den Hoven, 2015). Research has also found that personalised news recommender systems are likely to have a negative direct effect on knowledge gain (Beam, 2014).

Counterarguments suggest that the entire ecosystem of infomediaries, including search engines and social media, is associated with greater exposure to opposing perspectives (Flaxman, Goel & Rao, 2016). They also emphasise that algorithmically driven recommendations improve the user experience (Mitova et al., 2023) and have the potential to encourage political participation (Feezell, Wagner & Conroy, 2021). Further research found no evidence to support the filter bubble hypothesis in the case of Google News, although this system was found to over- or under-represent certain news outlets (Haim, Graefe & Brosius, 2018). While these counter-arguments highlight the positive aspects of algorithmic recommendations, it remains critical to assess their impact on information diversity and democratic discourse, considering the potential trade-offs and challenges they may pose in the era of artificial intelligence (AI). Also, new challenges have emerged recently, as advances in generative artificial intelligence have significantly increased the accessibility, speed and scalability of disseminating various forms of information disorders. In this context, detecting mis- and disinformation has become

increasingly stressful, given the potential for generating and spreading unintentional fake content stemming from artificial hallucinations, which can be described as generating content that does not correspond to ground-truth. Avoiding deception is also a pressing concern in news aggregation, as it is necessary to maintain an acceptable level of credibility and, therefore, trust. The need for robust and transparent systems to prevent the unintentional dissemination of harmful content is therefore crucial. This is all the more critical as the detection of manipulated content remains a multi-faceted and ongoing challenge, whether for humans or computer programs.

In addition, a new form of aggregation is emerging: the automated rewriting and reuse of previously published content. This new trend in AI comes from the UK, where news publisher Reach uses a system called Gutenbot to maximise traffic to its network of websites without duplicating content. The tool is also used to rewrite press releases, leading some news outlets to rely on AI-generated content rather than traditional reporting. Despite claims of editorial verification, there are concerns about the lack of links to the source material. Given the threats that AI-based systems pose to the accuracy, reliability and credibility of sources, assessing the quality of aggregated content has therefore become one of the potentially most critical issues to be explored in the near future, for both human curators and algorithmic aggregators.

Conclusion

The news process is inherently influenced by a complex conjugation of human factors, introducing subjectivity and objectivity at every stage — selecting topics, choosing sources, determining angles and deciding on storytelling approaches. Human curation of news adds a layer to all these possible human biases, as it does not eliminate selection bias. On the other hand, human curation is seen as offering a nuanced approach, combining expertise and contextual understanding that a computer program cannot have. Algorithmic news aggregation reflects the same challenges of selection bias and source credibility as they relate to previously made human choices. However, algorithmically driven aggregation processes highlight the need to assess information quality and promote transparency and accountability.

The dual nature of news aggregation presents both a threat and an opportunity to the democratic function of news media consumption. On the one hand, the process risks reinforcing bias and contributing to polarisation by selectively presenting content and limiting exposure to diverse perspectives. Conversely, news aggregators can be seen as agents of increased accessibility and engagement, potentially broadening audience participation in democratic discourse.

Amidst these dualities, which remain to be empirically explored, the emergence of generative AI has brought new challenges. For researchers, these include assessing and developing methods to detect and prevent the unintentional spread of misinformation and disinformation, including through both human and algorithmic aggregation.

References

Anderson, C. W. (2013). What aggregators do: Towards a networked concept of journalistic expertise in the digital age. *Journalism*, 14(8), 1008-1023.

Beam, M. A. (2014). Automating the news: How personalized news recommender system design choices impact news reception. *Communication Research*, *41*(8), 1019-1041.

Bozdag, E., & Van Den Hoven, J. (2015). Breaking the filter bubble: Democracy and design. *Ethics and Information Technology*, *17*, 249–265.

Chen, V. Y., & Pain, P. (2021). News on Facebook: How Facebook and newspapers build mutual brand loyalty through audience engagement. Journalism & Mass Communication Quarterly, 98(2), 366-386.

Chyi, H. I., Lewis, S. C., & Zheng, N. (2016). Parasite or partner? Coverage of Google News in an era of news aggregation. *Journalism & Mass Communication Quarterly*, 93(4), 789-815.

Diel, S., & Roberts, C. (2021). News story aggregation and perceived credibility. *Newspaper Research Journal*, *42*(2), 162-181.

Engelmann, I., Luebke, S. M., & Kessler, S. H. (2021). Effects of news factors on users' news attention and selective exposure on a news aggregator website. *Journalism Studies*, 22(6), 780-798.

Feezell, J. T., Wagner, J. K., & Conroy, M. (2021). Exploring the effects of algorithm-driven news sources on political behavior and polarization. *Computers in Human Behavior*, *116*, 106626.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298-320.

Haim, M., Graefe, A., & Brosius, H. B. (2018). Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital Journalism*, 6(3), 330-343.

Kavanaugh, A., Gad, S., Neidig, S., Pérez-Quiñones, M. A., Tedesco, J., Ahuja, A., & Ramakrishnan, N. (2012). (Hyper) local news aggregation: Designing for social affordances. In *Proceedings of the 13th Annual International Conference on Digital Government Research* (pp. 30-39).

Lee, A. M., & Chyi, H. I. (2015). The rise of online news aggregators: Consumption and competition. *International Journal on Media Management*, 17(1), 3-24.

Mitova, E., Blassnig, S., Strikovic, E., Urman, A., Hannak, A., de Vreese, C. H., & Esser, F. (2023). News recommender systems: A programmatic research review. *Annals of the International Communication Association*, 47(1), 84-113.

Molyneux, L., & Coddington, M. (2020). Aggregation, clickbait and their effect on perceptions of journalistic credibility and quality. *Journalism Practice*, 14(4), 429-446.

Wang, Q., & Keith, S. (2021). News aggregators and copyright in the European Union and the United States in the digital age: Evolution, comparisons, and implications. *First Monday*.