

## Beyond the algorithm: Measuring diversity in media recommender systems\*

Heritiana Ranaivoson, Aina Errando & Adelaida Afilipoaie

\*This post is based on the imec.icon TRENDIFY research project, funded by imec and Agentschap innoveren & ondernemen.

For a functioning democracy, having an informed citizenry is crucial, along with providing a public platform for expressing and discussing diverse ideas and opinions. Diversity is vital not only in media content but also in how media interact with digital technologies, including automated communication processes (Bernstein et al., 2021). Media diversity is an important objective, which is formative for democratic societies (Helberger, 2011). Algorithms' decisive editorial role (Beckett, 2019) influences content consumption on the receiving end of the communication process. Yet, it is important to understand that media diversity is multi-dimensional and goes beyond solely content diversity considerations. Moreover, while we could easily place the blame on the lack of audiences' exposure to diverse content on algorithms, we need to first look at the various dimensions that define media diversity, to understand how diversity can be assessed.

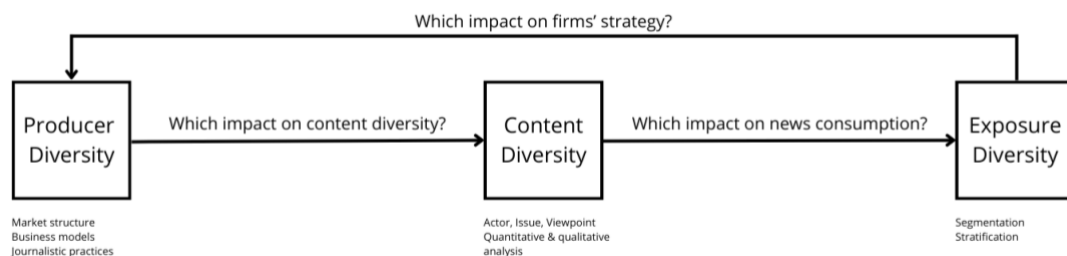
In the current scenario of polarization and fragmentation, we need to revisit the fundamental functions of diversity in society, and the impact of recommender systems therein (Bernstein et al., 2021). As Hendrickx et al. (2021) argue, the difficult tasks of characterizing and operationalizing news diversity within media markets and establishing a direct connection between it and democracy have been extensively discussed in scholarship but have never been fully achieved. Based on previous work (Ranaivoson, 2007, 2020), in the following lines, we briefly discuss a pragmatic definition of (media) diversity. The core of diversity makes it possible to **compare approaches across various disciplines** such as communication studies, social sciences, and computer science.

### Defining diversity

**The now well-established Stirling model** defines **diversity as a mix of variety, balance, and disparity** (Moreau & Peltier, 2004; Ranaivoson, 2007; Stirling, 2007). The model's main characteristic is to rely on a definition of diversity that goes beyond the sheer number (for example, of films, titles, etc.) to integrate for instance how each genre or origin is represented compared with the others, and how genres or artists that are very different from the others are present or not (Ranaivoson, 2020). More precisely, **the diversity of a system can only be assessed when its elements have been grouped into categories** (Ranaivoson, 2007). When this categorization has been achieved, *Variety* corresponds to the number of categories, *Balance* is the way the elements are spread among categories (for example, newspapers' market shares) and *Disparity* corresponds

to the level of difference between the categories (for example, how different the books are one from another or if they belong to the same genre).

Because defining diversity in a media context is a complex matter, it could be argued that Stirling’s model is too generic. For this reason, two additional, **more specific dimensions** can be added alongside Stirling’s model. The second dimension opposes **supplied diversity to consumed diversity** (Eaton & Lipsey, 1989; Van Cuilenburg & Van der Wurff, 2001). Supplied diversity corresponds to the diversity of what is made available. Consumed diversity refers to diversity as it is consumed depending on both consumer tastes and supplied diversity. A third dimension relies on Napoli (1999) who characterized diversity as having three aspects that correspond to different parts of the communication process: **source diversity, content diversity and exposure diversity**.



**Figure 1: Napoli’s Model and connections between concepts**

How can we assess media diversity? The triple M: Measures, Metrics, Methods

A limited number of studies (Australian Communications and Media Authority, 2020) combine the diversity of algorithmic suggestions and the diversity of consumption. Advancing our understanding of exposure diversity requires making conceptual distinctions, which can be further operationalised in research and policymaking. We propose to distinguish between measures, metrics, and methods.

### Measures

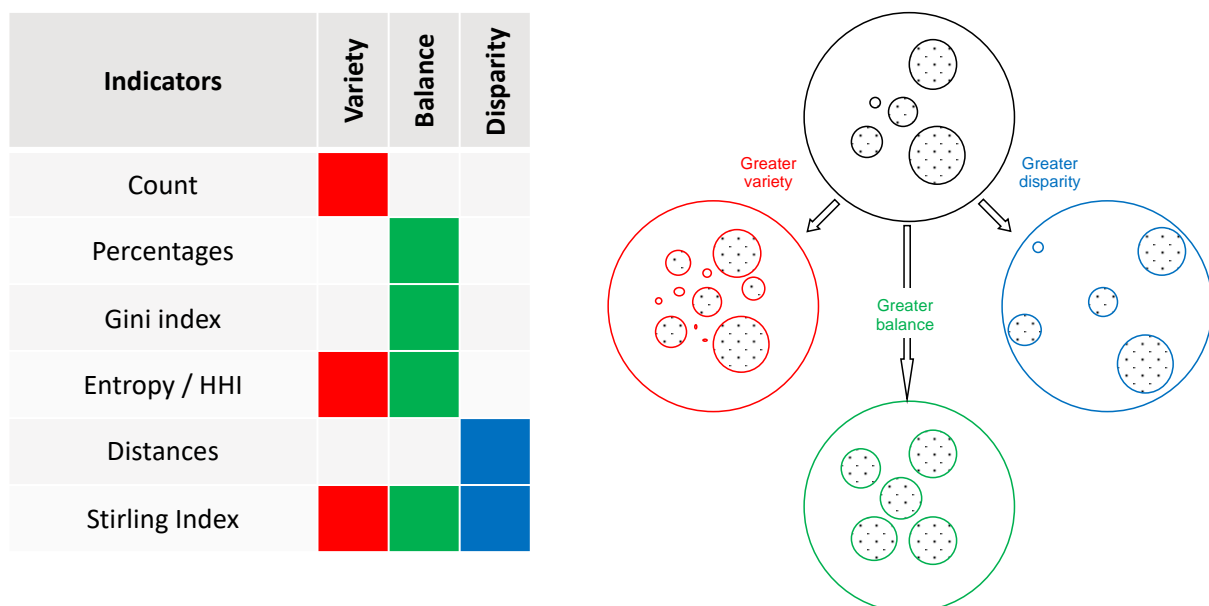
**Measures are interventions to achieve more diversity.** Measures may correspond to software design solutions or policies. Effective approaches to enhancing the diversity of exposure may be tools for alternative recommendation settings and technologies that make users aware of their filter bubble (Helberger, 2018).

### Metrics

**Metrics** or indicators **correspond to what is being measured**, be it diversity itself, or the barriers to diversity. Based on previous definitions, we identified four basic indicators to measure media diversity: Count, Percentages, Dual Indices and Distances (Ranaivoson & Domazetovikj, 2023).

- **Count**: corresponds to the number of categories or elements present, for example, the number of films available on Netflix).
- **Percentages** correspond to all indicators consisting in measuring the relative share of a category within a system (for example the Concentration Ratio used in competition policy).
- **Dual Indices** (Stirling, 2007), include notably the Entropy Index, based on Shannon (2001)'s work, and the Herfindahl-Hirschman Index (HHI), used to assess market concentration.
- Computer scientists assess diversity by using **Distances** (Kunaver & Požrl, 2017). Considering our definition, increasing the diversity of recommendations will result in suggesting items that are further from a user's preferences. For example, if somebody always reads about football, they could be proposed an article about another sport, because football is a subcategory of sport.
- The **Stirling Index** is an index that allows us to measure Variety, Balance, and Disparity at the same time. It has been rarely applied to diversity in the creative industries (see Farchy & Ranaivoson, 2011 for an example).

All these indicators can be mapped over the definition of the Stirling Model (Variety, Balance, Disparity). The following figure illustrates on the right side the impact of increasing the respective components of our first dimension of media diversity (the Stirling model). The left side visualises how it overlays with the metrics here.



**Figure 2: Mapping of the main indicators over the diversity components (source: author)**

## Methods

**Methods are ways to collect data with which to assess diversity.** Methods include experiments, surveys, and computational approaches. Some approaches focus on

analysing survey data. Assessment of diversity in the digital environment, as Helberger (2018) points out, must include new criteria, such as the extent to which users have options and the autonomy to choose between sources.

What should be the next steps?

**A major difficulty lies in how to best involve – or address – online platforms.** Online platforms have a crucial impact on all media sectors through their algorithmic gatekeepers (Helberger, 2019; Napoli, 2015), which they deploy to automatically filter, rank, and recommend content (Haim et al., 2018). Understanding their impact requires being able to measure the media diversity they make available and recommend. However, in doing so, many questions arise. This is why it is necessary to first be able to accurately measure existing media content diversity through the proposed triple M framework (measures, metrics, methods) and then compare it to the output generated by recommender systems. This is to be followed by research into audience consumption which would reveal the effects of algorithms and recommender systems on audiences' exposure diversity. Lastly, the findings can showcase how algorithmic gatekeepers are contributing to or hindering media diversity and where interventions are further needed.

A follow-up and more extensive overview of frameworks used to not only measure media diversity but, beyond, promote prominence, discoverability, and serendipity, will be key to ensure measurement frameworks of media diversity remain relevant. **That is why we propose the novel notion of epistemic welfare.** Epistemic welfare refers to the conditions and capabilities necessary for individuals and groups to proactively, deliberately, and autonomously engage with, contribute to, and influence the processes of knowledge production, modification, and dissemination. Citizens need diverse media (more media, reflecting different perspectives in a balanced manner) but such diversity may not be enough if citizens do not know which media to trust, or if the systems that recommend them content are not transparent.

Not only is diversity and its trajectory dependent on new technologies and recommender systems in the future, but it is also influenced by shifting social norms, usage patterns, laws, and commercial strategies. The media's constitutive role has always been to provide society with a range of accurate and relevant information for democratic debate. These days, platforms are partially fulfilling this function. Therefore, platforms and journalism must now join forces in this shared responsibility (Bernstein et al., 2021).

---

## References

- Australian Communications and Media Authority. (2020). News in Australia: Diversity and localism.
- Beckett, C. (2019). New powers, new responsibilities. A global survey of journalism and artificial intelligence. <https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/>
- Bernstein, A., Vreese, C., Helberger, N., Schulz, W., Zweig, K., Baden, C., Beam, M., Hauer, M., Heitz, L., Jürgens, P., Katzenbach, C., Kille, B., Klimkiewicz, B., Loosen, W., Moeller, J., Radanovic, G., Shani, G., Tintarev, N., Tolmeijer, S., & Zueger, T. (2021). *Diversity in News Recommendation*. 9. <https://doi.org/10.4230/DagMan.9.1.43>
- Eaton, B. C., & Lipsey, R. G. (1989). Product differentiation. In *Handbook of Industrial Organization* (p. 952-1009). Elsevier Science Publishers B. V.
- Farchy, J., & Ranaivoson, H. (2011). An international comparison of the ability of television channels to provide diverse programme: Testing the Stirling model in France, Turkey and the United Kingdom. *UIS Technical Paper*, 6, 77-138.
- 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), 330343. -<https://doi.org/10.1080/21670811.2017.1338145>
- Helberger, N. (2011). Diversity by Design. *Journal of Information Policy*, 1, 441-469.
- Helberger, N. (2018). Challenging Diversity-Social Media Platforms and a New Conception of Media Diversity. Digital Dominance: The Power of Google, Amazon, Facebook, and Apple, 153-175.
- Helberger, N. (2019). On the Democratic Role of News Recommenders. *Digital Journalism*, 7(8), 993-1012. <https://doi.org/10.1080/21670811.2019.1623700>
- Hendrickx, J., Smets, A., & Ballon, P. (2021). News Recommender Systems and News Diversity, Two of a Kind? A Case Study from a Small Media Market. *Journalism and Media*, 2(3), 515-528. <https://doi.org/10.3390/journalmedia2030031>
- Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems – A survey. *Knowledge-Based Systems*, 123, 154-162. <https://doi.org/10.1016/j.knosys.2017.02.009>
- Moreau, F., & Peltier, S. (2004). Cultural Diversity in the Movie Industry: A Cross-National Study. *Journal of Media Economics*, 17(2), 123-143.
- Napoli, P. M. (1999). Deconstructing the Diversity Principle. *Journal of Communication*, Autumn, 7-34.
- Napoli, P. M. (2015). Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers. *SPECIAL ISSUE ON THE GOVERNANCE OF SOCIAL MEDIA*, 39(9), 751-760. <https://doi.org/10.1016/j.telpol.2014.12.003>
- Ranaivoson, H. (2007). *Measuring cultural diversity: A review of existing definitions*. UNESCO.
- Ranaivoson, H. (2012). Does the consumer value diversity? How the economists' standard hypothesis is being challenged. In *The Consumption of Culture, The Culture of Consumption. A Collection of Contributions on Cultural Consumption and Cultural Markets* (p. 70-95). Lambert Academic Publishing.
- Ranaivoson, H. (2019). Online platforms and cultural diversity in the audiovisual sectors: A combined look at concentration and algorithms. In *Audiovisual Industries and Diversity* (p. 100-118). Routledge.

Ranaivoson, H. (2020). Cultural diversity. In R. Towse & T. Navarrete Hernández, *Handbook of Cultural Economics, Third Edition* (p. 183-191). Edward Elgar Publishing. <https://doi.org/10.4337/9781788975803.00026>

Ranaivoson, H. R., & Domazetovikj, N. (2023). Platforms and Exposure Diversity: Towards a Framework to Assess Policies to Promote Exposure Diversity. *Media and Communication*, 11(2), 379-391. <https://doi.org/10.17645/mac.v11i2.6401>

Shannon, C. E. (2001). A Mathematical Theory of Communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(1), 3-55. <https://doi.org/10.1145/584091.584093>

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15), 707-719.

Van Cuilenburg, J., & Van der Wurff, R. (2001). Impact of moderate and ruinous competition on diversity: The Dutch television market. *The Journal of Media Economics*, 14, 213-229.