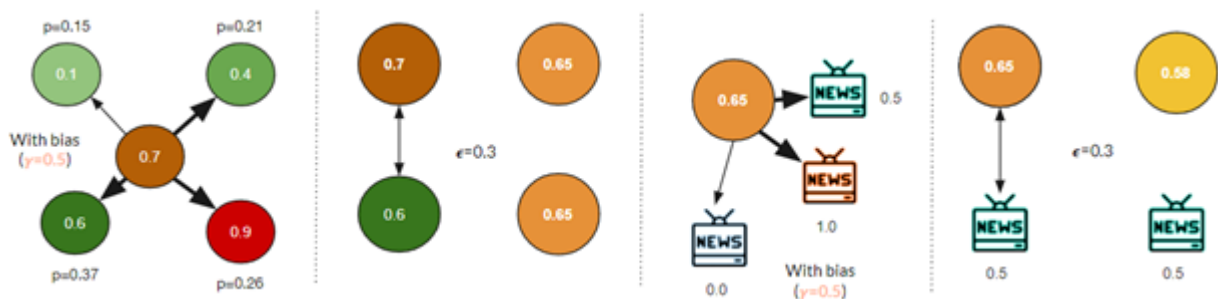


# Interplay between algorithmic bias and external information effects in opinion dynamics with bounded confidence

**Alina Sirbu**<sup>1</sup>, Valentina Pansanella<sup>2</sup>, Janos Kertesz<sup>4</sup>, Giulio Rossetti<sup>3</sup>

<sup>1</sup>University Of Pisa, Pisa, Italy, <sup>2</sup>Scuola Normale Superiore, Pisa, Italy, <sup>3</sup>National Research Council (CNR), Pisa, Italy, <sup>4</sup>Central European University, Vienna, Austria

Opinion formation in society is a complex process with multiple contributing mechanisms. At the base there are individual traits, peer interactions and interaction with mass media. In recent years, peer and mass media interactions have changed in nature, as we increasingly consume information through social networks. Due to their business oriented nature, social network platforms embed recommender systems to enhance user experience and maximise platform usage. These propose to users information that they are more likely to enjoy, which also means that they are more likely to agree with or approve the messages shown, a mechanism known as “algorithmic bias”. This may result in the formation of information bubbles, and appearance of fragmentation and polarisation of the public discourse. From the point of view of opinion dynamics, algorithmic bias in social networks have an important effect on two mechanisms: peer interaction and mass media consumption. In previous work [1] we have introduced a bounded confidence model in which algorithmic bias acts on peer interaction, by modifying the probability of selecting a peer to discuss with, changing from an uniform distribution to a distribution that depends on the distance between the agent’s opinions. We showed that this mechanism can increase fragmentation and polarisation, and slow down the dynamics. This applies both to complete social networks and to complex network topologies, where dynamics are further slowed down and fragmentation enhanced [2]. In this work we extend the previous model by introducing mass media (external) effects. The agents are endowed with a probability to interact with mass media, and they select among the available media using algorithmic bias, which makes it more probable to consume media that is closest to one’s opinion. We study several settings, ranging from one polarised or moderate external media to two and three media settings. Our results show that aggressive media campaigns (extreme message or higher probability to interact with a moderate message) tend to cause polarisation, agreeing with previous literature. The effect is enhanced in general by algorithmic bias. There are however settings where algorithmic bias appears to have a protective effect against polarisation and fragmentation: when the media landscape contains both polarised and moderate media outlets and the population is open minded (confidence bound is large). In this case, higher algorithmic bias impedes agents to interact with extremist media, facilitating consensus on the moderate media, albeit with much slower dynamics.



## References

- [1] A. Sirbu, D. Pedreschi, F. Giannotti, J. Kertész, Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. *PloS One* 14(3), e0213246 (2019).
- [2] V. Pansanella, G. Rossetti, L. Milli, From mean-field to complex topologies: network effects on the algorithmic bias model." In *Complex Networks & Their Applications X: Volume 2, Proceedings of the Tenth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2021* 10, pp. 329-340. Springer International Publishing, 2022.