

Manipulation Resistant Social Networks

Paolo Turrini

MathSys Project Proposal

Idea and Motivation

We use social networks all the time: for news, food, events and even charity. But how much can we trust them? Was the information provided to us truthful and accurate, or was it biased and manipulated?

The most advanced systems available now for decision support and recommendation are learning-based systems, which estimate the similarity between users to improve the accuracy of the suggestions. However, all these algorithms treat users' input as a given and are not equipped with the theoretical guarantees needed to ensure its truthfulness. This means that biased participants can potentially disrupt them providing false information – think for example of how easy it is to damage a business by posting a fake review on Amazon or TripAdvisor.

In an era where the reliability of the information available to us is increasingly put under scrutiny and fake news make the headlines, there is an urgent need for provably reliable platforms that foster trust by design.

The goal of this project is to construct social networks which discourage the deliberate injection of biased information, measuring such systems against formally specified requirements. The project studies manipulation as a strategic decision, looking at how to put together individuals' opinions so as to reduce the expected gains for malicious attackers, and that can happen in various forms: bribes, fake accounts, direct influence and the like. The key insight is the adoption of 'noisy combinations' of aggregation methods, identifying the least favourable from the attackers' point of view.

Key Research Areas

Artificial Intelligence, Algorithmic Game Theory, Computational Social Choice

Key References

- [Brandt 2017] Felix Brandt; *Rolling the Dice: Recent Results in Probabilistic Social Choice*, in Trends in Computational Social Choice (Ulle Endriss, editor); 2017.
- [Chistikov et al., 2020] Dmitry Chistikov, Grzegorz Lisowski, Mike Paterson and Paolo Turrini; *Convergence of Opinion Diffusion is PSPACE-complete*, Proceedings of the AAAI Conference on Artificial Intelligence (AAAI); 2020.
- [Grandi and Turrini, 2016] Umberto Grandi and Paolo Turrini; *A Network-based Rating Systems and its Resistance to Bribery*, Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI); 2016
- [Grandi et al., 2018] Umberto Grandi, James Stewart and Paolo Turrini; *The Complexity of Bribery in Network-based Rating Systems*, Proceedings of the AAAI Conference on Artificial Intelligence (AAAI); 2018.
- [Jackson, 2008] Matthew O. Jackson; *Social and Economic Networks*; 2008.
- [HL Paper 100] House of Lords Select Committee on Artificial Intelligence; *AI in the UK: ready, willing and able* (HL Paper 100), Parliament Publications; 2019.