PolyGraphs: Combating Networks of Ignorance in the Misinformation Age

Brian Ball, Alexandros Koliousis and Mike Peacey

1 Aims & Objectives

Misinformation is spreading through society, on social media networks and elsewhere, about a range of important topics. But in democratic societies, appropriate action requires large groups to form reasonable opinions, or *attitudes*, on what are often complex issues. We could attribute mistaken beliefs to cognitive biases (Kahneman, 2011); or, instead, to structural features of social networks (O'Connor & Weatherall, 2019). Our project explores the latter hypothesis. Thus, unlike prior work in the computational social sciences, we focus on socially networked individuals modelled as rational agents with evidence-based opinions. Moreover, where the spread of false belief within a population has been investigated before, its consequences for the knowledge (or ignorance) of the group as a whole have not; and the individuals modelled have been insufficiently evidentially sophisticated. By contrast, we aim to understand:

1. How group attitudes (descriptively) depend upon and (normatively) should be aggregated from individual ones.
2. Ways in which sensitivity to 'higher-order' evidence, such as peer disagreement and testimonial unreliability, can and should impact individual and group attitudes.

We want to understand how groups can remain ignorant – whether through error (i.e. false belief) or omission (group agnosticism) – and for how long, even when group members act rationally and assess evidence that itself points to the truth. The challenge is to construct a mathematical and algorithmic framework suitable for simulating the social phenomena in question.

2 Brief Plan

Our project comprises two research components. First, we will investigate the nature of groups and the relations between individual and group attitudes. Using the methods of plural logic and network science, we explore the hypothesis that whether a group believes or knows something depends not only on purely quantitative facts as how many of its members believe or know it, but also on how its members relate to one another, making the graphical representation of groups crucial. Second, we tackle a problem that has recently attracted considerable attention in philosophical circles (Skipper & Steglich-Petersen, 2019): how can an individual best use 'higher-order' evidence bearing on the question of whether it is rational for her to believe something?

Our project spans across three subject areas: philosophy, economics and computing – hence it has three co-investigators. We all work collaboratively on each component. Our methodology can be viewed as a sequence of translating philosophical questions on group attitudes into mathematical models of information sharing in social networks into graph computational algorithms. We also expect our philosophical investigations to yield new insights into information sharing models in economics and graph workloads in computing. We give a detailed operationalisation of our project in §4.

3 Methods & Techniques

Our primary experimental method is philosophical simulations – the use of computer simulations of large-scale, realistic social networks for philosophical investigations. A small group of pioneering philosophers already argue that simulations should be "a tool in every philosopher's toolbox" (Mayo-Wilson & Zollman, 2020). But philosophical simulations potentially pose a twin danger: too simple can be unrealistic; too complex, intractable. Thus, we build our simulation framework upon three well proven models, one per subject area; and extend them in §§3.1-3.2.

From philosophy, we draw on Bayesian epistemology, modelling individual rational beliefs, or *credences*, as coming in degrees that respect the principles of probability theory and evolve by conditionalization on new evidence. From economics, we build upon Bala & Goyal (1998) and model social dynamics using connected directed graphs. Vertices represent individuals (or agents), each with a credence (a real number between 0 and 1) in some target proposition. Edges represent information sharing channels from one agent to another. Agents periodically acquire new (first-order) evidence and share it with their direct neighbours; and they update their credences based on not just their own evidence, but also evidence received from neighbours. Analytic solutions of this model are computationally intractable for complex networks (Park et al., 2014), reinforcing the need for philosophical simulations.

Finally, from computing, we build upon the Bulk Synchronous Parallel (BSP) model of graph computations (Malewicz et al., 2010). Bala and Goyal's local learning algorithm suits a BSP-style, iterative program. At each iteration, the program (i) reads messages sent to each agent at the previous iteration; (ii) updates the credence of each agent, conceptually in parallel; and (iii) sends messages to its outgoing edges that will be received at the next iteration. We favour BSP for our computational analysis because synchronisation barriers in between stages (i)-(iii) enable us to check the correctness of our proposed algorithms while being scalable.

Put together, our simulation framework will produce results on a larger scale and with greater realism than is common in this emerging corner of philosophical practice. For example, we will use real-world graph datasets based on co-authorships that contain hundreds of thousands of vertices and edges (Leskovec & Krevl, 2014); and we will run our simulations on modern graph analytics and databases engines (Besta et al., 2020). By enhancing the evidential sophistication of agents in the groups we model, and accounting for structural features of the social groups formed, we opt to increase the fit between our models and the realities they are intended to capture, enhancing the empirical validity of our results.

3.1 Group Attitudes

Prior work on the spread of false belief among rational individuals has not explicitly considered what the findings mean for the group of agents as a whole. For example, O'Connor & Weatherall (2019) found that when agents distrust one another, discounting each other's evidence based on the difference in their credences, the result can be *polarisation* – agents in distinct sub-populations arrive at opposing views despite encountering similar evidence. It seems natural to say that in polarisation, the group as a whole remains agnostic. However, each individual has a strong opinion, one way or the other, and there is nothing in the model that explicitly represents the resulting group ignorance. Our work will fill this lacuna by introducing a special node in the network, the group agent, and considering a variety of methods for aggregating individual attitudes into collective ones, while exploring their effects on group knowledge (and ignorance).

We explore the effects of weighting the contribution of an individual agent's credence to the group's attitude in terms of (a) how many agents in the group 'listen' to her (i.e. access her evidence), and (b) how many agents she 'listens' to. In more sophisticated variants, we also explore (c) an aggregation hierarchy. For example, we introduce group agents representing subgroups, each with its own aggregated credence, and we aggregate these to arrive at the credence of the whole group. When building the proposed weighted aggregation hierarchy, we opt to augment local learning with input from algorithms that weigh agents based on authority (Page et al., 1999) or influence (Kempe et al., 2003) in the network. These algorithms are computationally similar to local learning.

Group attitudes may cause imbalances in the computational workload due to the large number of incoming edges to group agents. We observe that group attitudes do not affect local learning since edges are unidirectional. Thus, group credence updates need not be synchronous with individual ones at each iteration; nor are they totally asynchronous, since we are also interested in the iteration when group attitudes emerge. We propose to explore bounded staleness, where a group agent is guaranteed to read messages from individual agents no older than a fixed number of iterations.

3.2 Higher-Order Evidence

We propose two novel models – confessionals and testimonial reliability – to explore the effects on individual and group credences of various forms of rational sensitivity to higher-order evidence.

Confessionals. With confessional models, we study how higher-order evidence about peer attitudes might affect credences. Confessionals trade off first-order for higher-order evidence. Instead of sharing their observations from when acting according to their credences, agents share their credences directly, together with information about their incoming and outgoing connections to other agents. We explore a number of ways of updating an agent's credence, knowing her neighbours' credences and those of neighbours of neighbours. We compare the results with common intuitions about cases, and in respect of their conduciveness to true belief.

Testimonial reliability. In testimonial reliability models, we embellish Bala and Goyal-style models with higher-order evidence that some of the first-order evidence is unreliable. We model what happens when: (i) a certain fraction of an agent's neighbours report results untruthfully, and (ii) agents know this.

We explore a number of ways of updating an agent's credence, knowing that some of her neighbours' testimony is unreliable. For example, an agent might simply (a) accept all of her neighbours' (reliable and unreliable) testimony; (b) discount her neighbours' first-order evidence at a rate determined by their collective reliability; or, ideally, (c) accept all and only the accurate testimony to which she is exposed. We will explore the relative truth-conduciveness of these methods as the proportion of unreliable agents varies, bringing our results to bear on debates in the epistemology of testimony, and on the relative merits of principles related to the use of higher-order evidence (Dorst, 2020).

By updating credences using higher-order evidence, our models differ from Bala and Goyal's message passing model in two ways: (i) they no longer rely on information from one-hop neighbours, capturing more complex graph properties within a two-hop neighbourhood; and (ii) they introduce an additional relation (or information plane) among agents (testimonial reliability). We could model (i) and (ii) by nesting a localised neighbourhood query in the main learning algorithm that extracts structural and statistical information from neighbours for each agent.

4 Timeframe & Milestones

We propose three major dissemination milestones. One year into the project, we will open-source our simulation framework alongside two research papers, one on metaphysics of groups and one on group attitudes, marking our presence in the area. After one-and-a-half years, we will submit a second pair of research papers (and release code accordingly), one on confessionals and one on testimonials. By the end of second year, we will put the two components together in a major publication. Table 1 breaks the project into a series of (collaborative) tasks.

5 Anticipated Results

Our work will provide new insights into group knowledge and ignorance and the methods that might be deployed to combat misinformation. Our graph-theoretic approach to the metaphysics of groups will enhance our understanding of the aggregation of individual attitudes. Finding the most truth-conducive aggregation technique is open to investigation. We also hope to shed light on philosophical debates surrounding peer disagreement and

Timeframe	P	E	C	Task description
Project set-up & simple group attitudes (month 0 to 6)	•			Understand the metaphysics of groups and their attitudes
		•	•	Reproduce prior results (e.g. polarisation effect)
	•	•	•	Model a single-level group hierarchy
Complex group attitudes (6 to 12)	•	•		Generalise formation of group attitudes to multiple levels
			•	Rank agents by authority or influence
	•	•	•	Model multi-level group credence aggregation hierarchy
Confessionals & testimonial reliability (9 to 21)		•		Build probabilistic framework for higher-order evidence (HOE)
		•	•	Update credences based on multi-hop neighbour evidence
	•	•		Understand effect of HOE on group attitudes
	•	•	•	Vary sensitivity to HOE
Project wrap-up (21 to 24)	•	•	•	Contextualise results to influence policy

Table 1. Tasks in philosophy (**P**), economics (**E**) and computing (**C**)

the epistemology of testimony. For example, we aim to find the most knowledge-conducive ways of accommodating the testimony of others under various conditions.

As a result of our collaboration, philosophers will gain a new tool to design, run and reproduce simulations based on (mis)information diffusion models. Economists will gain new models of opinion aggregation, and information sharing, involving evidentially sophisticated agents comprising structured groups. And computer scientists will gain new graph workloads that can inform the design of systems in the area. These are just three examples of how our work can impact other research at the boundaries of the humanities, the social, and the computing sciences. The APEX award will be the seed to pursue further collaborations and funding that will deepen our research agenda in each subject area. For example, future work might explore temporal graph analytics as a way to model the dynamicity of groups.

Our investigations will impact society by bridging the divide between informing and engaging audiences (Wihbey, 2019). In the UK alone, a cross-parliamentary committee recently reported on disinformation and fake news, exploring routes for the regulation of social media platforms (Digital, Culture, Media & Sport Committee, 2019); and Cairncross (2019) considered the journalistic landscape, noting effects on democratic engagement. Our work relates to the issue of how social media networks and journalistic organisations could maximise group knowledge, and minimise misinformation, reinvigorating democratic decision-making in the process. In future work, we aim to extend our collaborations with organisations encountering real use cases and needing to combat misinformation.

References

Bala, V. & Goyal, S. (1998). Learning from neighbours. *The Review of Economic Studies*, 65(3), 595-621.

Besta, M., Peter, E., Gerstenberger, R., Fischer, M., Podstawski, M., Barthels, C., Alonso, G., & Hoefler, T. (2019). Demystifying graph databases: Analysis and taxonomy of data organization, system designs, and graph queries. arXiv:1910.09017 [cs.DB].

Cairncross, F. (2019). The Cairncross review: A sustainable future for journalism. House of Commons.

Digital, Culture, Media & Sport Committee (2019). *Disinformation and 'fake news': Final report*. Report HC1791, House of Commons.

Dorst, K. (2020). Evidence: A guide for the uncertain. *Philos. Phenomen. Res.*, 100(3), 586-632.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence through a social network. In *Proc. of ACM SIGKDD* (pp. 137-146).

Leskovec, J. & Krevl, A. (2014). SNAP Datasets: Stanford large network dataset collection. snap.stanford.edu.

Malewicz, G., Austern, M. H., Bik, A. J., Dehnert, J. C., Horn, I., Leiser, N., & Czajkowski, G. (2010). Pregel: A system for large scale graph processing. In *Proc. of ACM SIGMOD* (pp. 135-146).

Mayo-Wilson, C. & Zollman, K. J. (2020). The computational philosophy: Simulation as a core philosophical method. Preprint.

O'Connor, C. & Weatherall, O. (2019). *The misinformation age: How false beliefs spread*. Yale University Press.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the Web*. Technical Report 1999-66, Stanford InfoLab.

Park, I., Peacey, M., & Munafò, M. (2014). Modelling the effects of subjective and objective decision making in scientific peer review. *Nature*, 506, 93-96.

Skipper, M. & Steglich-Petersen, A. (2019). *Higher-Order Evidence: New Essays*. Oxford University Press.

Wihbey, J. P. (2019). *The social fact: News and knowledge in a networked world*. MIT Press.