Understanding Clique Formation in Social Networks - An Agent-Based Model of Social Preferences in Fixed and Dynamic Networks

Pratyush Arya(✉) and Nisheeth Srivastava

Indian Institute of Technology Kanpur, Kanpur 208016, India
parya22@iitk.ac.in

Abstract. This paper presents results from *in silico* experiments trying to uncover the mechanisms by which people both succeed and fail to reach consensus in networked games. We find that the primary cause for failure in such games is preferential selection of information sources. Agents forced to sample information from randomly selected fixed neighborhoods eventually converge to a consensus, while agents free to form their own neighborhoods and forming them on the basis of homophily frequently end up creating balkanized cliques. We also find that small-world structure mitigates the drive towards consensus in fixed networks, but not for self-selecting networks. Preferentially attached networks appear to show the highest convergence to one color, thereby showing resilience to balkanization of opinion in self-selecting networks. We conclude with a brief discussion of the implications of our findings for the representation of behavior in socio-cultural modeling.

Keywords: Social preference · preference learning · agent-based modeling · clique formation · balkanization · filter bubbles · polarization

1 Introduction

In the ever-evolving digital landscape of the 21st century, we confront socio-cognitive divides shaped by polarization, filter bubbles, and clique formation. Polarization reflects the increasing divergence of societal and political viewpoints, fragmenting ideological landscapes into opposing extremes [1]. In the realm of social media, this phenomenon is amplified by algorithmic personalization, transforming it into a potent force that drives societies towards divisiveness [2]. In the same vein, filter bubbles, a term birthed by Eli Pariser [3], encapsulate the unsettling reality of intellectual isolation, which now pervades the World Wide Web. Algorithmically generated digital echo chambers present users with content that aligns with their preexisting preferences, reinforcing self-confirming

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
R. Thomson et al. (Eds.): SBP-BRiMS 2023, LNCS 14161, pp. 231–240, 2023.
https://doi.org/10.1007/978-3-031-43129-6_23

information loops. Clique formation materializes as individuals, sharing common attributes or beliefs, clustering together in cyberspace, resulting in islands of homogeneity [4]. Homophily, an age-old sociological phenomenon, has experienced an exponential surge due to the reduction of friction in communication in cyberspace, exerting profound influences on society, politics, and cognitive processes [5].

Previous efforts to comprehend and address these phenomena have predominantly adopted social and cultural perspectives, examining how societal structures, media environments, and cultural contexts shape their manifestations [1,4]. However, there has been a noticeable lack of focus on how these phenomena impact and are influenced by individual cognitive and information processing mechanisms. This gap in the literature signals an uncharted frontier in our understanding of polarization, filter bubbles, and clique formation. The intricate interaction between external stimuli and internal cognitive processes is at the heart of how individuals navigate their social and informational environments. As such, understanding these phenomena from an information processing standpoint is crucial to understand why and when polarization is likely to result in networks of individuals.

In the context of this paper, we operationalize networks of individuals as graphs produced by three different mechanisms, two of which make sociological assumptions: Erdos-Renyi (ER), Barabasi-Albert (BA), and Watts-Strogatz (WS). ER graphs, characterized by random connections between nodes, offer a baseline mathematical graph model, with no socio-cultural appurtenances, for studying network dynamics. On the other hand, BA graphs, generated via a preferential attachment mechanism, exhibit power-law degree distributions, meaning the probability of encountering highly connected nodes is relatively higher. This aligns with the structure of social media and other digital networks, where influential individuals gather larger followings. BA graphs thus have a clear sociological connotation and provide insights into the dynamics of online communities wherein low social friction easily permits large inequalities in the degree distribution of connectivity between individuals. WS graphs, with their small-world architecture, strike a balance between local clustering and global connectivity, reflecting networks in the real-world wherein higher social friction reduces the range of degree distributions accessible to individuals. The small-world property further reflects the interconnected nature of real-world social networks, wherein individuals can establish connections with others through short paths, akin to the "six degrees of separation" concept, without requiring to make a large number of direct connections personally. We examine how social preferences vary over the course of networked consensus games in all three categories of graphs in this paper.

2 Empirical Background for this Work

The motivation, and the empirical background, for this work comes primarily from Michael Kearns' paper, "Behavioral Experiments on a Network Formation

Game" [6]. The paper talks about a series of behavioral experiments where 36 human participants had to solve a competitive coordination task (of biased voting) for monetary compensation. Communication, in these games, happens only via the game GUI, and only with individuals in one's assigned social neighborhood. It has been found that in such cases, where the social neighborhoods are explicitly fixed, and participants are then asked to achieve a collective goal, human participants tend to perform well - subjects are able to extract almost 90% of the value that is available to them in principle. This has led researchers to conclude that humans are quite good at solving a variety of challenging tasks from only local interactions in an underlying network [7].

However, when Kearns made a slight change to the game, human performance deteriorated. The slight change entailed participants having to build the network during the experiment, via individual players purchasing links whose cost is subtracted from their eventual task payoff. A striking finding is that the players performed very poorly compared to behavioral experiments in which network structures were imposed exogenously. Despite clearly understanding the biased voting task, and being permitted to collectively build a network structure facilitating its solution, participants instead built very difficult networks for the task. This finding is in contrast to intuition, case studies and theories suggesting that humans will often organically build communication networks optimized for the tasks they are charged with, even if it means overriding more hierarchical and institutional structures [8,9].

These results suggest that humans are able to achieve a collective goal if a network structure is imposed on them, and they are restricted to communicating within the fixed neighborhood itself; however, when they are free to choose people to communicate with, instead of selecting people that will maximize the chances of global coordination, human participants end up building sub-optimal networks and fail to coordinate effectively.

3 Social Preference Formation

Central to our model is the assumption that the inference of social preferences occurs through the same information processing mechanisms as the inference of individual preferences. Building upon this assumption, our account relies on two specific information-processing assumptions.

Firstly, we embrace the principle of inductive inference, which posits that individuals make decisions by inferring what to do based on their past choices involving similar options. In our model, agents exhibit this inductive reasoning by updating their color preferences based on previous interactions and outcomes, thereby gradually adjust their preferences over time, resulting in the emergence of distinct color clusters.

Secondly, our model incorporates the concepts of memory growth and memory decay. Inspired by the workings of human memory, we assume that agents' memories of past interactions can both strengthen and fade. Memory growth reflects the reinforcement of memory traces associated with interactions that led to similar color preferences, promoting the formation of social ties with like-minded individuals. On the other hand, memory decay represents the natural process of forgetting, allowing agents to adapt and respond to changing social dynamics. These memory dynamics contribute to the evolution of the network structure and the emergence of distinct color clusters in the dynamic network case.

By integrating inductive inference, memory growth, and memory decay into our model, we aim to provide a more comprehensive understanding of how cognitive processes shape social behavior. While our model is a simplified representation of complex human decision-making, it offers insights into the mechanisms underlying social preferences and network dynamics.

3.1 Preference Inference per Iteration

There is now substantial evidence to believe that inductive inference underpins the construction of several (if not all) mental attributes [10]. This Bayesian approach to cognition was recently applied to the problem of preference learning [11]. Following their notation, an agent's preference for an option is identical to the probability that it is desirable, p(r|x), and can be calculated by summing out across evidence of desirability observed in multiple contexts,

$$p(r|x) = \frac{\sum_{c \in C} p(r|x,c)p(x|c)p(c)}{\sum_{c \in C} p(x|c)p(c)} \tag{1}$$

Here C is the set of all contexts offering x as a possible choice. The desirability probability p(r|x, c) simply considers the frequency with which the agent had previously preferred option x in context c, the option probability p(x|c) expresses the frequency with which the option x is observed in context c, and the context probability p(c) expresses the base rate of context c in the agent's environment.

3.2 Memory Growth and Memory Decay Through Iterations

In the context of the model, memory decay and memory growth are parameters that control how the memory matrix evolves over time. The role of these parameters comes in particularly in the case of dynamic network.

Memory decay signifies the gradual decrease in the strength of an agent's memory of past interactions. It models the natural forgetting process in human memory. A higher memory decay rate means that memories of past interactions fade more quickly, while a lower decay rate means that memories persist for a longer time.

$$new\_memory = memory \times (1 - memory\_decay) \tag{2}$$

Memory growth, on the other hand represents the strengthening of an agent's memory of past interactions that have led to similar color preferences. It captures the idea that repeated experiences of similarity reinforce memory traces. A higher memory growth rate means that agents are more likely to remember and interact with agents who have similar color preferences, while a lower growth rate means that memory is less influenced by past interactions.

$$new\_memory[i,j] = memory[i,j] + (similar\_preferences[i,j] \times memory\_growth) \quad (3)$$

where:

- new_memory[i, j] is the updated memory value for agent i's memory of agent j,
- memory[i, j] is the previous memory value for agent i's memory of agent j,
- similar_preferences[i, j] is a measure of the similarity between agent i's and agent j's color preferences,
- memory_growth is a parameter controlling the rate at which memory is reinforced.

We introduce an exponential decay factor to the memory distances, which represents the influence of memory decay. The memory weights are then calculated as the product of the exponential decay factor and the corresponding memory values between agents. This way, we emphasize stronger memories while accounting for the decay process.

The use of the exponential decay factor ensures that closer memory distances and stronger memory values lead to higher memory weights, indicating a higher probability of selecting an agent as a neighbor. The normalization step ensures that the memory weights sum up to 1, providing a valid probability distribution for neighbor selection. In doing so, the neighborhood selection process takes into account both memory growth and memory decay, resulting in the formation of connections based on the strength and recency of agents' memories.

4 Demonstrations and Results

In a typical consensus game, members of a group are permitted to preferentially assign themselves one of a small set of colors, but the entire group is rewarded if it eventually converges to one color. Kearns [6] finds that people are very good at maximizing the group's welfare across a variety of network structures and incentives, so long as the set of their neighbors is held constant: human subjects achieved approximately 90% of the theoretically maximum payout attainable by a perfectly coordinated group.

To assess the behavior of our social preference learning agents, we simulated an environment containing 36 agents, each randomly endowed with one of four color preferences. In other words, for a given agent i, the initial $p_i(r|x) = 1$ for one x, and $= 0$ for the three other xs (colors). The agents could interact with any of the other agents in a sequence. The possible agents with which the initiator i interact with are, from his perspective, the context; thus, interaction partners (responders) are considered c and the interaction is selected by sampling the available neighbors. For simulations using fixed networks, each agent's

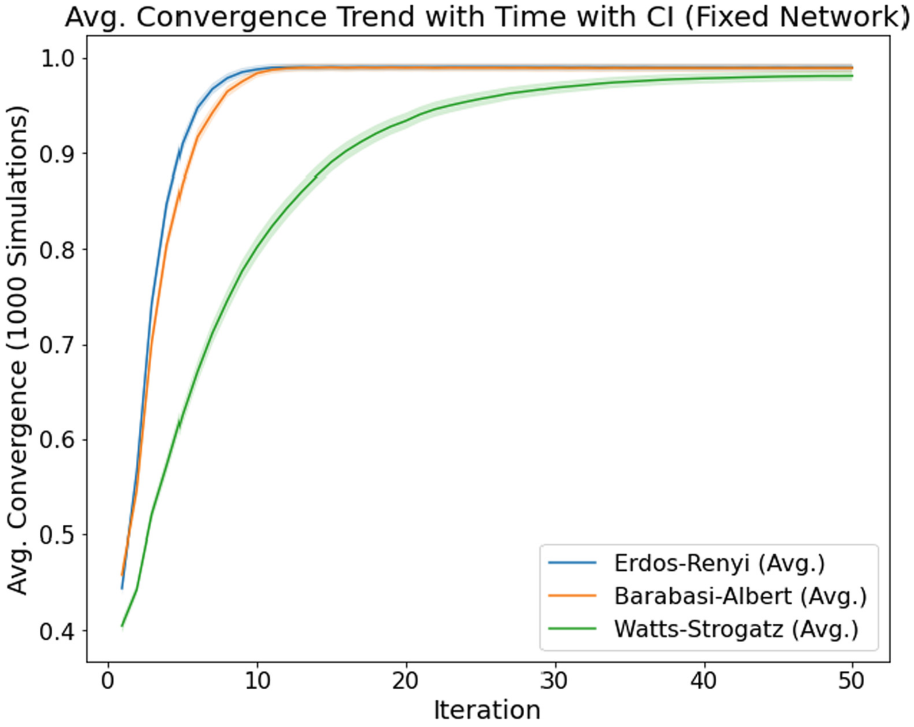
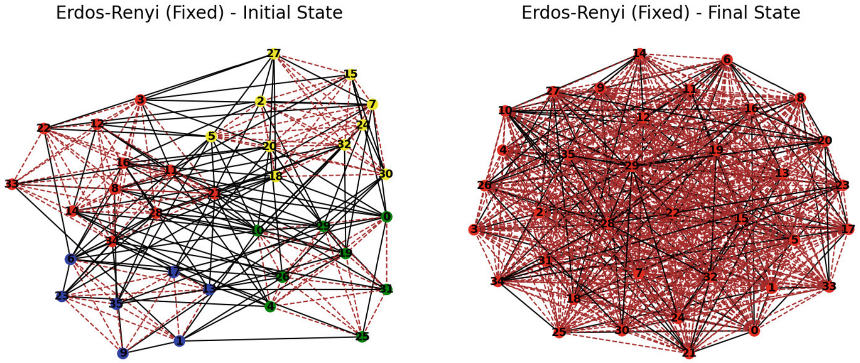


Fig. 1. The plot shows convergence over time for all the fixed network simulations for all three types of graphs. Shaded area represents 95% CI after 1000 simulations. We see convergence for all three graph types.

neighborhood was specified and it could not be changed during the course of the iterations. During an interaction, the responder indicates to the initiator his preferred color ($\arg\max[p(r|x,c)]$), and the responder received no information. At each time step, the initiator updates their own color preferences by marginalizing across the preferences expressed by their neighbors using the preference inference computation mentioned earlier.

We simulate neighborhoods randomly using all three types of graphs - ER, BA, and WS - 1000 times, and report results using the average convergence (the greatest number of nodes converging to a particular color divided by the total number of nodes in the graph at any point in time) obtained for 50 iterations of the consensus game played on each graph for all three categories of graphs in Fig. 1. Even in the absence of an explicitly specified reward for group consensus, our simulation results show that individual agents use the preferences of their neighbors to change their personal preferences, until consensus is reached.

Consistent with the existing literature [12], we find that the color with the greatest representation in the initial condition of each graph wins most frequently (this result simply verifies that under a fixed network structure, our model appropriately propagates beliefs). We also find that the rate of convergence to consensus is directly proportional to the degree of nodes on average in all three types of graphs.


Erdos-Renyi (Fixed) - Initial State

Erdos-Renyi (Fixed) - Final State

Fig. 2. On the left is the initial network for the dynamic network condition in one of the trials, using the ER graph. On the right is the final network after 50 iterations. We see balkanization and cliques formation based on color preference similarities.

But what happens when agents are free to choose their neighbors? When Kearns [7] relaxed the fixed network structure, such that subjects could select which of their neighbors they wished to receive information about, they found that coordination suffered massively, with efficiency dropping to about 40%. It turns out that while humans are extremely good at adapting their preferences to existing network structures, something about the process of social link formation causes this facility of coordination to break down.

We find similar results from our simulation experiment across a broad range of parameter values for memory growth and memory decay. Since network connections were now permitted to be dynamic, agents updated their neighborhoods using encounter information throughout the simulation. At each model iteration, the propensity for interacting with other agents changed, and so did their current preference, using the computation for p(r|x) as above. See Fig. 2 - agents start out with a fixed network, and are then allowed to sample from other agents to update their neighborhood and connections. As a result of this, the final network state (on the right of Fig. 2) turns out to be balkanized.

When updating preferences in fixed network conditions, agents performed the computation as suggested by Eq. 1, and that was enough to get them to global convergence - even in absence of any specified rewards. However, in the case of dynamic networks, when agents were free to choose their neighbors in every iteration, agents retain memories of past interactions, enabling them to recall and potentially favor agents with whom they have had shared color preferences in the past. This memory retention allows for the persistence of social ties and the potential formation of clusters based on shared preferences. This contributes

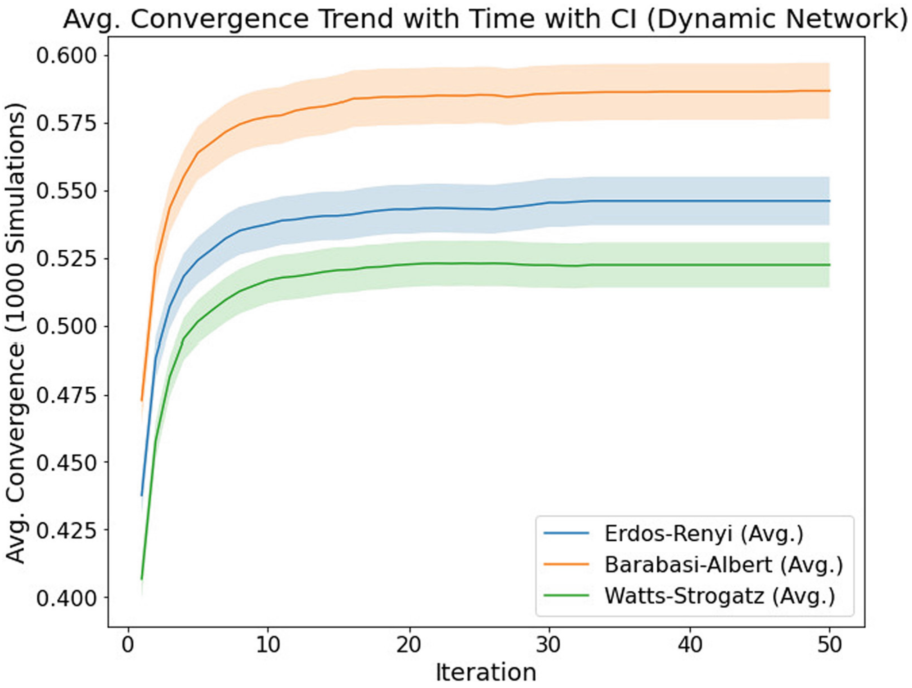


Fig. 3. The plot shows convergence over time for all the fixed network simulations for all three types of graphs. Shaded area represents 95% CI after 1000 simulations.

to the reinforcement of existing social ties, potentially leading to the emergence of cohesive clusters of agents with similar color preferences. This is the case for ER, BA, as well as WS graphs. However, there is a curious differentiation that can be observed when we look at the convergence asymptote value for the three types of graphs across all simulations and all iterations - see Fig. 3 above.

We see that Barabasi-Albert networks show convergence to a higher asymptotic value compared to Watts-Strogatz as well as Erdos-Renyi networks. Considering the structural differences in how the three graphs are generated, we find an interesting explanation for this difference. What makes the BA graph different from the other two is its degree distribution, which follows a power law - thereby increasing the probability of finding nodes that are thickly connected with many neighbors, compared to ER graphs, where the degree distribution is binomially (approximately normally) distributed. Likewise, with WS, we have a small world structure, yielding a close to uniform degree distribution.

For the consensus game, all that matters is the local neighborhood - so, if a node is thickly connected, there is a high chance that it is connected to nodes that have varying colors. If such a node switches over, it's going to have a lot of impact on the rest of the graph. Since we are more likely to see this sort of highly trusted or highly influential node in a BA network than in ER or WS graphs, we see a higher convergence asymptote for BA than ER or WS graphs.

Thus, we find that the same algorithm, when allowed to work with a fixed network structure, performs information coordination efficiently, whereas when allowed freedom to preferentially create local network neighborhoods, agents behave in locally optimal ways that reduce global coordination. We believe these findings explain to a considerable extent the mysterious gap in coordination performance in Kearns' networked game experiments: Agents, and likely humans, assure themselves that they have equilibrated to the consensus preference through sampling the preference of their neighbors. When forced to consider all neighbors, they must necessarily engage with all the information present in their neighborhood; when free to choose, they end up restricting communication with neighbors who share their preference.

5 Conclusion

In this paper, we used a memory-based model of social preference learning to reproduce both the success and failure of agents to attain consensus in a networked game, based on whether agents were permitted to select their social neighborhood. We showed that networks of agents forced to play with neighborhoods assigned to them nearly always converged to a consensus color in the game, although this process was slower for Watts-Strogatz small-world neighborhoods. We also showed that networks of agents permitted to create their own neighborhoods failed to converge to a consensus, with Barabasi-Albert style preferentially attached networks reaching more majority consensus than alternative types.

One alternative to the memory model we used is instance-based learning (IBL). IBL assumes that decision making is based on remembering past experiences and generalizing from these to new situations [13]. The Adaptive Control of Thought - Rational (ACT-R) model is another alternative. ACT-R posits that cognition is composed of a set of basic modules (e.g., visual and auditory), a single production system that coordinates interactions among the modules, and a single declarative memory system that stores factual knowledge [14]. However, while each of these models focuses on different aspects of cognition, they are ultimately just vehicles for the assumptions - and it is these assumptions that determine how accurately the model can predict phenomena in the real world. The choice of model does not fundamentally change our conclusions, so long as the assumptions that guide our model are valid and are themselves representative of the phenomena we seek to understand.

Our findings have theoretical as well as practical implications for enhancing group efficiency and cohesion, particularly in addressing the challenges posed by clique formation and balkanization. By understanding the mechanisms underlying network dynamics and their impact on group behavior, we can also design social media platforms and online communities that foster a less balkanized environment. In particular, our results show that it is not necessary to impose fixed networked structure to prevent balkanization. The presence of highly connected nodes in networks also protects communities from failures in consensus, so long as these nodes are open to changing their colors based on observing their local neighborhood's majority view. Interestingly, these results are consistent with recent empirical work showing that the effect of filter bubbles in large-scale social media may be overstated [15].

Naturally, our current model is highly simplified, and ignores the possibility of alternative reward structures influencing the opinions of individual nodes in the graph. Exploring these possibilities constitutes a clear direction for future work in this project.

References

1. DiMaggio, P., Evans, J., Bryson, B.: Have American's social attitudes become more polarized? Am. J. Sociol. **102**(3), 690–755 (1996)
2. Borgesius, F.J.Z., et al.: Should we worry about filter bubbles? Internet Policy Rev. **5**, 1–16 (2016)
3. Pariser, E.: The Filter Bubble: What the Internet Is Hiding from You. Penguin Press, New York (2011)
4. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: homophily in social networks. Ann. Rev. Sociol. **27**, 415–444 (2001)
5. Sunstein, C.R.: #Republic: Divided Democracy in the Age of Social Media. Princeton University Press, Princeton (2017)
6. Kearns, M., Judd, S., Vorobeychik, Y.: Behavioral experiments on a network formation game. In: Proceedings of the ACM Conference on Electronic Commerce (2012)
7. Kearns, M., Judd, S., Tan, J., Wortman, J.: Behavioral experiments on biased voting in networks. Proc. Natl. Acad. Sci. U.S.A. **106**, 1347–1352 (2009)
8. Burns, T., Stalker, G.M.: The Management of Innovation. Oxford University Press, Oxford (1994)
9. Nonaka, I., Nishiguchi, T.: Fractal design: self-organizing links in supply chain management. In: Knowledge Creation: A Source of Value, pp. 199–230. Ed. St. Martin's Press (2009)
10. Tenenbaum, J.B., Kemp, C., Griffiths, T.L., Goodman, N.D.: How to grow a mind: Statistics, structure, and abstraction. Science **331**, 1279–1285 (2011)
11. Srivastava, N., Schrater, P.: Rational inference of relative preferences, Proceedings of Advances in Neural Information Processing Systems 25, vol. 26 (2012)
12. Tang, J., Wu, S., Sun, J.: Confluence: conformity influence in large social networks. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 347–355 (2013)
13. Gonzalez, C., Lerch, F.J., Lebiere, C.: Instance-based learning in dynamic decision making. Cogn. Sci. **27**, 591–635 (2005)
14. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychol. Rev. **111**, 1036 (2004)
15. Dahlgren, P.M.: A critical review of filter bubbles and a comparison with selective exposure. Nordicom Rev. **42**(1), 15–33 (2021)