How Context Mobility Might Shape Online Discussion Agenda: Using Agent-Based Models to Simulate Topic Community Formation

CMPLXSYS 530 Winter 2020 - Final Report

Yijing Chen

Abstract

This project simulated the process of online community formation in a multi-class topic space, and examined how users' context mobility and bias against unlike-minded others can affect the community structure. The modeling results from two theoretical networks shows that a crowd with higher context mobility is likely to form less stable and looser interest-based communities, and a higher level of discrimination against unlike-minded users will result in better and clearer partitions. The model also reveals that community strength responds differently to the change of probabilities of breaking/connecting at different stickiness level of current connections.

1 Introduction

Social media with varying levels of socio-technical affordances has been facilitating today's political discussions in cyberspace. By breaking the limitations of time and distance that used to circumscribe people's interaction scope, digital platforms now allow users to engage not only in local, real-time discussions, but also in global, asynchronous conversations. Meanwhile, users with similar political orientations or overlapping interests tend to interact more frequently and form online communities, which could potentially turn into echo-chambers where exchanges of fresh perspectives are rare and in-group ideologies get reinforced. The increased possibilities of connecting with distanced others and meanwhile clustering with like-minded people have been shaping the topic agenda and attention distribution in online political discussions. Thus, this project plans to simulate such a process with hypothetical variations in agents' behavioral characteristics.

Past research has empirically examined the qualitative notion of opinion polarization on social media (e.g., Conover et al., 2011; Morales et al., 2015), and usually narrated such debates in a binary framing (i.e., left-leaning or right-leaning). Under this framework, past agent-based modeling (ABM) practices that simulated online user interaction have been adopting a binomial labeling system for model configuration (Li and Tang, 2015; Song and Boomgaarden, 2017). However, political discussions are often multi-faceted, engaging participants with a wide spectrum of opinions on various topics. The binary partition structure oversimplifies the discussion agenda and attention distribution and fails to capture the nuances and subtleties in online discussions. Therefore, this project hopes to capture a finer-grained discussion agenda by introducing a novel ABM approach with multi-class label assignments.

While some users are more actively changing and updating their online connections, and feel more comfortable discussing fresh topics in foreign contexts, others prefer to remain in their current social circles and maintain the status quo by sticking to one or a few topics they are familiar with. I conceptualize this behavioral difference as context mobility, or topic mobility - to measure the extent to which a user is able/willing to freely move around either through establishing new links with others focused on a different topic, or directly switching their topic focus into another one. I further extend this concept to discriminative mobility by introducing a penalizing factor, which essentially represents the degree of discrimination against unlike-minded people with a different topic focus. To examine how these two levels of mobility impacts a macro-level distribution of public attention, this project used agent-based models (ABMs) to observe how the macro-dynamics change under different configurations of mobility parameters.

The simulation result shows that the overall mobility level can pose a negative impact on the com-

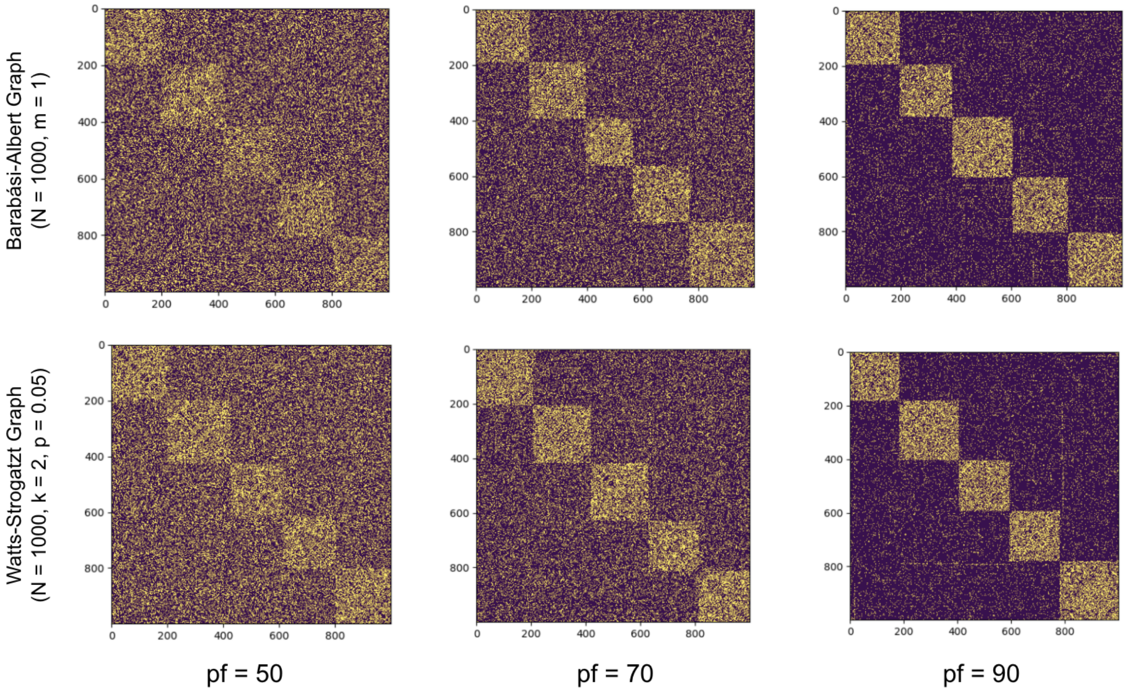

Figure 1: Adjacent matrices generated by the agent-based model. The upper three were simulated in a BA graph and the bottom simulated in a WS graph. From left to right the penalizing factor increases from 50 to 70.

munity strength, the stability of community structure and the convergence speed, and the level of bias towards unlike-minded others can positively influence the community strength.

2 Methods

2.1 Model Description

This project will use PyCX to build an agent-based model, where each agent represents a single user, and the corresponding node is colored based on the topic the user is interested in. The model will update synchronously with discrete timesteps.

2.1.1 Environment

To replicate the scale-free and small-world characteristics of real-world social networks, I tested my model on a Barabasi-Albert (BA) and a Watts-Strogatz (WS) undirected graph with 1000 nodes and similar average degrees[1], such that in both environments, two equal-size groups of people will be interacting with each other at a similar level of interaction intensity. However, as I will discuss in detail in section 3.1, the network dynamics under these two settings do not vary significantly from each other, which implies that the difference in the distribution of original degrees (i.e., whether individual-level interaction intensity is a uniform constant) does not systematically impact the community formation process simulated by this model. Therefore, I performed my parameter sensitivity test only on the BA graph, and I will present the results in Section 3.2.

2.1.2 Agents and Interactions

Each agent has an initial topic focus with a certain level of context mobility. In each timestep, an agent will probably break up/connect with other agents, or change its topic focus. A user's context mobility is reflected by P_b (probability of breaking up a link), P_c (probability of establishing a link) and P_s (probability of changing a topic); and his/her bias level is captured by the penalizing factor (PF). The penalizing factor controls the extent to which he/she discriminates against those who are interested in a different topic when making decisions about establishing/breaking a link or switching to a new topic. By definition,

[1] Graph hyperparameter setup: BA graph - number of edges to attach from a new node to existing nodes (m) = 1, average degree = 1.998; WS graph - number of nearest neighbors each node connected to (k) = 2, probability of rewiring each edge (p) = 0.05, average degree = 2

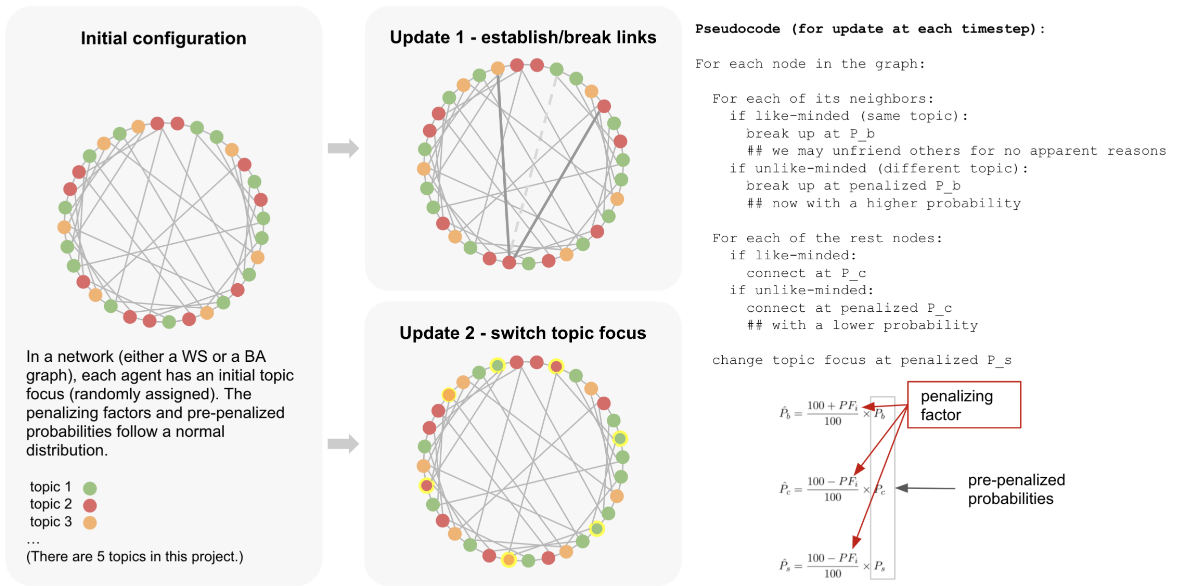

Figure 2: Model diagram

high-mobility users will update their neighborhood and their topic more frequently, and highly-biased users will be more likely to break up with its current neighbors who do not share a common topic, less likely to connect with strangers interested in different topics, and less likely to switch to a new topic.

For example, let X_i be the i_{th} user interested in topic X, and P_{bi}, P_{ci}, P_{si} and PF_i be the corresponding pre-penalized probabilities and penalizing factor. Suppose A_1 is currently connected to A_2 and B_1, but not connected to A_3 and B_2. The probabilities of A_1 breaking up with A_2 and B_1 are P_{bi} and respectively $\hat{P_{bi}}$, where:

$$\hat{P_{bi}} = \frac{100 + PF_i}{100} \times P_{bi} \tag{1}$$

Similarly, the probability of A_1 connecting with A_3 and B_2 are P_{ci} and $\hat{P_{ci}}$, where:

$$\hat{P_{ci}} = \frac{100 - PF_i}{100} \times P_{ci} \tag{2}$$

Finally, the probability of A_1 switching to another topic is $\hat{P_{si}}$, where:

$$\hat{P_{si}} = \frac{100 - PF_i}{100} \times P_{si} \tag{3}$$

I captured the individual differences in behavioral preferences and context mobility levels by sampling the pre-penalized probabilities and penalizing factors from four normal distributed parameter space, which means:

$$P_b \sim \mathcal{N}(\mu_b, \sigma_b^2) \tag{4}$$

$$P_c \sim \mathcal{N}(\mu_c, \sigma_c^2) \tag{5}$$

$$P_s \sim \mathcal{N}(\mu_s, \sigma_s^2) \tag{6}$$

$$P_f \sim \mathcal{N}(\mu_{pf}, \sigma_{pf}^2) \tag{7}$$

2.1.3 Initialization and Model Parameters

To answer the question of how users' context mobility and bias shapes the community structure, the parameters of interest include the mean of pre-penalized probabilities and penalizing factors (i.e., μ_b, μ_c, μ_s and μ_{pf}). The pre-penalized probabilities reflect a general level of users' behavioral preferences of connecting and breaking up with others regardless of whether or not they share a common interest, whereas the penalizing factors determine the extent to which users are likely to cuddle with like-minded people and to avoid those with a divergent focus. For other less relevant parameters, the project sets a topic number of 5 and a universal σ ($\sigma_b = \sigma_c = \sigma_s = 0.1, \sigma_{pf} = 10$).

At the initialization stage, each agent randomly chooses a topic, with its pre-penalized probabilities and its penalizing factor sampled from Gaussian distributions with specified means and standard deviations.

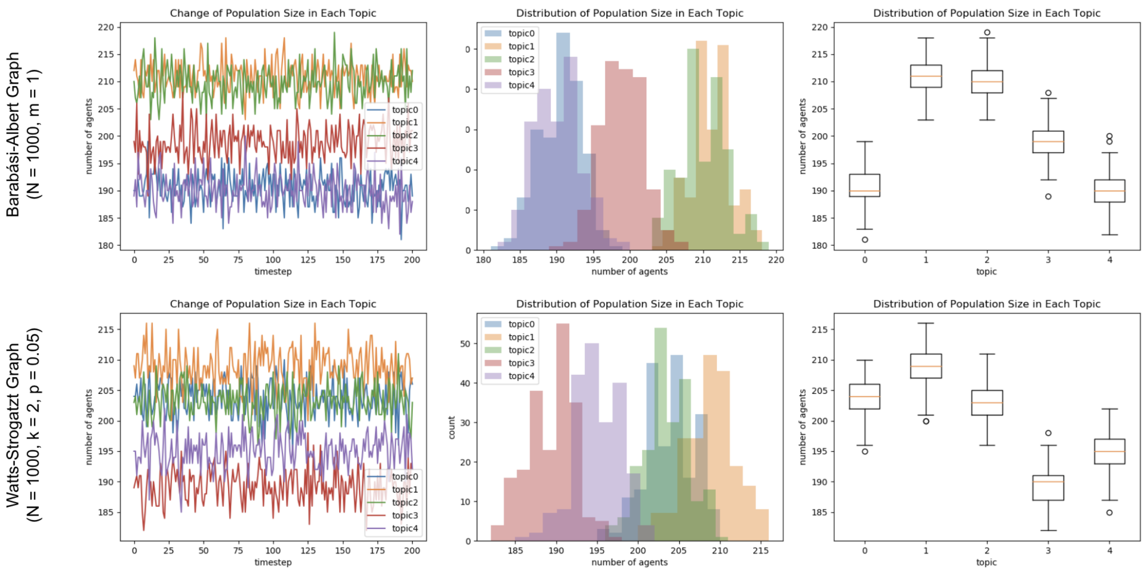

Figure 3: The distribution of topic attention in a BA graph and a WS graph.
Initial topic assignment for the BA graph: 0: [193], 1: [219], 2: [193], 3: [195], 4: [200];
Initial topic assignment for the WS graph: 0: [204], 1: [209], 2: [203], 3: [189], 4: [195]

2.1.4 Model Scheduling

At each time step, the model iterates through each node to decide whether or not it will establish new links, break existing links, or change its topic focus. In the parameter sweeping step, I ran the model for a sufficiently long time to capture general patterns in agent behaviors.

2.2 Model Analysis

To observe whether the model can replicate the process of real-world online community formation, I visualized the network using an adjacent matrix, so that it could clearly exhibit the clusters along the diagonals.

I first ran the model with several configurations of different means of penalizing factors (mu_{pf}) and pre-penalized probabilities (mu_b and mu_c) - to qualitatively evaluate its performance. Following a visual impression that a higher average of penalizing factor contributes to forming a better-partitioned network (see Figure 2) and varying levels of pre-penalized probabilities will produce different but less conclusive results, I calculated the partition coverage (PC) and partition performance (PP) to quantitatively evaluate the sensitivity of cluster strength in terms of PF, mu_b and mu_c. Note that I care about both the partition coverage and performance, as these two ratios measure slightly different aspects of community strength. The partition coverage score reflects the proportion of intra-community edges in total edges, which focus exclusively on the intra-community coherence; the partition performance score sums up the proportion of intra-community edges and inter-non-community edges, and thus takes community separation into consideration.

Moreover, as I noticed that the number of timesteps required to reach a dynamic equilibrium varies with different pre-penalized probabilities, I was motivated to further analyze the change of clustering coefficient through simulation with different mu_b and mu_c.

3 Results and Evaluations

3.1 General Dynamics

The ABM I built can replicate some aspects of online community formations, as it shows that users with a common topic focus tend to establish denser links in between. A model with higher penalizing factors yielded more tightly-clustered communities, with the number of intra-community links significantly larger than that of inter-community links, which makes the topic cluster more salient in the adjacent matrix (see Figure 2). While the community structure appears to reach a dynamic equilibrium after initialization, the community membership keeps fluctuating around a constant mean level that equals to

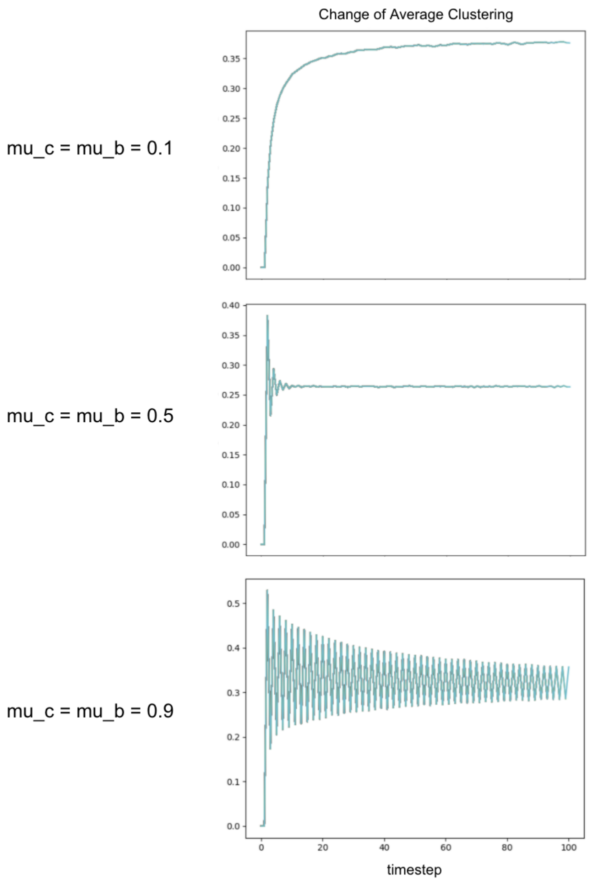

Figure 4: Plot of average clustering with different value sets of P_c and P_b

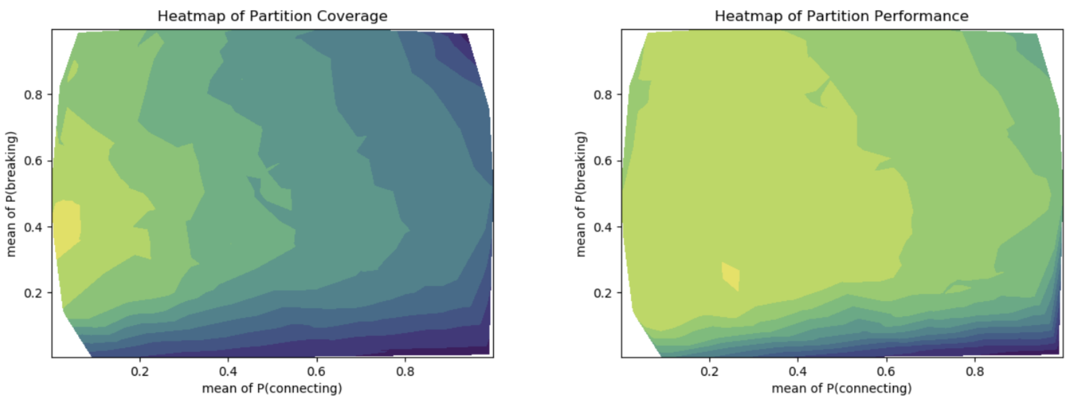

Figure 5: Heatmaps produced by parameter sweeping. A lighter color means a higher value.

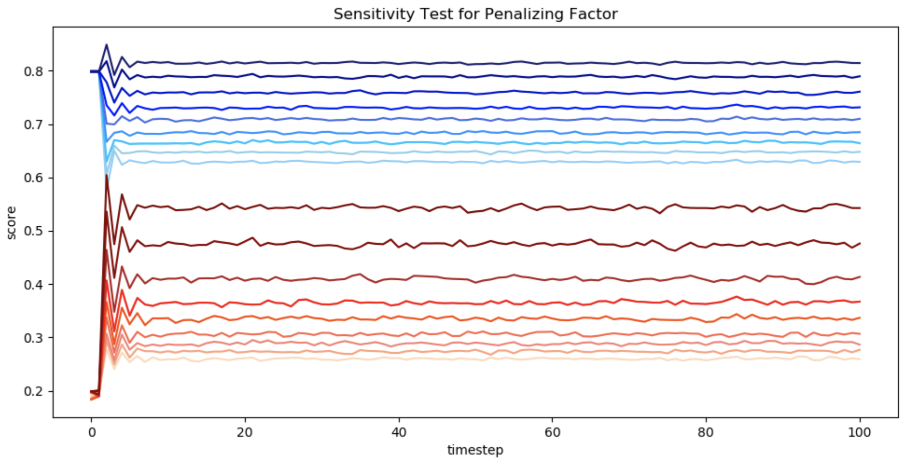

Figure 6: Sensitivity plot of partition scores with regard to the penalizing factor. Blue lines represent partition performance; red lines represent partition coverage. The lighter the line color, the smaller the penalizing factor.

the initial population size of a certain community (See Figure 3). The population size of each topic spreads out in an approximate normal distribution over time.

These patterns can be observed from both the Barabasi-Albert and the Watts-Strogatz graph, which implies that the initial degree distribution does not greatly impact the ultimate community structure or size. Therefore, I used the Barabasi-Albert graph for later parameter sensitivity tests and will discuss the results for BA graph only.

3.2 Sensitivity Tests

After eyeballing the network dynamics, I performed two sets of sensitivity tests to quantitatively measure the extent to which community strength corresponds to the change of penalizing factor and pre-penalized probabilities.

3.2.1 Pre-penalized probabilities

Pre-penalized probabilities determine people's general level of context mobility, that is, the probability of connecting and breaking up with others regardless of whether or not there is interest overlap. When manually tuning these probability parameters, I observed that a model with a higher P_b and P_c could produce topic clusters with more unstable structures, as the salience of diagonal clusters would change dramatically over each timestep. In line with this observation, Figure 4 shows three different convergence patterns of clustering coefficient. With a lower probability of connecting and breaking up with others, the clustering coefficient slowly approaches an upper limit with a diminishing incremental increase. The clustering coefficient would go through a fluctuation phase when the average $mu_b = mu_c = 0.5$; the range of fluctuations and the length of time before the model converge tend to increase as mu_b and mu_c become larger.

In regard to final community strength, Figure 5 shows the final partition coverage (PC) and partition performance (PP) after 150 timesteps. I notice that both PC and PP are very sensitive to the probability of breaking up with neighbors (P_b) and that probability is small; with a higher level of P_b, the change of PC and PP become more responsive to different values of P_c, which is the probability of initiating random new connections.

3.2.2 Penalizing Factor

Since the penalizing factor reflects the individual bias against unlike-minded people, it is by intuition another critical parameter that can influence the community strength and the overall partition quality. Figure 6 shows that community strength can be enhanced by a higher penalizing factor in terms of partition coverage and partition performance, as we can see that the greater the penalizing factor (darker-colored lines), the higher the partition coverage and the partition performance. Moreover, partition coverage appears to be slightly more sensitive to the change in penalizing factor than partition performance, as suggested by a wider overall range of partition coverage score in Figure 6.

4 Discussions

4.1 Interpretations of Results

The project shows the theoretical impact of users' context mobility and preferential biases on the stability and strength of community structures. From the parameter sweepings for context mobility, we can see that the overall level of social and topical fluidity negatively influences the strength and stability of community structures. This confirms our intuition that when people are more aggressively connecting and breaking up with one another, they are less likely to form tightly clustered communities, and it takes longer for the community structure to stabilize.

More interestingly, when agents are in general not very likely to break up with neighbors, which signals a greater level of stickiness to existing connections, the community strength is very sensitive to a small change in their break-up probability. In contrast, when agents are less likely to maintain to their current connections (with a break-up probability higher than 0.2), the community structure no longer actively respond to changes in P_b, and becomes sensitive to the probability of connecting with others. The partition coverage and partition performance showed similar results with regard to the sensitivity patterns at different levels of P_b.

Meanwhile, it is worth noting that the partition performance is less sensitive to the P_c at a higher P_b level, and more sensitive to P_b with a smaller P_b value. This means that when we consider both intra-community coherence and inter-community separation, the enhancement in community strength is not that significant when users (with a moderately high level of P_b) decrease their probability of connecting with others. Similarly, the enhancement in community strength is much more significant when users (who value existing connections very much) increase their probability of breaking up with neighbors.

The sensitivity tests for penalizing factor shows that the discrimination against unlike-minded others would result in greater separation. This again proved our intuition that a preferential bias that people hold towards those who share the same or similar interests would strengthen both the intra-community coherence and inter-community separation. The more people prefer to cuddle with like-minded others, the stronger community they are likely to form.

4.2 Limitations and Future Work

Although the model managed to reasonably display some realistic aspects of community formation, this project has two major limitations that point out directions for future research.

First, this project assumes that the level of context mobility and pre-penalized probabilities follow a normal distribution and set the standard deviations without supports of empirical evidence. Due to the novelty of these concepts, there is no benchmark for reference; future researchers can help to collect these data from real-world digital traces on social media.

Moreover, the single topic assignment for each agent is an unrealistic assumption that each user could only have one topic focus at a time, which apparently oversimplified the dynamics in interest-based community formation. Future studies can try building more sophisticated multi-layer agent-based models, and vary the numbers of topic focus for each agent to capture the individual difference in the span of interests.

References

Michael D Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *Fifth international AAAI conference on weblogs and social media*.

Zhenpeng Li and Xijin Tang. 2015. From global polarization to local social mechanisms: A study based on abm and empirical data analysis. In *Smart modeling and simulation for complex systems*, Springer, pages 29–40.

AJ Morales, Javier Borondo, Juan Carlos Losada, and Rosa M Benito. 2015. Measuring political polarization: Twitter shows the two sides of venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25(3):033114.

Hyunjin Song and Hajo G Boomgaarden. 2017. Dynamic spirals put to test: An agent-based model of reinforcing spirals between selective exposure, interpersonal networks, and attitude polarization. *Journal of Communication* 67(2):256–281.