



DEPARTMENT OF COMPUTER SCIENCE

Link Recommendation Algorithms Affect the Virality of Social Media Influencers

Jadesola Bejide

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Bachelor of Science in the Faculty of Engineering.

Thursday 9th May, 2024

Chapter 3

Base Model

We use the heterogeneous adaptive social network model proposed by Bullock and Sayama (2023) [3] as a base model for our project. This section formally defines the model and its hyperparameters (parameters of the model which are predefined and fixed) and its outcome measures which evaluates the effects of the opinion dynamics on the social network.

3.1 Model Description

The topology of the model is initialised as a complete graph G with a set of nodes V connected by a set of weighted directed edges E excluding self-connections where $|V| = N, |E| = N(N - 1)$. Each node $i \in V$ represents an individual agent with an opinion $x_i \in \mathbb{R}$ such that the opinion is updated continuously. Each node is influenced by each member of its in-neighbourhood and the strength of the influence exerted by each neighbour j on node i is $w_{ij} \in \mathbb{R}_{\geq 0}$, where the network edge is $(j, i) \in E$. All network edge weight are non-negative. Equation 3.1 defines the opinion update policy for each node. One can observe that a node's conformity update policy parameter governs the *rate* of which its opinion approaches the social norm subject to random drift. Equation 3.2 defines the weight update policy of each node such that a node updates the influence each of its immediate neighbours has on it. Equation 3.3, $\langle x \rangle_i$, denotes the weighted local social norm of a node, based on its immediate neighbours, normalised by the node's total in-strength; it may become undefined where an agent's in-strength is zero. If undefined, neither conformity or homophily apply.

The opinion dynamics are defined as:

$$\frac{dx_i}{dt} = c_i(\langle x \rangle_i - x_i) + \epsilon \quad (3.1)$$

$$\frac{dw_{ij}}{dt} = h_i F_h(x_i, x_j) + a_i F_a(\langle x \rangle_i, x_j) \quad (3.2)$$

$$\langle x \rangle_i = \frac{\sum_{j \in N_i} w_{ij} x_j}{\sum_{j \in N_i} w_{ij}} \quad (3.3)$$

where F_h, F_a represent the behavioural functions for homophily and neophily respectively where

$$F_h(x_i, x_j) = \theta_h - |x_i - x_j|$$
$$F_a(\langle x \rangle_i, x_j) = |\langle x \rangle_i - x_j| - \theta_a$$

θ_h, θ_a are fixed population-wide hyperparameters which threshold opinion distances. For the homophily behavioural function F_h , if the opinion distance between an agent i and its neighbour j is less than θ_h , then the edge weight w_{ij} is strengthened, otherwise it is weakened. This represents two individuals building stronger connections with each other if their opinions are similar. Conversely, for the neophily behavioural function F_a , if the distance between the local social norm of agent i and its neighbour j is greater than the threshold θ_a then the edge weight w_{ij} . This represents an individual's attention to novelty, where they are likely to strengthen connections to individuals with different opinions. The effects of both behavioural functions are subject to the individual's homophily and neophily update policy.

As the model uses Euler Forward Integration to progress the model, equations (3.1) and (3.2) can be reformulated as:

$$dx_i = dt \cdot (c_i(\langle x \rangle_i - x_i)) + \epsilon \quad (3.4)$$

$$dw_{ij} = dt \cdot (h_i F_h(x_i, x_j) + a_i F_a(\langle x \rangle_i, x_j)) \quad (3.5)$$

Note that equation 3.4 does not extend multiplying the RHS of equation 3.1 by dt to ϵ . This is not explicitly stated in the original paper, however we found this approach led our MLR results to be closer to the results of the original paper.

3.2 Model Settings

We use the same settings described in **Simulation Settings** of the paper [3]. That is, the number of nodes in the model, N , is set to 1000; the homophily and neophily thresholds are set to $\theta_a, \theta_h = 0.03$; each weight is initialised such that $\forall(j, i) \in E : w_{ij} \sim \mathcal{U}[0, 1]$; each agent opinion is initialised such that $\forall i \in N : x_i \sim \mathcal{N}(0, 1^2)$. We replicate the Uniform Heterogeneous version of the adaptive social network model such that for each node i , $c_i, h_i, a_i \sim \mathcal{U}[0.01, 0.3]$. This is because the original paper quantitatively outlines the results for Uniform Heterogeneity unlike Non-Uniform Heterogeneity as they explain that Uniform Heterogeneity exhibits more statistically significant results. The simulation uses Forward Euler Integration to update the model where $\Delta t = 0.1$ for $t = \{0, \dots, 100\}$. The ϵ in equation 3.1 denotes the stochastic drift of each agent's opinion and is sampled from $\mathcal{N}(0, 0.1^2)$ every time step.

3.2.1 Model Algorithms

Based on our understanding of the **Model** section of Bullock and Sayama's 2023 paper, we have formulated a set of algorithms to make the details of our replication workflow explicit, and these algorithms can be found in Appendix B. This level of detail is not encapsulated in the original paper but we believe it may aid reader understanding of our exact implementation.

3.3 Model Outcome Metrics

This section summarises each outcome measure of the adaptive social network model.

3.3.1 Node Individual Properties

In-Strength and Out-Strength

Each node's in-strength and out-strength are the sum of its incoming weights and outgoing weights respectively. They are defined as

$$\text{in-strength}_i = \sum_{j \in V, i \neq j} w_{ij}$$

$$\text{out-strength}_i = \sum_{j \in V, i \neq j} w_{ji}$$

Each node's in-strength represents the total influence exerted on it by its immediate neighbours and each node's out-strength represents the total influence it exerts on others.

Within Community In-Strength and Within Community Out-Strength

Within community in-strength and within community out-strength are similar to in-strength and out-strength except the set of weights considered are restricted to edges that connect an individual node i to each member, j , of its community, C_i , node j [3]. They are defined as

$$\text{within community in-strength}_i = \sum_{j \in C_i, i \neq j} w_{ij}$$

$$\text{within community out-strength}_i = \sum_{j \in C_i, i \neq j} w_{ji}$$

3.3.2 Deviance and Within-Community-Deviance

Each node's deviance measures how different its opinion is to the average network-level opinion and its within-community-deviance measures how different a node's opinion is to the average opinion of its community. Let $\lceil G \rceil$ denote the average network opinion, where G is the social network model, and $\lceil C_i \rceil$ denote the average opinion of node i 's community. Node deviance and within-community deviance are then defined as

$$\begin{aligned} \text{node-deviance}_i &= |x_i - \lceil G \rceil| \\ \text{within community deviance}_i &= |x_i - \lceil C_i \rceil| \end{aligned}$$

3.3.3 Community Measures

Bullock and Sayama also consider the impact of a node's update policy on its community's size (the number of agents in C_i), average edge weight and the range, standard deviation and deviance of the opinions of the agents in C_i [3]. These are summarised as

$$\begin{aligned} \text{community size}_i &= |C_i| \\ \bar{W}_i &= \frac{1}{|E_{C_i}|} \sum_{k,j \in C_i} w_{kj} \\ \text{range}_i &= \max(x_j) - \min(x_k) \quad \text{where } j, k \in C_i \\ \sigma_i &= \sqrt{\frac{\sum_{k \in C_i} (x_k - \lceil C_i \rceil)^2}{|C_i|}} \\ \text{community deviance}_i &= |\lceil C_i \rceil - \lceil G \rceil| \end{aligned}$$

where E_{C_i} is the set of edges in the community that node i belongs to. The average edge weight of a community defines how tight knit a community is where larger values denote more tight knit communities. The community range defines the diversity of opinions within a community and the community deviance measures how different the average opinion of the community is to the average network-level opinion.

3.3.4 Assortativity

The paper extends the assortativity measure for directed and weighted networks introduced by Yuan et al. [5] to use attributes of the nodes as feature α and β . Specifically, they use the Cartesian 2-product of the following node attributes: opinion x_i , conformity c_i , homophily h_i and neophilicity a_i such that there are 16, 2^4 , assortativity values. s_i^α and s_i^β are updated from the in-degree and out-degree of node i respectively to:

$$\begin{aligned} s_i^\alpha &= \sum_{j \in (\Gamma_i^{in} \cup \Gamma_i^{out})} \alpha_j - \alpha_i \\ s_i^\beta &= \sum_{j \in (\Gamma_i^{in} \cup \Gamma_i^{out})} \beta_j - \beta_i \end{aligned}$$

where α_i is the value of feature α for node i , β_i is the value of feature β for node i and $\Gamma_i^{in}, \Gamma_i^{out}$ are the in-neighbourhood and out-neighbourhood of node i respectively. The symbol \Leftrightarrow refers to bidirectional assortativity which can be positive, $\stackrel{+}{\Leftrightarrow}$, or negative, $\stackrel{-}{\Leftrightarrow}$. \Leftarrow or \Rightarrow refers to the unidirectional assortativity of two network features.

Assortativity is measured at a community-level, as part of the *Community Measures*, and at network-level, as part of the *Population Measures*. Community-level assortativity is calculated by taking the community partition dictionary object produced by `community_louvain.best_partition(graph)` and creating subgraph objects populated with nodes from the same community. Each subgraph is used as input for the weighted assortativity calculations to get an assortativity for each community. For network-level assortativity, the entire social network is used as input.

```

def get_community_assortativities(graph: nx.DiGraph, communities: dict) -> dict:
    assortativities = {}

    unique_communities: set = set(communities.values())

    for community in unique_communities:
        agents = list(filter(lambda agent: communities[agent] == community, graph.nodes))
        features = ['opinion', 'conformity', 'homophily', 'neophily']
        assortativities[community] = directed_weighted_assortativity(
            graph.subgraph(agents),
            features
        )
    return assortativities

```

Figure 3.1: A code snippet of how community-level assortativity is calculated and stored.

We only found one library for the directed weighted assortativity as described by Yuan et al and it is written in R¹. Consequently, we implement the equations ourselves in Python based on the equations outlined in Yuan et al's paper [5] and the method described in Bullock and Sayama's paper to measure the difference between the features.

3.4 Collecting Results & Multiple Linear Regression (MLR)

Bullock and Sayama interpret the results of the model using Multiple Linear Regression. The MLR models for each node's outcome measures and network-level assortativity are trained on 10 final network configurations, amounting to 10×10^3 data entries, as each network has 1000 nodes. The explanatory variables of the MLR models are conformity c_i ; homophily h_i ; neophily a_i ; the interaction between conformity and homophily $c_i h_i$; the interaction between conformity and neophily $c_i a_i$ and the interaction between homophily and neophily $h_i a_i$. The p -values for network-level assortativity of weights w_{ij} are obtained by running a 1-sample t -test on the population-level assortativity. As discussed in section 2.5.4, using a large sample size helps ensure sufficient statistical power. Our replication results and evaluation are presented in chapter 4.

After each simulation, we save the graph object as a .pickle file using the Python library **pickle**. This is later opened by a data collection Python script which computes each outcome measure for each node in the network and compiles the results in a .csv file. We separate simulating the model from data collection as we find that performing both tasks in one supercomputer job is computationally expensive. This allowed for more efficient simulation runs and higher priority on the supercomputer queues, aiding productivity.

We analyse the results using Jupyter Notebooks and utilise Python libraries numpy and Pandas. We use the Python package **statsmodels** to train the multiple linear regression models due to its integration of statistical tooling which enables us to efficiently evaluate the statistical significance of each node update policy. We also use **scipy.stats.ttest_1samp**, with $n = 10$, to perform the one sample t -test for network level assortativity of weights w_{ij} . Alternatively, we considered using **scikit-learn** - whilst both libraries produce the same coefficients for each MLR model, scikit-learn does not have the same useful statistical tools that **statsmodels** and **scipy** provides. Going a step further than the original paper, we repeat this five times to assess the variability of the results. We do this because the model has many sources of randomness, from its assignment of node weights and attributes, the random drift of each agent's opinion and the Louvain Community Detection algorithm.

3.5 Model Development and Criticism

We use Python for development based on previous extensive experience with the language and to take advantage of the graph library *NetworkX*. These advantages come at the expense of runtime, as it is an interpreted, slow language. Static, compiled languages such as C++ or Rust are alternative languages that we could have used which prioritise runtime. To mitigate the performance of Python, we use *Pythonic* features such as generators and list comprehension. Such features are optimised for performance. We also

¹<https://cran.r-project.org/web/packages/wdnet/index.html>

identify and exploit loop invariants and use data memoisation to further speed up code. Our replication models were run on the University of Bristol’s Blue Pebble 1 supercomputer, where the simulation is run in parallel using array jobs. Two nodes were assigned to run five network simulations, each in parallel, to achieve task-level parallelism. The average runtime for the simulation of each network, excluding data collection, is around 80 minutes where the use of Pythonic features helped reduce the runtime by around 45 minutes on average.

The primary cause of the remaining runtime is because the network topology is a complete graph. The code for the replication model consists of for loops where each node is looped over every time step t , and a further nested for loop for each edge connecting the node to a member of its in-neighbourhood. This nested for loop alone has a time complexity of $O(n^3)$, where n is the number of nodes, and this is repeated $T = t/\delta t$ times amounting to a time complexity of $\omega(Tn^3)$.

In addition to the general ambiguity about the exact implementation details of the model, the order that nodes should be updated is also unclear. Various potential update schemes are described in section 2.5.2; our replication model employs random order asynchronous behaviour because it produces results most similar to the original paper’s MLR coefficients.

Finally, we found problems with weights having a lower bound of 0 during development. The adaptive social network model is a directed graph but the Louvain Community Detection algorithm requires an undirected graph. To remediate this, an approximate undirected graph is derived, where each edge is the mean of the edge weights of the bidirectional edges between nodes at the expense of a loss of information. For example, this means that a node i and j could have the associated edges $(i, j, 50)$ and $(j, i, 0)$ between them in the directed network. The approximated edge weight for the undirected network is 25 and the Louvain detection algorithm is likely to put the nodes in the same community. Whilst this is valid, as node i could be very interested in node j but this interest is not reciprocated, if this relationship is the same for all other nodes in the same community such that the within-community-out-degree of node j is 0, then $\sigma_{tar}^\beta = 0$, such that $\rho_{\alpha,\beta}(C_j)$ (where C_j is the community of node j) is undefined. We found that this problem is magnified as the size of the community becomes smaller. Thus, network assortativity is only likely to be well defined when computed on the network as a whole rather than on each community.

Chapter 4

Base Model Results

In this chapter, we begin by summarising Bullock and Sayama’s findings of the relationship between each node’s update policy parameters, and their interactions, on each node’s individual-level, community-level and the network level outcome measures. We compare this to our results where we go a step further from the original paper and assess the effect size of each statistically significant result. Our results indicate that we generally successfully replicate Bullock and Sayama’s model.

There are several sources of randomness in the adaptive social network model. As a result, we present our results using variability analysis, where we perform Multiple Linear Regression for each outcome measure five times to verify the consistency of our findings. At the end of this chapter, we conduct an experiment that demonstrates the inconsistency of the Louvain Community Detection algorithm and validates our need to perform variability analysis.

4.1 Original Findings



Figure 4.1: Network visualisations depict (*left*) a representative network formed under conditions of Uniform Heterogeneity with marker shape indicating community membership and, (*right*), the same network with letters indicating the highest policy parameter for nodes with relatively extreme update policies. In all cases, nodes are coloured by their opinion (using a consistent colour scale) and node placement simulates the effect of (invisible) spring-like edge weights using a standard Fruchterman-Reingold force-directed algorithm. [3]

4.1. ORIGINAL FINDINGS

The left social network visualisation in Figure 4.1 colours nodes based on a consistent scale which represents their opinions [3]. Nodes with similar opinions tend to be placed relatively close to each other such that they are positively assortated on opinion. There is some evidence that nodes with similar opinions end up in the same community but there are also counterexamples of this [3]. In the right social network visualisation, highly neophilic nodes appear to be clustered at the centre of the network such that they are strongly connected with each other. Highly homophilic and highly conformist nodes can be found throughout the network [3]¹.

4.1.1 Multiple Linear Regression Analysis

Bullock and Sayama generalise the findings of their Multiple Linear Regression models, by grouping results from a node's individual-level and community-level outcome measures (Tables 4.1, 4.2 and 4.3) and separate findings for network-level assortativity (Table 4.4) [3]. This includes building a MLR model for community size. However, community size is count data which is not suitable for linear regression [64]. Poisson regression, for example, would have been a more appropriate regression model for this target variable [69]. Nonetheless, for the sake of model replication we also use linear regression for community size.

Tight or loose knit communities are associated with positive and negative average edge weight coefficients respectively; opinion diversity relates to opinion range and opinion standard deviance. The authors of the paper use the magnitude, parity (sign) and statistical significance of each coefficient to make inferences on the effects of each node update policy on individual node and community measures [3].

Table 4.1: Multiple linear regression of a node's *individual-level* outcome measures on its update policy parameters (c_i , h_i , a_i) and their interactions (bottom three rows). Models were built on data from 10^4 nodes obtained from 10×1000 -node networks formed under conditions of *Uniform Heterogeneity*. Statistically significant coefficients are indicated with asterisks (*: $p < 10^{-2}$; **: $p < 10^{-3}$; ***: $p < 10^{-4}$; etc.) [3].

Outcome measure	in-strength	within-community in-strength	out-strength	within-community out-strength
const.	-5571****	-1770****	5134****	1838****
c_i	1494	-2634****	-11720****	-4462****
h_i	17913****	3518****	20606****	8083****
a_i	114823****	43952****	-8784****	-2706****
$c_i h_i$	-16104****	16649****	22008****	7930**
$c_i a_i$	25205****	-3695*	-20315***	-9147***
$h_i a_i$	-441920****	-154036****	-15849*	-6867*

Outcome measure	deviance	within-community deviance
const.	1.55****	0.97****
c_i	-2.47****	-3.74****
h_i	7.07****	-0.24
a_i	-3.94****	4.14****
$c_i h_i$	5.37*	13.81****
$c_i a_i$	-6.01*	0.02
$h_i a_i$	-5.09*	-9.14****

¹Fruchterman-Reingold force-directed algorithm: https://networkx.org/documentation/stable/reference/generated/networkx.drawing.layout.spring_layout.html

Table 4.2: Multiple linear regression of a node’s *community-level* outcome measures on its update policy parameters (c_i, h_i, a_i) and their interactions (bottom three rows). Models were built on the same data as Table 4.1 Statistically significant coefficients are indicated with asterisks (*: $p < 10^{-2}$; **: $p < 10^{-3}$; ***: $p < 10^{-4}$; etc.) [3].

Outcome measure	community size	average edge weight	community deviance	opinion range	opinion std. dev.
const.	337.9****	4.99****	9.9****	1.65****	1.15****
c_i	65.3****	-11.57****	-3.67****	-1.34****	-2.49****
h_i	21.3	-0.83	-0.53	-0.07	4.33***
a_i	-68.1****	12.88****	3.57****	1.45****	-0.07
$c_i h_i$	-295.8****	43.55****	14.12****	4.91****	4.14*
$c_i a_i$	38.6	-0.28	0.74	0.06	7.03****
$h_i a_i$	118.5	-29.05****	-8.57****	-3.34****	-15.21****

Table 4.3: Multiple linear regression of a node’s *community’s assortativity* measures on its update policy parameters (c_i, h_i, a_i) and their interactions (bottom three rows). Models were built on the same data as Tables 4.1 and 4.2. Statistically significant coefficients are indicated with asterisks (*: $p < 10^{-2}$; **: $p < 10^{-3}$; ***: $p < 10^{-4}$; etc.) [3].

Outcome measure	assortativity on opinion	assortativity on conformity (c_i)	assortativity on homophily (h_i)	assortativity on neophilia (a_i)
const.	0.026****	0.004****	0	0.026****
c_i	-0.077****	-0.014****	-0.004****	-0.047****
h_i	-0.006	-0.002	0.001	-0.001
a_i	0.078****	0.013****	0.002	0.054****
$c_i h_i$	0.269****	0.052****	0.004	0.171****
$c_i a_i$	0.021	0.004	0.013*	0.005
$h_i a_i$	-0.163****	-0.032****	-0.003	-0.127****

4.1.2 Individual-Level and Community-Level Measures and Community Assortativity

Each node's individual-level and community-level measures and their community's assortativity aim to reveal the relationship between each node's update policy parameters (alongside their interactions) and the dynamics of the social network model [3]. They evaluate the effect of each node's social psychological traits on their behaviour in an adaptive social network model and the properties of the communities they belong to.

Conformity

Nodes with a higher conformity parameter, c_i , tend to have significantly lower out-strength and community in-strength as shown in Table 4.1 [3]. This means that they tend to have less influence on other nodes and tend to be less influenced by members of their community. Nodes with higher conformity are less deviant, with an opinion close to the average network opinion and close to the average opinion of their community. All of these results were reported to be highly significant with $p < 10^{-4}$ [3]. Highly conformist nodes tend to belong to larger, less tightly-knit and less deviant communities with reduced opinion diversity as shown in Table 4.2. Communities with high levels of conformity are less assorted on opinion, conformity, homophily and neophily as shown in Table 4.3. Conformity is not found to be statistically significant for a node's in-strength such that we cannot explain the relationship between these two variables.

Homophily

Nodes with a higher homophily parameter, h_i , are reported to have significantly higher in-strength and out-strength. This means that they tend to both be more influenced by other nodes and exert more influence on other nodes; with a more deviant opinion at network-level but not within their own community as shown in Table 4.1. The relationship between homophily and opinion diversity; the size and how tight-knit the communities that they belong to is not found to be statistically significant as shown in Table 4.2 [3].

Neophily

Nodes with a higher neophily parameter, a_i , tend to have higher in-strength but lower out-strength at a network and community level as shown in 4.1 [3]. This means that novelty seekers are more influenced by other nodes but exert less influence on other nodes. Their opinions are typically more conformist to the social network but more deviant within their community. This means that nodes attracted to novelty tend to strengthen their connections with other nodes with opinions different to their own; this is consistent with the definition of neophily. Novelty seekers tend to belong to smaller, more tight-knit communities with high opinion diversity as shown in Table 4.2 [3]. We believe there to be a typo in the original paper where they report that nodes are more assorted on opinion, homophily and neophily whereas 4.3 shows that nodes are more assorted on opinion, conformity and neophily.

4.1.3 Network Level Assortativity

Network level assortativity of weights w_{ij} reveals the extent to which the adaptive social network is positively or negatively assorted based on opinion, conformity, homophily and neophily.

At a network level, nodes are significantly positively assorted with respect to their opinion due to the combined effects of conformity and homophily causing nodes with similar opinions to cluster together and form similar opinions [3]. Nodes are significantly positively assorted with respect to conformity. Nodes are also significantly positively assorted with respect to neophily such that that novelty seekers tend to flock together. Nodes are significantly negatively assorted with respect to homophily as highly homophilic nodes separate from each other unless they share the similar opinions.

Conformity and neophily are reported to be bidirectionally negatively assorted with one another such that nodes with a high c_i tend to not strengthen their connection with nodes with a high a_i and vice versa [3].

High homophily nodes tend to have more influence on conformist nodes and this relation is unidirectional [3]. Nodes with a lower homophily update policy parameter, h_i , tend to have more influence on highly neophilic nodes but highly neophilic nodes have more influence on highly homophilic nodes. This means that nodes with high neophily tend to strengthen their connections with nodes with low

Table 4.4: Network level assortativity of weights w_{ij} for various properties of upstream node j and downstream node i . Calculations were based on the same data as Tables 4.1, 4.2 and 4.3. Assortativity coefficients significantly different from zero (one sample t -test, $n=10$) are indicated with asterisks (*: $p < 10^{-2}$; **: $p < 10^{-3}$; ***: $p < 10^{-4}$; etc.) [3].

Property	x_j	c_j	h_j	a_j
x_i	0.0359****	0.0004	0.0011	-0.0011
c_i	0.0005	0.0077****	0.0148***	-0.0226****
h_i	0.0004	-0.0007	-0.0023**	0.0027**
a_i	0.0004	-0.0164****	-0.0267****	0.0407****

homophily but nodes with high homophily tend to strengthen their connections with nodes with low neophily. This may be because nodes with high neophily have an opinion different from nodes with high homophily, such that they strengthen their connections with these nodes.

4.2 Comparing Results

We now report the results of our variability analysis based on our own implementation of the model. We run each MLR model five times with each of the five sets of 10×10^3 datasets and compute the mean and standard deviations of the model coefficients. We define a replicated result to be a coefficient of the original model that is within **one standard deviation** of the mean coefficient. This is visualised succinctly in Appendix C. We assume the results from our variability analysis are normally distributed to plot the normal curve and illustrate the variability of our results at the expense of a loss of information of the true distribution of the variability of the data. Wider curves highlight that the results of our experiment were more variable and tighter curves suggest less variability. The actual results from original paper are indicated with a bold vertical line, where a **green** line suggests that we successfully replicate the results and a **red** line suggests that we did not successfully replicate the results.

We add more depth to the multiple linear regression analysis by also reporting the *effect sizes* of each coefficient. The effect size of each explanatory variable, X_i , is calculated by

$$f^2 = \frac{R^2_{\text{included}} - R^2_{\text{excluded}}}{1 - R^2_{\text{included}}}$$

where R^2_{included} is the R^2 computed when the explanatory variable of interest is included and R^2_{excluded} is where the explanatory variable is excluded [68]. We also report the f^2 for when all explanatory variables are present in X . The original results indicate statistical significance with increasing asterisks. Similarly, when reporting the effect sizes of our replicated models, we use the plus symbol '+' such that + : $0.02 \leq f^2 < 0.15$, ++ : $0.15 \leq f^2 < 0.35$ and +++ : $0.35 \leq f^2$. We do not report the effect sizes for constants or coefficients which are not statistically significant.

4.2.1 Individual-Level Outcome Measures

Table 4.5: Mean coefficients of a node's *individual-level* outcome measures on its update policy parameters (c_i , h_i , a_i) and their interactions from variability analysis. Statistically significant coefficients are indicated with asterisks (*: $p < 10^{-2}$; **: $p < 10^{-3}$; ***: $p < 10^{-4}$; etc.). Effect sizes are indicated with plus signs (+: small; ++: medium; +++: large).

Outcome Measure	in-strength		within-community in-strength		out-strength	
	f^2	11.13	f^2	8.26	f^2	0.49
const.	−5382.11****			−1675.51****		5533.29****
c_i	950.52			−2827.92****	0.01	−13488.53****
h_i	18252.09****	0.07+		3597.69****	0.016	18751.57****
a_i	113757.97****	2.8+++		43064.51****	2.36+++	−9277.90****
$c_i h_i$	−16960.43****	0.003		16114.76****	0.018	29896.93****
$c_i a_i$	28331.38****	0.009		−2274.88		−18617.44**
$h_i a_i$	−442946.44****	2.30+++		−152739.52****	1.60+++	−14983.34*

Outcome Measure	within-community out-strength		deviance		within-community deviance	
	f^2	0.38	f^2	0.32	f^2	0.07
const.	2004.87****			1.60****		1.16****
c_i	−5231.53****	0.012		−2.58****	0.003	−2.41****
h_i	7226.96****	0.023+		6.51****	0.02+	3.90****
a_i	−3028.98****	0.004		−3.99****	0.008	−0.13
$c_i h_i$	11278.07****	0.003		6.96		4.74
$c_i a_i$	−7546.51**	0.0014		−6.50*	0.001	6.14***
$h_i a_i$	−6865.77*	0.0001		−3.62		−12.37****

We replicate up to $\sim 50\%$ of the coefficients across each MLR model relating to a node's individual-level outcome measures within one standard deviation as illustrated by Figure C.1. All other coefficients are on the tail of our variability analysis, excluding coefficients for within community deviance. All individual-level outcome measures excluding those related to deviance have high standard deviations such that the results are highly variable. This shows that the adaptive social network does *not* find consistent results about a node's individual-level outcome measures which is likely to make model replication difficult to assess using regression analysis, especially as Bullock and Sayama only build each MLR model once so we are unaware of the variability of their model's results.

We find very small p -values for several of the coefficients relating to a node's individual-level outcome measures. For example, we observe $p < 10^{-143}$ for the effect of homophily on in-strength. Such small p -values may be attributed to the large amount of data that we train the models on. So, it is important to also report the effect sizes which are independent of the sample size of the MLR model to understand the practical significance of each node property on each outcome measure.

Homophily has a small effect size on the in-strength and neophily and the interaction between homophily and neophily to have a large effect size of 2.8 and 2.30 respectively such that these parameters largely impact the in-strength of each node. Nodes which are highly neophilic or highly homophilic but not both are more likely to strengthen their connections with other nodes. Nodes with both high homophily and neophily are less likely to strengthen their connections with their neighbours, at a network and community level, as there is a contradiction in their preference of the distance between their opinion and the opinions of their neighbours. High homophily also has a small effect on the out-strength, within community out strength and deviance of each node. This affirms that high homophily leads to deviant, fragmented communities. Whilst, we report similar statistical significance of each node update policy parameters on out strength, we find that no other node update policy parameter is of practical significance. We also report no practical significance of the effects of node update policies on the within

4.2. COMPARING RESULTS

community deviance. This shows that, whilst we report statistical significance for a result, the effect size is often negligible such that the result lacks practical significance.

4.2.2 Community-Level Outcome Measures

Many coefficients relating to a node's community-level outcome measures are successfully replicated within one standard deviation, as illustrated in Figure C.2. This excludes opinion std. which is surprising as we mostly replicate the findings for opinion range. Neither we or Bullock and Sayama report statistical significance for the relationship between the interaction between conformity and neophily, $c_i a_i$, on most community-level outcome measures. As it is not significant, we believe the ability to replicate it within one standard deviation is reduced which explains why coefficients for $c_i a_i$ are generally *not* replicated.

High conformity has a small effect on producing loose-knit communities and the interaction between conformity and homophily has a small effect on producing tight-knit communities with reduced opinion diversity. This affirms that the combined effect causes nodes with similar opinions to cluster together and consequently form similar opinions. Perhaps surprisingly, the interaction between conformity and homophily has a small effect size of increasing the diversity of opinions within a community with a larger coefficient than the one for high neophily - which also has a small effect size. This may arise due to the approximation of the social network as an undirected graph which is required to use the Louvain Community Detection algorithm results in a loss of information of the asymmetry of the edge weights. A highly conformist node is likely to strengthen their connections with highly homophilic nodes but not necessarily vice versa. However, the mean edge between a highly conformist and highly neophilic node is still high overall such that the two nodes end up in the same community.

Table 4.6: Mean coefficients of a node's *community-level* outcome measures on its update policy parameters (c_i , h_i , a_i) and their interactions from variability analysis. Statistically significant coefficients are indicated with asterisks (*: $p < 10^{-2}$; **: $p < 10^{-3}$; ***: $p < 10^{-4}$; etc.). Effect sizes are indicated with plus signs (+: small; ++: medium; +++: large).

Outcome Measure	community size		average edge weight		community deviance		opinion range	
	f^2	0.06		0.17		0.04		0.15
	coef.	f^2	coef.	f^2	coef.	f^2	coef.	f^2
const.	340.97****		5.08****		10.39****		1.68****	
c_i	91.58*	0.009	-12.49****	0.03 ⁺	-3.67****	0.001	-1.37****	0.03 ⁺
h_i	1.78		-0.66		-0.19		-0.07	
a_i	-113.19****	0.013	12.90****	0.03 ⁺	3.41**	0.007	1.38****	0.03 ⁺
$c_i h_i$	-352.08		46.45****	0.02 ⁺	13.39****	0.006	5.07****	0.02 ⁺
$c_i a_i$	23.59		3.33		1.98		0.44	
$h_i a_i$	262.46*	0.004	-33.27****	0.01	-9.75*	0.003	-3.61****	0.01

Outcome Measure	opinion std. dev.	
	f^2	0.16
	coef.	f^2
const.	1.00****	
c_i	-3.99****	0.03 ⁺
h_i	-0.21	
a_i	4.01****	0.03 ⁺
$c_i h_i$	14.67****	0.02 ⁺
$c_i a_i$	1.31	
$h_i a_i$	-10.38****	0.01

4.2.3 Community and Network Level Assortativity

Table 4.7: Mean coefficients of a node's *community's assortativity* measures on its update policy parameters (c_i , h_i , a_i) and their interactions from variability analysis. Statistically significant coefficients are indicated with asterisks (*: $p < 10^{-2}$; **: $p < 10^{-3}$; ***: $p < 10^{-4}$; etc.). Effect sizes are indicated with plus signs (+: small; ++: medium; +++: large).

Outcome Measure	assortativity on opinion		assortativity on conformity (c_i)		assortativity on homophily (h_i)		assortativity on neophily (a_i)	
f^2	0.13		0.03		0.14		0.06	
	coef.	f^2	coef.	f^2	coef.	f^2	coef.	f^2
const.	0.0255****		0.0011****		0.0255****		0.0043****	
c_i	-0.0787****	0.02+	-0.0072****	0.006	-0.0430****	0.02+	-0.0174****	0.012
h_i	-0.0057		0.0001		-0.0016		-0.0006	
a_i	0.0808****	0.02+	0.0041		0.0457****	0.03+	0.0162****	0.01
$c_i h_i$	0.2942****	0.016	0.0158		0.1598****	0.017	0.0613****	0.008
$c_i a_i$	0.0183		0.0144*	0.001	0.0110		0.0113	
$h_i a_i$	-0.1973****	0.007	-0.0129		-0.1167****	0.009	-0.0477****	0.005

Table 4.8: Network level assortativity of weights w_{ij} variability analysis. Statistically significant coefficients are indicated with asterisks (*: $p < 10^{-2}$; **: $p < 10^{-3}$; ***: $p < 10^{-4}$; etc.).

Property	x_j	c_j	h_j	a_j
x_i	0.0348****	0.0001	0.0005	0.0003
c_i	0.0003	0.0082**	-0.0181****	-0.0012***
h_i	-0.0000	-0.0216****	0.0401****	0.0030
a_i	0.0002	0.0151****	-0.0279****	-0.0025***

We successfully replicate the findings of opinion assortativity within one standard deviation of our identified mean, excluding the coefficient which explains the relationship between opinion assortativity and the interaction between homophily and neophily. On the other hand, we do *not* successfully replicate the findings for assortativity on conformity, homophily or neophily. We also report minimal practical significance of a node's update policy parameters on their community's assortativity. We report that each node's conformity, c_i , and neophily, a_i (but not their interaction) has a small effect on the negative assortativity on opinion and homophily but we report no other practical significance for the other node update policy parameters despite high statistical significance. Additionally, all coefficients have small magnitudes which approach zero such that there is a weak relationship between each node update policy parameter, their interactions and community assortativity.

For network level assortativity of the weights w_{ij} , we also do not successfully replicate the findings for $h_i \Rightarrow c_j, a_i \Rightarrow c_j, c_i \Rightarrow h_j, h_i \Rightarrow h_j, c_i \Rightarrow a_j$ or $a_i \Rightarrow a_j$. All such coefficients are not within the distribution of our variability analysis. Assortativity, $\rho_{\alpha,\beta}(G)$, is in the range [-1,1] where values which approach -1 suggest negative assortativity and values which approach 1 suggest positive assortativity. Each network level assortativity value for opinion, conformity, homophily and neophily approaches zero, and thus are neither moderately positively nor negatively assortated. Given our results are generally consistent with the original paper, we believe that the assortativity of the social network is perhaps overstated.

4.2.4 Summary

Overall, whilst we replicate most coefficients for each MLR model, we are less confident in the adaptive social network model's ability to exhibit consistent results about the effect of each node update policy parameters and their interactions on the individual-level and community-level outcome measures. Additionally, we do not find it valuable to report assortativity at a network-level and community-level for assortativity values which approach zero. For all MLR models, there are many statistically significant results however only few results have a reasonable effect size to suggest practical significance. Nonetheless, we have implemented an adaptive social network model with heterogeneous opinion dynamics as a

simple representation of the mechanics of a real online social network. We believe that this model is still a useful base model for our research on the effects of recommendation algorithm on social networks and influencer performance.

4.3 Louvain Community Detection is Inconsistent

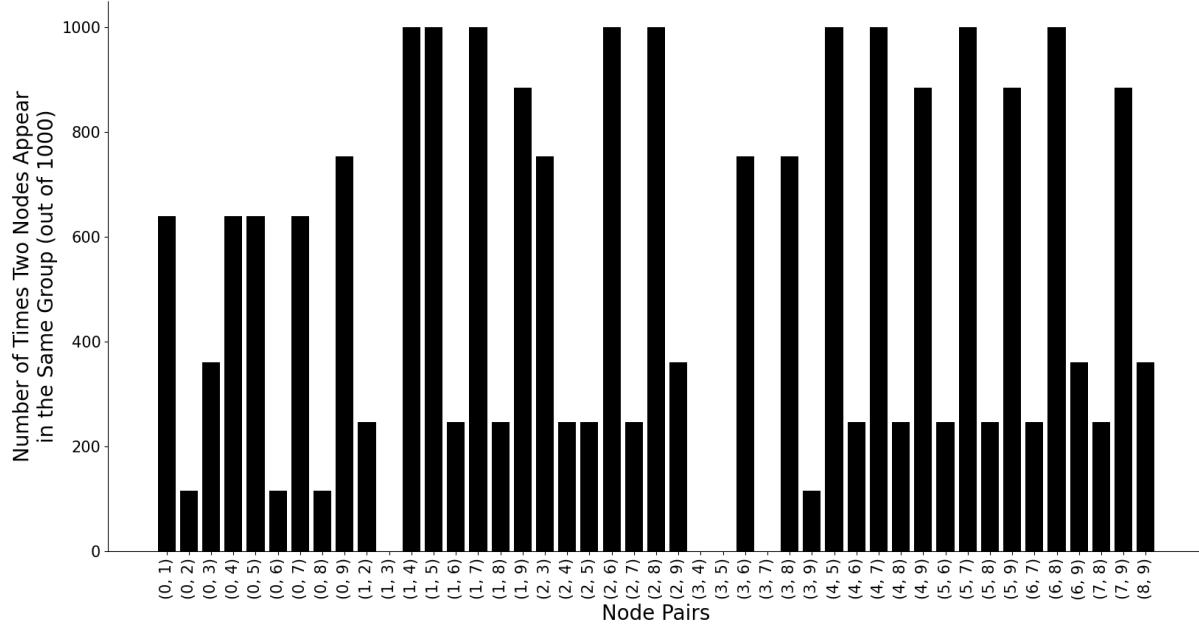


Figure 4.2: 1000 runs of the Louvain detection algorithm on the same graph with 10 nodes.

The Louvain Community Detection algorithm is heuristic with many local optima such that its results are potentially inconsistent. To affirm this, we conduct an experiment where Louvain Community Detection is applied to the same social network 1000 times, and assess the number of times each node pair appears in the same community. The algorithm is consistent if each node pair appears in the same community 1000 times (always) or 0 times (never). In Figure 4.2, we find that the number of times pairs of nodes appear in the same community is *not* consistent. For example, node pair **(1,4)** always appears in the same community over 1000 computations of the Louvain Community Detection algorithm but node pair **(0,9)** only appears in the same community under 80% of the time.

Thus, we are less confident in community-level or a node's community assortativity outcome measures as the inconsistency of the Louvain Community Detection method hinders replication of model results. This is especially as the network modularity of the adaptive social network approaches zero [3], such that the formation of communities is no better than random [4]. However, one could increase the number of times that the adaptive social network model is run, as we do in our variability analysis, to hopefully converge to some mean result that better captures the outcomes of the model.

Bibliography

- [1] Council of the European Union, “Regulation (eu) 2022/2065 of the European Parliament And of the Council,” 2022.
- [2] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, p. P10008, Oct. 2008.
- [3] S. Bullock and H. Sayama, “Agent Heterogeneity Mediates Extremism in an Adaptive Social Network Model,” 2023.
- [4] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Phys. Rev. E*, vol. 69, p. 026113, Feb 2004.
- [5] Y. Yuan, J. Yan, and P. Zhang, “Assortativity measures for weighted and directed networks,” *Journal of Complex Networks*, vol. 9, no. 2, p. cnab017, 2021.
- [6] R. Lempel and S. Moran, “SALSA: the stochastic approach for link-structure analysis,” *ACM Trans. Inf. Syst.*, vol. 19, p. 131–160, apr 2001.
- [7] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2022.
- [8] A. Barinka, “Meta’s Instagram Users Reach 2 Billion, Closing In on Facebook.” <https://www.bloomberg.com/news/articles/2022-10-26/meta-s-instagram-users-reach-2-billion-closing-in-on-facebook>, October 2022. (Accessed on 01/05/2024).
- [9] “Year on Tiktok 2023: Scroll back with our community.” <https://newsroom.tiktok.com/en-sg/year-on-tiktok-2023-sg>, 2023. (Accessed on 01/05/2024).
- [10] F. P. Santos, Y. Lelkes, and S. A. Levin, “Link recommendation algorithms and dynamics of polarization in online social networks,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 50, p. e2102141118, 2021.
- [11] C. Goantă and S. Ranchordás, *The regulation of social media influencers*. Elgar law, technology and society, Cheltenham, UK ;: Edward Elgar Publishing, 2020.
- [12] R. M. Samantha Kay and J. Parkinson, “When less is more: the impact of macro and micro social media influencers’ disclosure,” *Journal of Marketing Management*, vol. 36, no. 3-4, pp. 248–278, 2020.
- [13] E. Djafarova and C. Rushworth, “Exploring the credibility of online celebrities’ Instagram profiles in influencing the purchase decisions of young female users,” *Computers in human behavior*, vol. 68, pp. 1–7, 2017.
- [14] T. Lorenz, “The Lil Tay Saga Reaches Its Logical Conclusion,” *The Atlantic*, 2018.
- [15] S. Javed, “Andrew Tate shares ‘final message’ after being banned from social media,” *The Independent*, 2022.
- [16] “Controversial influencer Andrew Tate banned from Facebook and Instagram,” *Sky News*, 2022.

BIBLIOGRAPHY

- [17] M. R. Dickey, “The 22 Key Turning Points In The History of Youtube,” *Business Insider*, vol. 15, 2013.
- [18] Oxford English Dictionary, “Gig Economy,” July 2023.
- [19] M. De Veirman, V. Cauberghe, and L. Hudders, “Marketing through instagram influencers: the impact of number of followers and product divergence on brand attitude,” *International journal of advertising*, vol. 36, no. 5, pp. 798–828, 2017.
- [20] S. Kusumasondjaja and F. Tjiptono, “Endorsement and visual complexity in food advertising on Instagram,” *Internet Research*, vol. 29, no. 4, pp. 659–687, 2019.
- [21] Merriam-Webster, “Viral,” April 2024. (Accessed on 06/05/2024).
- [22] N. Confessore, “Cambridge Analytica and Facebook: The scandal and the fallout so far,” *The New York Times*, vol. 4, p. 2018, 2018.
- [23] A. W. Palmer, “How Tiktok Became a Diplomatic Crisis.,” *International New York Times*, 2023.
- [24] D. Garcia, “Social Data Science.” <https://dgarcia-eu.github.io/SocialDataScience/>, 2023. Accessed: 2024-01-29.
- [25] G. Simmel, “Fashion,” *American Journal of Sociology*, vol. 62, no. 6, pp. 541–558, 1957.
- [26] R. K. Merton, “The Matthew Effect in Science, ii: Cumulative Advantage and the Symbolism of Intellectual Property,” *Isis*, vol. 79, no. 4, pp. 606–623, 1988.
- [27] A. Ferrara, L. Espin-Noboa, F. Karimi, and C. Wagner, “Link recommendations: Their impact on network structure and minorities,” in *14th ACM Web Science Conference 2022*, WebSci ’22, ACM, June 2022.
- [28] M. Sherif, “The psychology of social norms.,” 1936.
- [29] S. Milgram and C. Gudehus, “Obedience to authority,” 1978.
- [30] R. Hackforth, *Plato: Phaedrus*. No. 119, Cambridge University Press, 1972.
- [31] H. Stewart, “‘Andrew Tate is a symptom, not the problem’: why young men are turning against feminism,” *The Guardian*, 2024.
- [32] T. Veblen, *The theory of the leisure class*. Routledge, 2017.
- [33] P. ERDdS and A. R&wi, “On random graphs I,” *Publ. math. debrecen*, vol. 6, no. 290-297, p. 18, 1959.
- [34] P. Erdős, A. Rényi, *et al.*, “On the evolution of random graphs,” *Publ. math. inst. hung. acad. sci*, vol. 5, no. 1, pp. 17–60, 1960.
- [35] A. A. Saberi, “Recent advances in percolation theory and its applications,” *Physics Reports*, vol. 578, pp. 1–32, 2015. Recent advances in percolation theory and its applications.
- [36] R. Motwani and P. Raghavan, “Randomized algorithms,” *ACM Computing Surveys (CSUR)*, vol. 28, no. 1, pp. 33–37, 1996.
- [37] L. Page, S. Brin, R. Motwani, T. Winograd, *et al.*, “The pagerank citation ranking: Bringing order to the web,” 1999.
- [38] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical Review E*, vol. 69, Feb. 2004.
- [39] M. E. Newman, “Assortative mixing in networks,” *Physical review letters*, vol. 89, no. 20, p. 208701, 2002.
- [40] J. A. Holyst, K. Kacperski, and F. Schweitzer, “Social impact models of opinion dynamics,” *Annual Reviews Of Computational PhysicsIX*, pp. 253–273, 2001.

- [41] H. Xia, H. Wang, and Z. Xuan, “Opinion Dynamics: A Multidisciplinary Review and Perspective on Future Research,” *IJKSS*, vol. 2, pp. 72–91, 10 2011.
- [42] R. Axelrod, “The dissemination of culture: A model with local convergence and global polarization,” *Journal of conflict resolution*, vol. 41, no. 2, pp. 203–226, 1997.
- [43] H. Sayama, “Enhanced ability of information gathering may intensify disagreement among groups,” *Physical Review E*, vol. 102, no. 1, p. 012303, 2020.
- [44] H. Sayama, “Social Fragmentation Transitions in Large-Scale Parameter Sweep Simulations of Adaptive Social Networks,” in *International Conference on Parallel Processing and Applied Mathematics*, pp. 459–469, Springer, 2022.
- [45] J. Bryden, S. Funk, N. Geard, S. Bullock, and V. A. Jansen, “Stability in flux: community structure in dynamic networks,” *Journal of The Royal Society Interface*, vol. 8, no. 60, pp. 1031–1040, 2011.
- [46] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.
- [47] D. Liben-Nowell and J. Kleinberg, “The link prediction problem for social networks,” in *Proceedings of the twelfth international conference on Information and knowledge management*, pp. 556–559, 2003.
- [48] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh, “Wtf: The who to follow service at twitter,” in *Proceedings of the 22nd international conference on World Wide Web*, pp. 505–514, 2013.
- [49] “Twitter’s Recommendation Algorithm,” Mar 2023.
- [50] A. Grover and J. Leskovec, “node2vec: Scalable feature learning for networks,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864, 2016.
- [51] A. Grover and J. Leskovec, “node2vec: Scalable Feature Learning for Networks,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
- [52] K. Henderson, B. Gallagher, T. Eliassi-Rad, H. Tong, S. Basu, L. Akoglu, D. Koutra, C. Faloutsos, and L. Li, “Rox: structural role extraction & mining in large graphs,” in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1231–1239, 2012.
- [53] A. A. Hagberg, D. A. Schult, and P. J. Swart, “Exploring Network Structure, Dynamics, and Function using Networkx,” in *Proceedings of the 7th Python in Science Conference* (G. Varoquaux, T. Vaught, and J. Millman, eds.), (Pasadena, CA USA), pp. 11 – 15, 2008.
- [54] Y. Shmargad and S. Klar, “How partisan online environments shape communication with political outgroups,” *International Journal of Communication*, vol. 13, p. 27, 2019.
- [55] R. Levy, “Social media, news consumption, and polarization: Evidence from a field experiment,” *American economic review*, vol. 111, no. 3, pp. 831–870, 2021.
- [56] E. M. Daly, W. Geyer, and D. R. Millen, “The network effects of recommending social connections,” in *Proceedings of the fourth ACM conference on Recommender systems*, pp. 301–304, 2010.
- [57] J. Su, A. Sharma, and S. Goel, “The effect of recommendations on network structure,” in *Proceedings of the 25th international conference on World Wide Web*, pp. 1157–1167, 2016.
- [58] M. Zignani, S. Gaito, G. P. Rossi, X. Zhao, H. Zheng, and B. Zhao, “Link and triadic closure delay: Temporal metrics for social network dynamics,” in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, pp. 564–573, 2014.
- [59] J. C. Butcher, *Numerical methods for ordinary differential equations*. John Wiley & Sons, 2016.
- [60] D. Cornforth, D. G. Green, and D. Newth, “Ordered asynchronous processes in multi-agent systems,” *Physica D: Nonlinear Phenomena*, vol. 204, no. 1-2, pp. 70–82, 2005.

BIBLIOGRAPHY

- [61] M. Gardner, “The Fantastic Combinations of John Conway’s New Solitaire Game ‘Life’,” *Sc. Am.*, vol. 223, pp. 20–123, 1970.
- [62] M. A. Nowak, S. Bonhoeffer, and R. M. May, “Spatial games and the maintenance of cooperation.,” *Proceedings of the National Academy of Sciences*, vol. 91, no. 11, pp. 4877–4881, 1994.
- [63] I. Harvey and T. Bossomaier, “Time out of joint: Attractors in asynchronous random boolean networks,” in *Proceedings of the fourth European conference on artificial life*, pp. 67–75, Citeseer, 1997.
- [64] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [65] B. Lantz, “The large sample size fallacy,” *Scandinavian Journal of Caring Sciences*, vol. 27, no. 2, pp. 487–492, 2013.
- [66] G. M. Sullivan and R. Feinn, “Using effect size—or why the P value is not enough,” *Journal of graduate medical education*, vol. 4, no. 3, pp. 279–282, 2012.
- [67] J. Cohen, “Things I have learned (so far).,” in *Annual Convention of the American Psychological Association, 98th, Aug, 1990, Boston, MA, US; Presented at the aforementioned conference.*, American Psychological Association, 1992.
- [68] J. Cohen, “Statistical power analysis for the behavioral sciences (revised ed.),” 1977.
- [69] P. Roback and J. Legler, *Beyond multiple linear regression: applied generalized linear models and multilevel models in R*. Chapman and Hall/CRC, 2021.
- [70] X. N. Lam, T. Vu, T. D. Le, and A. D. Duong, “Addressing cold-start problem in recommendation systems,” in *Proceedings of the 2nd international conference on Ubiquitous information management and communication*, pp. 208–211, 2008.
- [71] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.
- [72] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.

Appendix B

Base Model Algorithms

Algorithm 2 Initialises the node attributes of each agent in the social network.

```
function INITIALISEPARAMETERS(uniform, agent)
    if uniform then
        Set agent homophily as  $h_i \sim \mathcal{U}[0.01, 0.3]$ 
        Set agent neophilia as  $a_i \sim \mathcal{U}[0.01, 0.3]$ 
        Set agent conformity as  $c_i \sim \mathcal{U}[0.01, 0.3]$ 
    else
        Set agent homophily as  $h_i \sim \mathcal{N}(0.05, 0.025)$ 
        Set agent conformity as  $c_i \sim \mathcal{N}(0.05, 0.025)$ 
        Set agent neophilia as  $a_i \sim \mathcal{N}(0.25, 0.025)$ 
    end if
end function
```

Algorithm 3 Initialises the social network as a complete digraph and each agent's opinion.

```
function INITIALISENETWORK( $N$ , uniform)
    let the social_network be a complete weighted directed graph with no self-connections of size  $N$ 
    for each agent $_i$  in the social_network do
         $x_i \sim \mathcal{N}(0, 1^2)$ 
        INITIALISEPARAMETER(uniform, agent $_i$ )
    end for
    return social_network
end function
```

Algorithm 4 Finds the local social norm of an agent

```
function SOCIALNORM(social_network, agent)
    let total_weights be the sum of all in-weights of INNEIGHBOURHOOD(agent)
    let opinion_weighted_sum be the sum of  $w_{ij} \cdot x_j$  of all neighbours in INNEIGHBOURHOOD(agent)
    if total_weights = 0 then
        return undefined
    end if
    return opinion_weighted_sum ÷ total_weights
end function
```

Algorithm 5 Updates the weight of the network edge between an agent and each member of its in-neighbourhood.

```

function UPDATEWEIGHTS(social_network, agent, social_norm,  $\theta_h$ ,  $\theta_a$ ,  $\Delta t$ )
    for neighbour in INNEIGHBOURHOOD(agent) do
        if social_norm is undefined then
             $w_{ij} = \max(0, w_{ij} + (\Delta t \cdot (\theta_h * (|x_i - x_j|))))$ 
        else
            let homophily_effect =  $h_i \cdot (\theta_h - |x_i - x_j|)$ 
            let neophily_effect =  $a_i \cdot (|\langle x \rangle_i - x_j| - \theta_a)$ 
             $w_{ij} = \max(0, w_{ij} + (\Delta t \cdot (\text{homophily\_effect} + \text{neophily\_effect})))$ 
        end if
    end for
end function

```

Algorithm 6 Simulates the adaptive social network using opinion dynamics and Euler's Forward Integration Method

Require: Network size N , number of iterations n , Δt

```

function SIMULATEMODEL( $N$ ,  $n$ ,  $\Delta t$ , uniform)
    let network = INITIALISENETWORK( $N$ , uniform)
    let  $\theta_h, \theta_a = 0.03$ 
    let  $t = 0$ 
    while  $t < n$  do
        for agenti in SHUFFLE(network.nodes) do
            let social_norm = SOCIALNORM(agenti, social_network)
            if social_norm is defined then
                 $x_i = x_i + \Delta t \cdot (c_i \cdot (\langle x \rangle_i - x_i))$ 
            end if
            let  $e_i \sim \mathcal{N}(0, 0.1)$ 
             $x_i = x_i + e_i$ 
            UPDATEWEIGHTS(network, node,  $\theta_h, \theta_a$ , social_norm)
        end for
         $t = t + \Delta t$ 
    end while
end function

```

Appendix C

Base Model Variability Analysis Visualisations

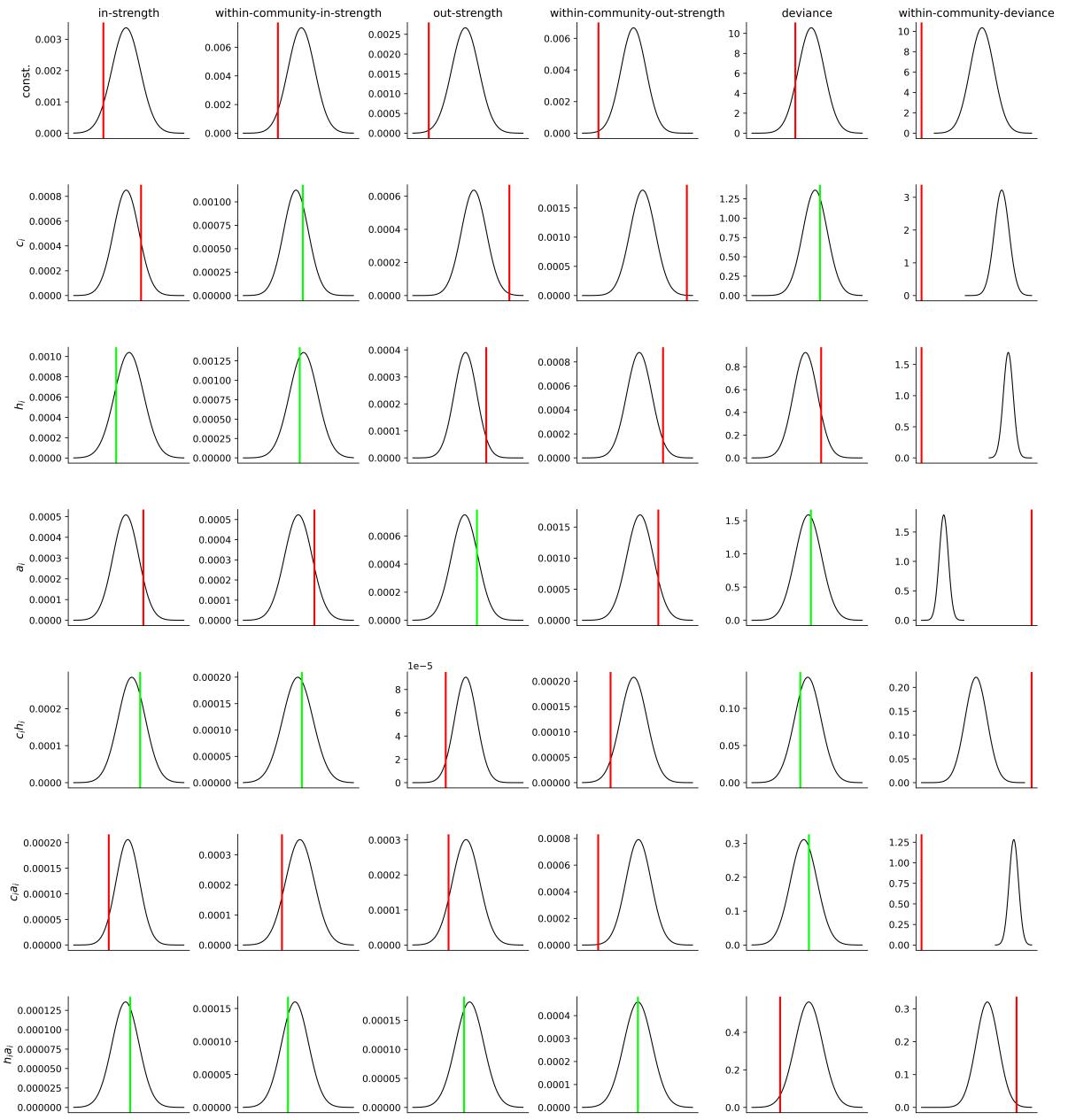


Figure C.1: A visualisation of our variability analysis a node's *individual-level* outcome measures on its update policy parameters. A green vertical line suggests successful replication and a red vertical line suggests unsuccessful replication.

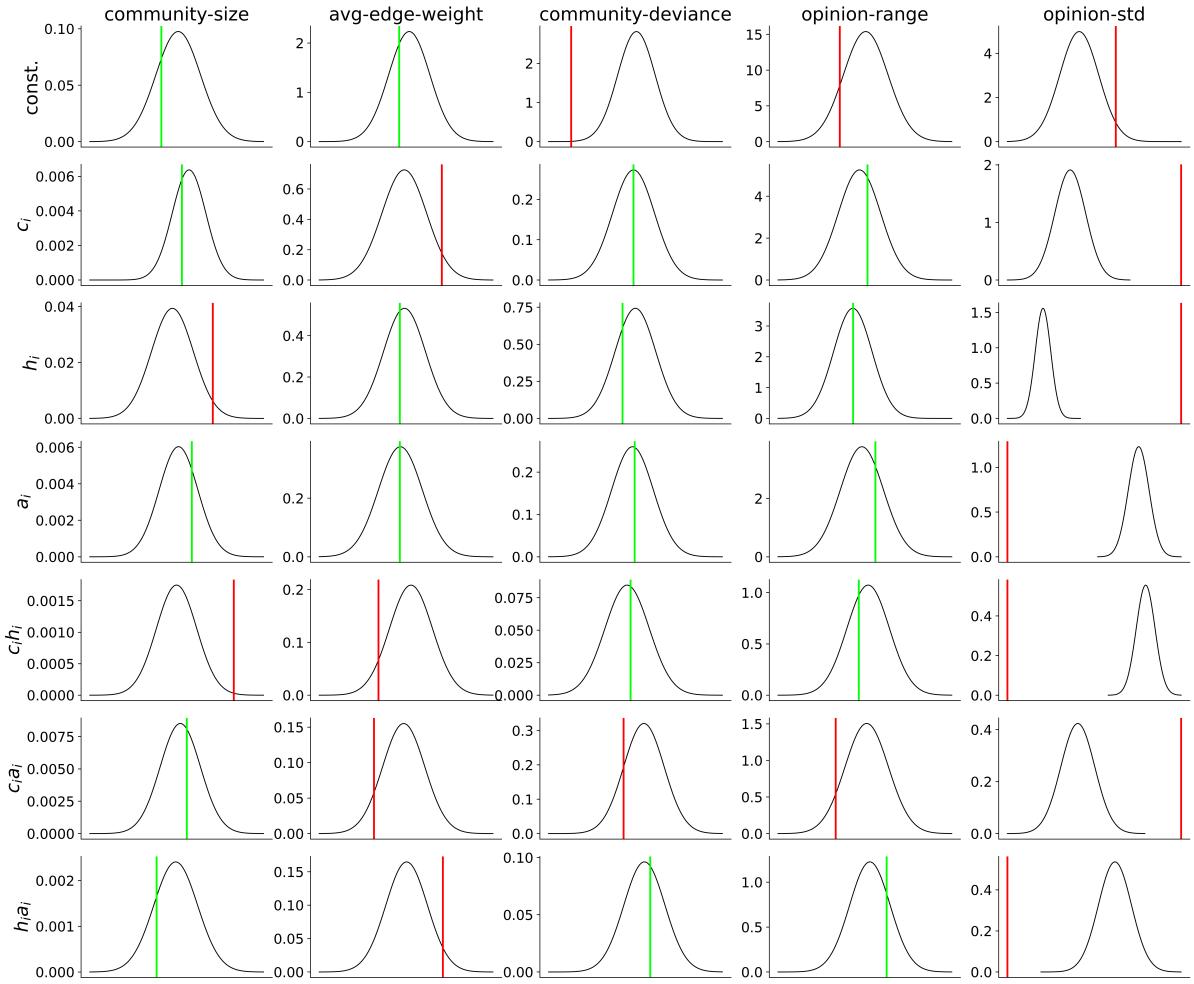


Figure C.2: A visualisation of our variability analysis of a node’s *community-level* outcome measures on its update policy parameters. A green vertical line suggests successful replication and a red vertical line suggests unsuccessful replication.

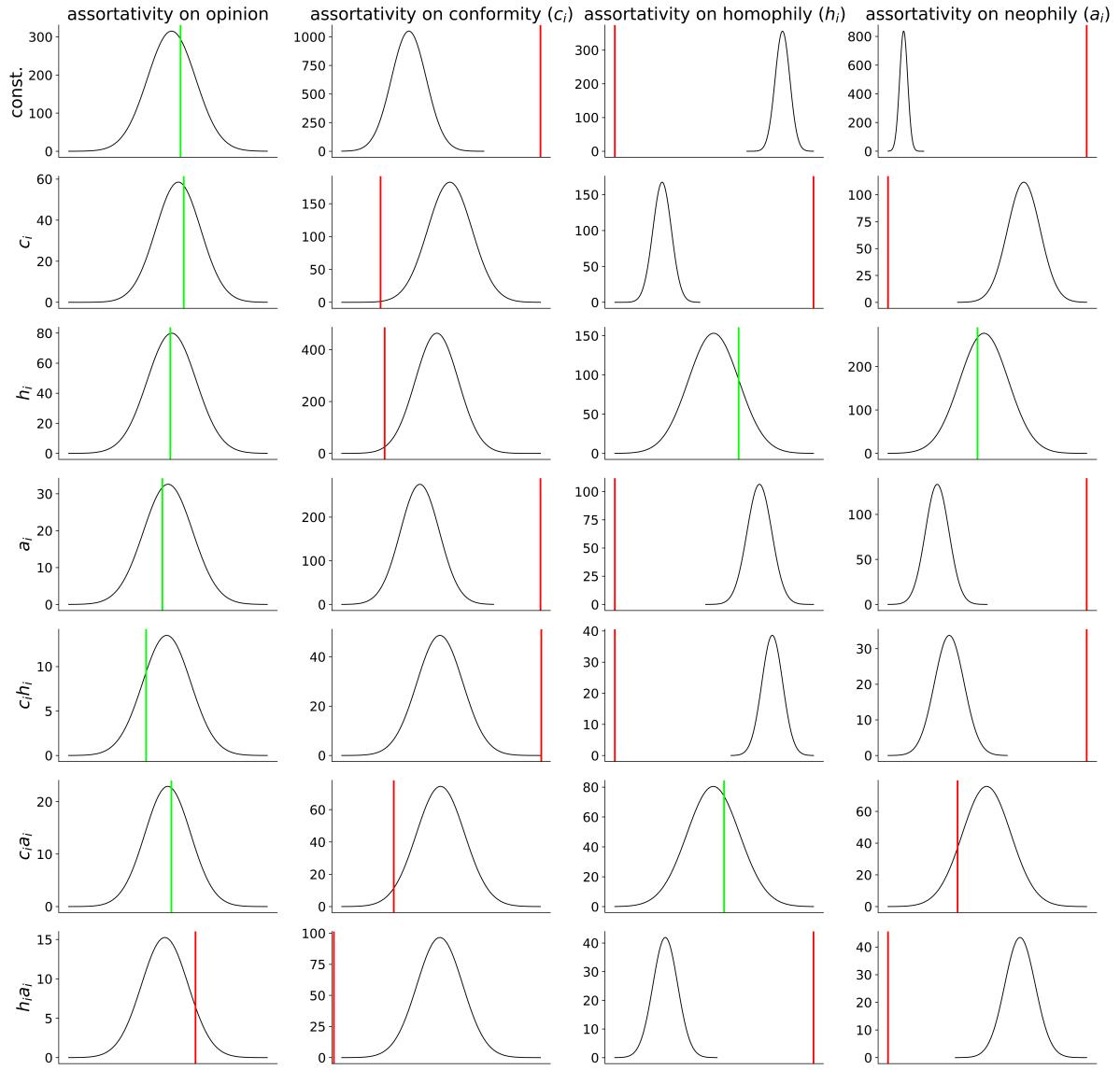


Figure C.3: A visualisation of our variability analysis of a node's *community's assortativity* measures on its update policy parameters. A green vertical line suggests successful replication and a red vertical line suggests unsuccessful replication.

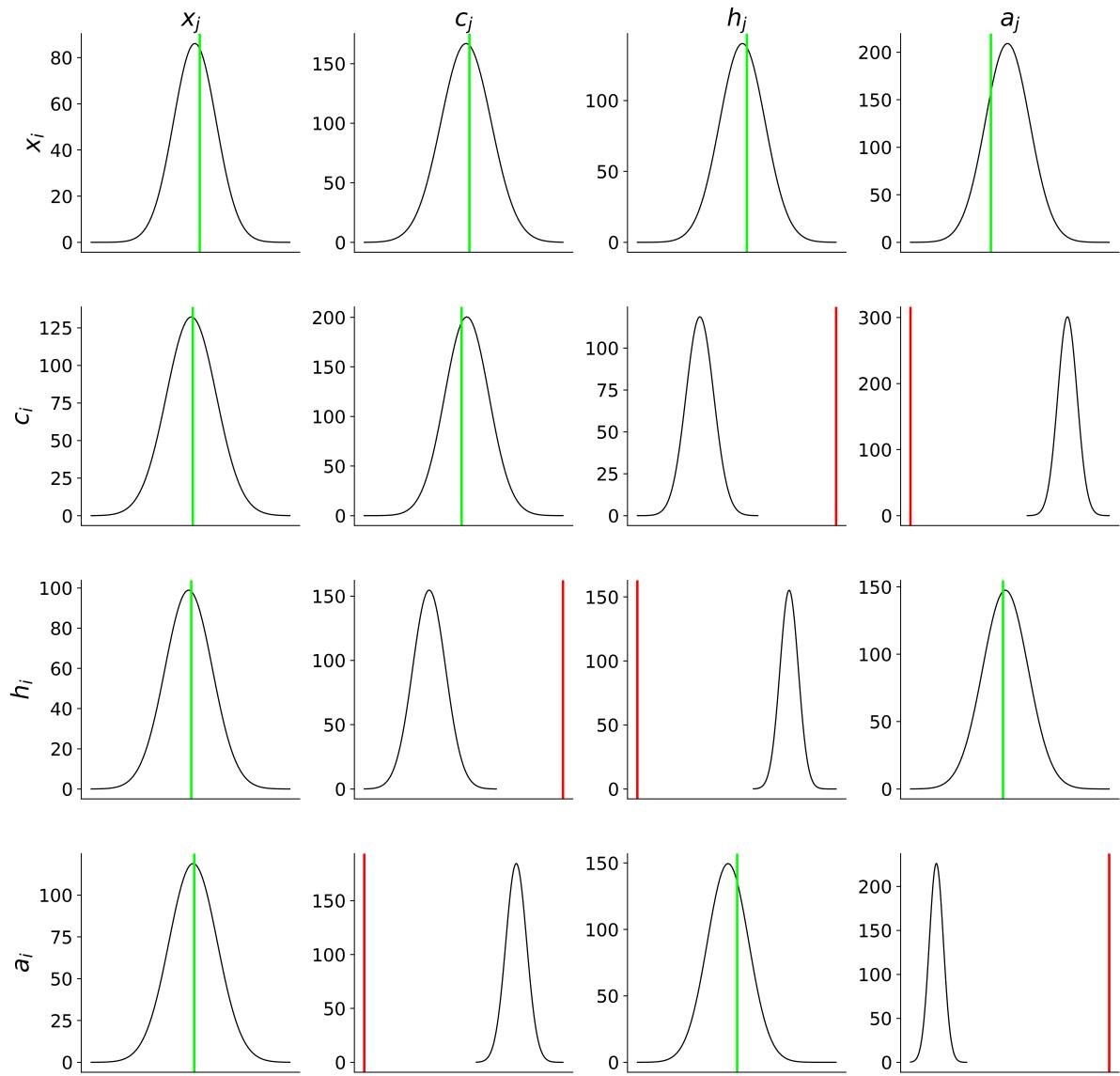


Figure C.4: A visualisation of our variability analysis of network level assortativity of weights w_{ij} . A green vertical line suggests successful replication and a red vertical line suggests unsuccessful replication.