Coexistence of cultural diversity and disassortative structure

Dhamma Kimpara

July 22, 2020

1 Introduction

1.1 Overview

While homophily in social networks has been widely studied and observed, mechanisms to attempt to explain the emergence of homophily are equally important. One particular challenge, but also point of interest, in modeling network formation in this context are the simultaneous dynamics of formation/destruction of social ties and changes of individual states. Out of this conjunction of model attributes arises a complex process that can be thought of as a feedback loop between individual states and network structure. More generally this can be seen as the of the interaction of the micro (individual) and macro (structure) levels of a process. In other words, network structure depends on individual states and individual states depend on network structure. We seek to improve on stylized models that study one process abstracted from the other.

Adaptive networks are the general class of dynamical networks whose topologies and states coevolve (**?**). In the context of cultural dis(integration) processes, we draw upon the adaptive network model in (**?**) which investigates a mechanism that models the emergence of both cultural diversity and high connectivity. Previous models of adaptive social network dynamics did not demonstrate how such an outcome could occur. Either they showed homogenization and well connectedness or cultural diversity and fragmented structure. Though the model in (**?**) has not yet been confirmed by empirical data, a different (much simpler) dynamic adaptive network model in (**?**) has been supported by empirical social network data.

In this project we replicated the theoretical results in (**?**), investigated an extension, and a potential weakness. The rest of the project is organized as follows: section**??** describes the adaptive network model. Section **??** is a brief summary of our results. Section **??** confirms the results in (**?**). Section **??** investigates a weakness in the original results and re-analyzes the original model. Section **??** anaylzes an extension of the model.

2 Methods

2.1 Model

Following (**?**), we will simulate the dynamics of a social network with an initial configuration of two culturally distant groups. The goal is to investigate what population attributes give rise to particular final configurations of the network in terms of cultural diversity and network structure. Particularly, we are interested in the case where there is both high cultural diversity and non-fragmented (highly connected) structure.

In this model, each group consists of 50 individuals (nodes). Directed edges (representing information flow) are initiated at random with an edge density of 0.2 and 0.02, within group and between groups, respectively (directed stochastic block model). Edge weights are initiated by drawing from $U[0, 1]$.

To represent culture, each individual's state or culture is represented by a vector in $\mathbb{R}^{10}$ with cultural distance measured by the euclidean distance. This choice is based on empirical studies on measuring organizational culture. Each culture will have its own "center" cultural vector, C_i. The two centers are initially separated by 3.0 (arbitrary units) in the cultural space. Then for each individual in each group i, their cultural vector is drawn from $C_i + [N(0,0.1), \ldots, N(0,0.1)]^T$.

Now in each iteration of the simulation, each individual will interact with one other individual and update their culture and corresponding edge. Our interpretation was that the sequence of nodes performing actions would be randomized in each iteration. Each node, v, when it is their turn, performs as follows: first, with probability 0.99, v selects u from its in-neighbors with probability proportional to the edge weight. Or, with the remaining probability (0.01), v selects any u from the connected component in which v belongs. Now, v decides whether to accept or reject u's culture. The probability of acceptance, P_A is given by

$$P_A = \frac{1}{2}^{\frac{|c_v - c_u|}{d}}$$

.

Where c_v and c_u are v and u's cultural vectors, and d is the cultural tolerance, or the cultural distance at which $P_A = 0.5$. If u's culture is accepted, c_v is updated as

$$c_v = (1 - r_s)c_v + r_s c_u$$

where r_s is the rate of cultural change, and the edge weight from u to v, w_{uv} is updated as

$$w_{uv} = \text{logistic}(\text{logit}(w_{uv}) + r_w)$$

where r_w is the rate of edge weight change. However, if the u's culture is rejected, v does not update their culture but the edge weight is updated in the opposite direction as

$$w_{uv} = \text{logistic}(\text{logit}(w_{uv}) - r_w)$$

When an edge weight falls under 0.01, the edge is removed from the network.

In all experiments, we use $d = 0.5, r_s = 0.5$, and $r_w = 0.5$ as the mean values. In our extension of (**?**), we include d, r_s, and r_w in culture vector c_v for all individuals (resulting in $c_v \in \mathbb{R}^{13}$).

2.2 Experiments

For each experiment, we drew the behavior parameters of each individual, d, r_s and r_w, from $N(0.5, \sigma$
$)where\sigma$
$correspondstotheparameterbeingsampled$(d, r$_s$ or r_w). The standard deviations of d, r_s and r_w, with values between 0.0 and 0.5 at interval 0.1 (6 values for each standard deviation resulting in 216 unique settings). For each setting of the experimental standard deviations, we performed 100 independent trials. Each trial simulated 500 iterations of the process described in section **??** and analyzed for the following statistics:

Definition 1. Cultural Distance (CD). Let c_v be the culture vector of individual v and A_1, A_2 be the disjoint sets of individuals in each group in directed network G then

$$< CD > := \frac{1}{|A_1||A_2|} \sum_{v \in A_1} \sum_{u \in A_2} |u - v|.$$

Where $|\cdot|$ is the euclidean norm.

Definition 2. Average Shortest Path Length (SPL).

$$< SPL > := \frac{1}{n(n-1)} \sum_{v \neq u} d(u, v)$$

Where n is the number of nodes or individuals in G and $d(u,v)$ is the length of the shortest path from u to v, where all edges have weight/distance 1. By convention (**?**), $d(u,v) = 0$ if v cannot be reached from u.

Here we note an ambiguity in the model: when d, r_s and r_w are drawn, they can be below zero or greater than one. For certain parameters, such as d, one or both of these outcomes makes the parameter nonsensical. We chose to clip the values where appropriate in order for the model to make sense since (**?**) does not mention how to handle these cases. In experiments with high standard deviations (close to σ

$=0.5), thistendstoclusterparametersateither 0.01 or 0.99.$

2.3 Summary of Results

First, we found that we were able to confirm the results, when analyzing the data as described in **??**. However, while doing this, an ambiguity in how to handle resulting graphs that were disconnected presented a potential weakness. We investigated this ambiguity and found that when the analysis of the shortest path length was confined to only either the largest weakly connected component or the largest strongly connected component, the results in terms of SPL diverged from the original results. We then re-analyze the original experiments and perform this same analysis on our extension where we include every individual's behavior parameters (d, r_s, r_w) in their culture vector.

3 Replication of Results

Figure 1 shows plots for $(< CD >, < SPL)$ plots for standard deviations of d, r_s and r_w. We observe that our desired network configuration, high diversity and low SPL is achieved for certain parameter settings. Note that high diversities of cultural tolerance (d) and rate of culture change (r_s) help maintain high cultural cultural distance. In terms of SPL, a high diversity of d helps lower SPL. These results confirm the results we replicated.

3.1 Linear Regression

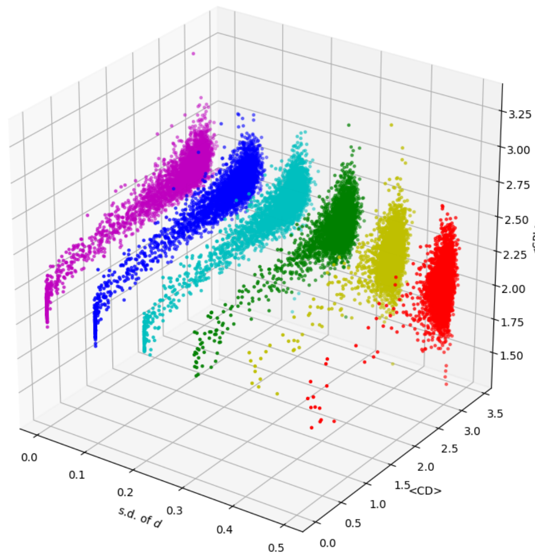
$$< CD > \;\sim\; 2.27 + 1.73\sigma_d + 1.85 - 2.14\sigma_{r_w} - 5.31\sigma_d\sigma_{r_s} + 2.84\sigma_d\sigma_{r_w} + 2.86\sigma_{r_s}\sigma_{r_w} \tag{1}$$

$$< SPL > \;\sim\; 2.47 - 0.66\sigma_d + 0.21 - 0.13\sigma_{r_w} - 0.44\sigma_d\sigma_{r_s} + 0.48\sigma_d\sigma_{r_w} + 0.18\sigma_{r_s}\sigma_{r_w} \tag{2}$$

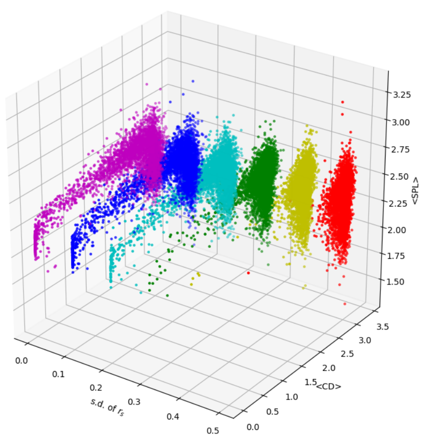
The linear terms in equation (**??**) show that CD is kept high when the diversities of d and r_s are high. CD however is lowered by a high diversity of s_w. This equation corresponds in form an interpretation to the original paper's results however, for us the magnitude of each term is lower. This implies that our experiments yielded less change from an equivalent change in standard deviations.

In equation (**??**), the most significant factor that decreases SPL is the standard deviation of cultural tolerance. In contrast to the equation for CD, our equation corresponds closely to the same equation in the original paper.

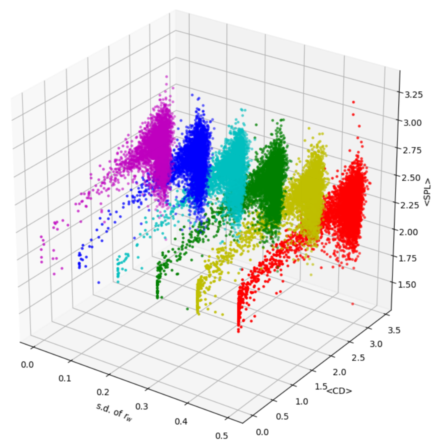
The ANOVA tables in our results also showed that all terms were extremely statistically significant, confirming the results.



(a)



(b)



(c)

Figure 1: 3D scatter plots showing the effect on SPL and CD of each standard deviation of (a) d, (b) r_s and (c)r_w. Each dot is the result of one simulation run, colored according to the standard deviation value.

4 Calculation of Average Shortest Path Lengths

Here we present a potential weakness of the original paper. While investigating the results of the previous section, we noticed that some trials resulted in a disconnected graph. However, these were only a small proportion (2%) of the trials so the disconnected graphs were ignored in the analysis. The same protocol was also followed in the original paper.

We first noticed that a significant number of graphs were disconnected when we ran the same experiment but with our extension where we included the three behavior parameters in each individual's culture vector. This trend was also observed when the group sizes were unbalanced (10 and 40 individuals). We concluded that in order to compare results between differing initial conditions and models, we needed to think carefully about how to handle the the disconnected graphs.

In an organization with required performance goals, it is reasonable to assume that if the organization size becomes too small, it will not be able to perform the functions it is required to do. Hence, it is likely that some external intervention will be initiated. We propose that if the organization size remains at greater than 80% of the original, the organization is still viable. We performed our further analyses with this assumption.

There are however, two ways to determine organizational size in a directed graph. One is to look at the largest strongly connected component (LSCC) and the other is to look at the largest weakly connected component (LWCC). We will see in section **??** that in our experiments, the size of the LSCC is highly correlated with the size of the LWCC.

Definition 3 (SPL_s and SPL_w:)**.** We define SPL_s and SPL_w on a network G as the average shortest path length in G of the largest strongly connected component and the undirected version of the largest weakly connected component, respectively.

We then applied this new analyses, analyzing the largest connected component to the original experiment.

4.1 Comparison Plots

Figure 2 shows the same plot as done in the previous section, on the same experiment data, except now we have added SPL_s and SPL_w. Note that both SPL_s and SPL_w do not decrease as σ_d increases, contrary to the trend of SPL. However, the trend in CD remains similar.

4.2 Component sizes

Figure 3 shows surface pltos of different combinations of the standard deviations and the average LSCC and LWCC (higher surface) of all the trials for each pair of values. We observe that these plots imply that σ_d have a large negative correlation with the LSCC and LWCC. The regression analysis on the average LSCC confirms this observation:

$$|LSCC| \sim 1.04 - 0.40\sigma_d - 0.01 - 0.02\sigma_{r_w} - 0.11\sigma_d\sigma_{r_s} + 9,16\sigma_d\sigma_{r_w} + 0.01\sigma_{r_s}\sigma_{r_w} \tag{3}$$
$$\tag{4}$$

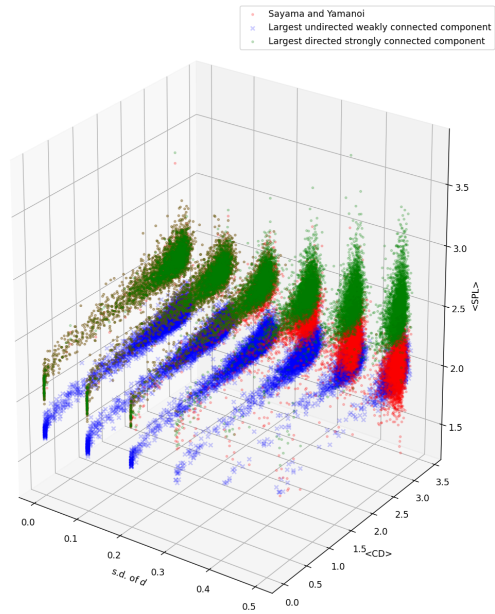
The resulting R-squared was 0.7 and all terms were extremely significant according to the ANOVA analysis except for $\sigma_{r_s}\sigma_{r_w}$. Note that in **??**, we show that the size of the LSCC and LWCC are highly correlated so we focus only on the LSCC.

4.3 Definition of SPL

From the two previous sections, we have seen that SPL_s and SPL_w do not correlate with the trend of SPL when σ_d increases. We have also seen a trend that the networks become significantly less strongly connected as σ_d increases. Here we suspect that a nuance in the definition of SPL maybe the cause of this discrepancy.

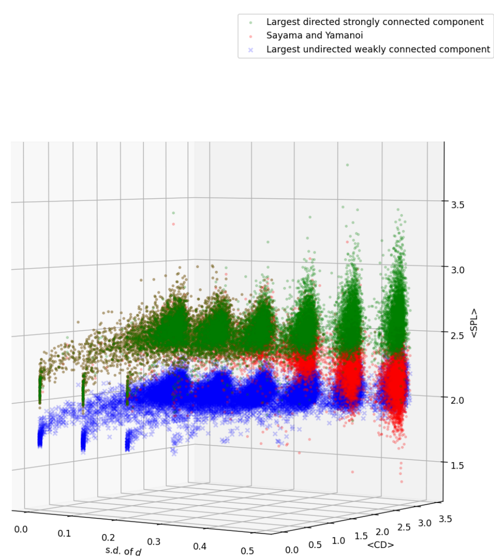
Notice that in the definition of SPL, the distance from node u to v is 0 when v is not reachable from u. This is also reflected in the standard library algorithms for calculating SPL in networkx. The downward trend in SPL can be explained by this fact and the fact that the size of the LSCC decreases as σ_d increases. As the size of the LSCC decreases, while the graph may remain connected, less and less nodes are reachable from each other. Therefore more and more pairs in the summation of the SPL calculation are 0. Hence, we do not think that these experiments necessarily point to increased structural connectivity as σ_d increases. If it did, we would expect to also see a downward trend in SPL_s and SPL_w.

Our position is that in the case of analyzing a connected, but not necessarily strongly connected directed graph, the SPL metric does not make sense. Even if many nodes are not reachable from each other, the SPL will still be low and could even be below 1. There is no way to distinguish between strong connectivity and high fragmentation.



view 1.png

(a)



view 2.png

(b)

Figure 2: 3D scatter plots showing CD, each standard deviation of d, and SPL in red, SPL_s in green, and SPL_w in blue.

4.4 Regression for SC vs WC

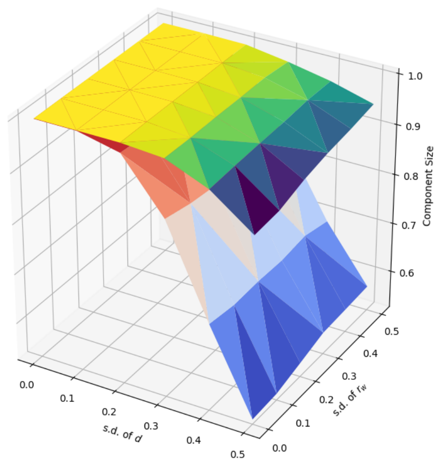
5 Changing all parameters

5.1 Plots

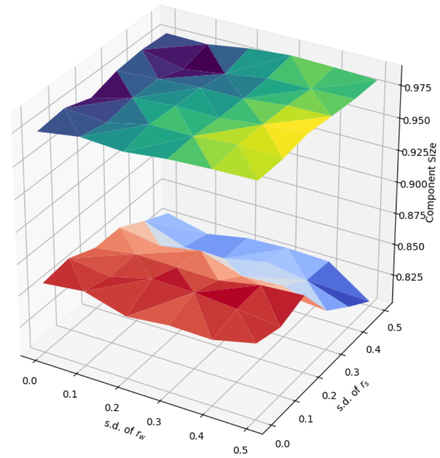
5.2 Regression and Comparison with Base

- rs - compsize.png.png

(a)



(b)



(c)

Figure 3: 3D surface plots showing the interactions between the diversities of combinations of σ_d, and σ_{r_w} on the average size of the LSCC and WSCC (higher surface).

References

Hilla Brot, Lev Muchnik, Jacob Goldenberg, and Yoram Louzoun. Feedback between node and network dynamics can produce real-world network properties. *Physica A: Statistical Mechanics and its Applications*, 391(24):6645–6654, 2012.

Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

Hiroki Sayama and Junichi Yamanoi. Beyond social fragmentation: Coexistence of cultural diversity and structural connectivity is possible with social constituent diversity. In *International Conference on Network Science*, pages 171–181. Springer, 2020.

Hiroki Sayama, Irene Pestov, Jeffrey Schmidt, Benjamin James Bush, Chun Wong, Junichi Yamanoi, and Thilo Gross. Modeling complex systems with adaptive networks. *Computers & Mathematics with Applications*, 65(10):1645–1664, 2013.