# COMPLEXITY MEASUREMENT OF MACROSCOPIC OPINION DYNAMICS

# TO INFER MECHANISMS WITHIN SOCIAL INFLUENCE NETWORKS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Michael J. Garee

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2020

Purdue University

West Lafayette, Indiana

# THE PURDUE UNIVERSITY GRADUATE SCHOOL
# STATEMENT OF DISSERTATION APPROVAL

Dr. Mario Ventresca, Co-Chair

School of Industrial Engineering at Purdue University

Dr. Hong Wan, Co-Chair

Department of Industrial & Systems Engineering

at North Carolina State University

Dr. Steven Landry

School of Industrial Engineering at Purdue University

Dr. Susan M. Sanchez

Department of Operations Research

at the Naval Postgraduate School

**Approved by:**

Dr. Abhijit Deshmukh

Head of the Graduate Program

TABLE OF CONTENTS

LIST OF TABLES

Table         Page

LIST OF FIGURES

# ABSTRACT

Garee, Michael J. Ph.D., Purdue University, May 2020. Complexity Measurement of Macroscopic Opinion Dynamics to Infer Mechanisms within Social Influence Networks. Major Professors: Mario Ventresca and Hong Wan.

Social influence networks are collections of entities dealing with a shared issue on which they have individual opinions. These opinions are dynamic, changing over time due to influence from other entities. Mechanisms within the network can affect how influence leads to opinion change, such as the strength and number of social ties between agents and the decision models used by an individual to process information from its neighbors. In real-world scenarios, these mechanisms are often hidden. Much effort in social network analysis involves proposing models and attempting to replicate target output data with them. Can we instead use the evolution of opinions in a network to infer these mechanisms directly?

This work explores how opinion change in social influence networks can be used to determine characteristics of those networks. Broadly, this is accomplished by simulating social influence networks using various designs and initial conditions to generate opinion data, and then identifying relationships between response variables and changes to the simulation inputs. Key inputs include the population size, the influence model that controls how agents change their opinions, the network structure, the activation regime that controls the sequencing of opinion updates, and probability distributions for communication errors. Analyzing the opinions of individual agents can provide insights about the individuals (microscopic), but in this work, focus is on insights into the social influence network as a complete system (macroscopic), so opinion data is aggregated according to each response variable.

Response variables are designed through the lens of complexity theory. Three types of complexity measurements are applied to opinion data: regression, entropy, and a new complexity measure. In each case, relationships between design factors and response variables are diverse. The influence model and the distribution of communication errors—a factor often omitted from the literature—are consistently impactful, with their various settings producing distinct profiles in time series plots of the measurements. Activation regime is impactful to some entropy measures. Network structure has little impact on the new complexity measure, and population size has little impact in general. Overall, distinctive relationships can exist between opinions and design factors. These relationships, as well as the measures and problem-solving approaches used in this work, may be helpful to analysts working to infer the properties of real-world social influence networks from the opinion data those systems generate.

# 1. INTRODUCTION

What if we could understand the inner workings of our social networks simply by tracking everyone's opinions? Perhaps some form of analysis on how opinions change over time could help us devise a model of how members of a social network make decisions. Analysis of social networks provides benefits to diverse sectors of research, such as epidemiology, political science, economics, psychology, and defense. As a subset of the broader world of social networks, social *influence* networks are more narrowly focused on changes in opinions or beliefs over time. Social influence is "the power [that] relevant others might exercise over an individual through authority, deference, and social conformity pressure" [1]. Models of social influence appear in the literature across many fields of study, including psychology, sociology, political science, statistics, mathematics, physics, and engineering [2]. As an example, in van Maanen and van der Vecht [3], messages on the social media platform Twitter are analyzed for user preferences about candidates in a televised talent show, and the authors simulate various decision models applied to the actual network structure to try to replicate the real-world data and gain insights about how the actual users might make decisions. This is an example of a generative approach to problem solving: manipulate inputs (in this case, the decision model) in order to generate desired outputs (user preferences).

In this dissertation, work proceeds in the reverse direction by adopting an inferential approach: infer the inputs to a process directly, given a set of desired outputs. Following the previous example, this means that given only the set of user preferences for the talent show candidates, we would seek to infer the decision model without repeated manipulation of the possible models. Such an approach is useful in scenarios where we have limited information about or control over a model's inputs, which is common in many real-world scenarios.

The primary tools and techniques used in this work are agent-based simulation, experimental design, social influence network analysis, time-series cluster analysis, and complexity theory. Simulation-based experimentation is used to generate a large body of opinion data from a wide array of different social influence networks. Opinion data is aggregated across the population (directly or through various complexity measures) because the research focus is on insights into macroscopic behavior of networks rather than microscopic behavior of individuals. Results emphasize qualitative relationships between changes in opinion data over time and various experimental factors.

The remainder of this chapter proceeds as follows. I begin with a discussion of my research philosophy, which motivates this dissertation (Section 1.1), before detailing the overall research objective, scope, and relevance of this work (Section 1.2). Section 1.3 contains a brief literature review of the topic; each of the next four chapters contains a more focused literature review. Elements of the methodology for the enclosed articles are outlined in Section 1.4. Finally, the articles themselves are summarized in Section 1.5 to give an overview of their specific research questions and results.

## 1.1   Research philosophy

My research philosophy guides the work that is presented throughout this dissertation. It can be expressed as:

*Insights can be revealed about the inner workings or generative mechanisms of a complex system by analyzing the current system state.*

Complex systems are described in more detail below (Section 1.3), but briefly, they are systems with a relatively high degree of difficulty in describing or creating the system, or they have a difficult "level of organization" [4]. In the philosophy statement above, *inner workings* and *generative mechanisms* refer to specifications, variables, or parameters of a complex system that affect how the system is created or operates,

and the *system state* is the product of such operations. So, changing the inputs may potentially change the outputs. This relates to Kalman's concept of *observability* in control theory [5], which measures a system to be observable if "all state variables can be determined exactly" using a finite amount of output data. My philosophy can be viewed as a claim about a weak form of observability, in that I consider distributions or qualitative insights about system variables, in addition to exact values, to be meaningful insights.

It is evident that in situations where we have full knowledge of the inputs and outputs of a complex system, such as with a computer simulation, we can search for relationships between the two. For example, regression modeling can quantify these relationships, within appropriate confidence intervals. Unfortunately, real-world complex systems rarely come with complete details about their inputs. With sufficient sample data to inform a robust understanding of input-output relationships for a complex system, could we begin to reveal *insights* about the system's inputs by using only its outputs? This is similar to the question asked by statistical inference methods (e.g., Bayesian inference), which work to deduce probability distributions for a population from sample data. In simple cases, the insights I am concerned with could be probability distributions (e.g., if random noise is in a communication channel, what is its underlying distribution?), but in more complicated scenarios, they may take form of a choice between one of several algorithms (e.g., is a network structure generated by a scale-free or random graph approach?) or similar elements.

Regardless, it is my research philosophy that given the outputs of a complex system, it is possible to infer details about the inputs or internal mechanisms of the system. Examining this philosophy in the context of social influence networks and the changing opinions of their members is the underlying theme of this dissertation.

## 1.2   Research objective

This section discusses this dissertation's research objective, the research scope, and the work's relevance to the greater field of social network analysis.

The overall objective of this dissertation is to study the degree to which opinion change in social influence networks can be used to determine characteristics of those networks (such as influence model, network structure, relevant probability distributions). In other words, given social network outputs, infer the inputs. An ideal outcome would be to find a method to precisely identify the attributes of a social influence network, given only the opinion values of the network members. Since social influence networks are complex systems with stochastic elements, this is an unlikely outcome. More realistic—yet still highly desirable—results would be qualitative relationships between measures computed on opinion data and input variables. For example, if we computed some function $f(x)$ on the opinion data from several networks that differ only in the algorithm that generates their network structure, a good result would be to observe different patterns in $f$ that depend on the network structures.

Broadly, this study proceeds by simulating social influence networks using a variety of initial conditions and specifications, computing response variables on agent opinion data, and identifying relationships between the response variables and changes to the initial conditions of the simulation. Simulation is used for two reasons: comprehensive information on real-world influence networks is lacking, and simulation allows fine control of the various model elements. Key model elements from existing literature include population size, influence model (controls how agents change their opinions), network structure model, activation regime (schedules the activity of agents), and the probability distributions for various numerical values in the simulations (e.g., initial agent opinions, communication noise between neighbors, or edge weights). These and other related elements are referred to as *simulation design factors* or *experimental design factors* throughout the text.

Agents within social influence networks can be influenced in two ways, either by their network neighbors or by sources external to the network [6]. In this research, focus is only on influence sources within the network. Agents individually have opinions, but in this dissertation, opinions are studied at the macroscopic level, either by aggregating (e.g., sum or average) the opinions of the population directly or through a complexity measure. Analyzing individual opinions could provide insights about individual agents, but here, the focus is on insights into the behavior of the social network as a whole.

Much effort in social network analysis is focused on selecting the correct model to replicate target data. For example, what network generation algorithm can produce an observed network structure? The work in this dissertation contributes to those efforts by approaching the problem backward: instead of testing different models to produce target data, this research works toward starting with the target data and directly inferring the model. In this way, my research has relevance to existing activity in the problem domain.

## 1.3   Literature review

This literature review explores the current state of knowledge in the fields relevant to the research objective discussed previously. Two primary topics in this section are social influence networks and complexity theory. Background reviews within each article are more tightly focused on that article's work and supplement the remarks below.

A social network is a collection of social entities—people, organizations, nations, and so on—that are connected by social ties, like friendships, business partnerships, or international agreements [7, 8]. In popular culture, social networks are online social sites like Facebook and Twitter, but in the literature, social networks describe not only social media [9], but sets of web sites [10], journal articles [11], and more (e.g., [12, 13]). Social networks, online or otherwise, are a worthy subject of study, as they

are "primary conduits of information, opinions, and behaviors" [14] and are "important in determining how diseases spread, which products we buy, how we vote, as well as whether we become criminals, how much education we obtain, and our likelihood of succeeding professionally" [15]. The study of these networks, broadly termed *social network analysis*, emphasizes measures based on the structure of the network (e.g., node degree, path length, centrality) [15]. Such analysis can give insights about the importance of nodes, connectivity between individuals, likely communities within the network, and mechanisms responsible for network formation. Social network analysis finds use in diverse areas of study, including (but not limited to) epidemiology [16–18], political attitudes [19], economic networks [20], viral marketing [21], immigration projections [12], counter-terrorism [22], and diffusion of information and innovation [23–27].

Social influence networks are a subset of social networks, in which agents exchange information with their neighbors to influence opinions about some shared issue [28]. Social influence network theory "[formalizes] the social process of attitude changes that unfold in a network of interpersonal influence" and "advances a dynamic social cognition mechanism, in which individuals are weighing and combining ... positions on an issue in the revision of their own positions" [2]. An *influence model* defines the interaction and changes of opinions [2]. Perhaps the earliest mathematical model of social influence is that of DeGroot [29], where the social group is working to come to a consensus (such as members of a committee that must work together). Every individual begins with their own subjective distribution $F_1, \ldots, F_n$ of a target parameter $\theta$, which is modified as a result of consulting with other members of the group. The DeGroot model assumes the result of this modification is a linear combination of the others' subjective distributions, such that individual $i$ changes their distribution from $F_i$ to $F_{i1}$ according to

$$F_{i1} = \sum_{j=1}^{n} p_{ij} F_j, \tag{1.1}$$

where $p_{i1}, \ldots, p_{in}$ are weights chosen by individual $i$ prior to interacting with the group, based on some criteria (e.g., perceived expertise). As this process is iterated,

the social group may converge to the same distribution if certain conditions about the network structure are satisfied.

DeGroot's linear updating model reappears in many works on social influence, directly or as the foundation for a new influence model. However, its popularity does not guarantee its real-world utility, since interactions between individual often proceed in highly nonlinear ways, in part because the way an individual responds is often also based on their identity in the group, not just their initial opinion [30]. In other words, realistic social networks are not simply collections of homogeneous agents [23, 31]. The nonhomogeneous nature of social groups makes agent-based modeling and simulation, the approach used in this dissertation, a tool well-suited for social influence network analysis [12, 19, 32–36].

Another domain where agent-based simulation is valuable is complexity analysis. *Complexity* does not have just a single definition or a single technique to measure it [37–39]. Despite the lack of consensus in the field, complexity generally aims to capture the degree of difficulty in describing and creating a system, along with its level of organization [4]. Preeminent complexity measures include Kolmogorov complexity for the minimum description length of a system [40], computational complexity for algorithmic time/space costs [41] (i.e., big-O notation), self-dissimilarity [42], and Shannon entropy [43].

Complex systems analysis is relevant to studies of social influence networks [44]. In fact, complexity theory has seen occasional use in this area, briefly covered in the following examples. Algorithmic complexity can be applied to analysis of social network structures [45], as can graph entropy [46]. Steeg and Galstyan [47] introduce *content transfer* as a new information theory measure to trace the strength of influence between different users' content in an online social network. Many studies analyze static networks, but social networks can grow and evolve, so dynamic network analysis can help bring understanding about the processes that make this evolution occur [35].

As a complexity measure, entropy can help us characterize social influence networks. As one example, Kulisiewicz et al. [48] compute entropy in human interaction

networks based on event sequences, time-ordered lists of events, with each event capturing a single time-stamped pairwise interaction. This means that only the occurrence of communication is measured, not the content or effect of the communication. Their main finding is that the frequency and duration of communications in real-world human networks is not strictly random (e.g., Poisson), and they observe that their entropy measure can be useful in community detection due to the often varied interaction profiles of different subgroups. Chapters 3 and 4 focus heavily on various entropy measures as applied to social influence networks.

## 1.4 Methodology

This dissertation follows an article-based format, with each of the subsequent chapters (except the concluding chapter) containing a published or publication-ready research article. The core of each article is built around one or more response variables calculated using simulated opinion data from a variety of social influence networks. The opinion data is generated via discrete-time (rather than discrete-event) agent-based simulation (e.g., [49]) using custom software written in the Python programming language. Simulating the social influence networks gives direct control of their properties and required inputs, which we manipulate according to experimental design techniques [50] in order to explore a wide array of possible network designs [51], simulation parameters [52], and mechanisms of social influence [53]. Experimental design factors are chosen based primarily on what appears to be factors of interest in the literature (e.g., the influence model), but some factors naturally emerge as variables while designing the agent-based simulation (e.g., population size and the probability distribution for creating edge weights).

Noise in agent interactions is rarely mentioned in the social influence literature, but it captures a fundamental aspect of social influence: communication is a noisy process, so analysis must account for that noise [44, 54, 55]. To do so, we modify each influence model to incorporate random noise (or error) and use the noise distribution's

variance as a design factor. Accounting for the perturbations caused by random noise is one benefit of using agent-based simulation for generating opinion data, versus directly computing results of analytical influence models (e.g., DeGroot [29]).

Much of the analysis in the included research articles is qualitative in nature, using the shapes or profiles of time series plots of the different response variables. Although it may not appear as rigorous as numerical methods, analyzing time series data based on its shape is an accepted approach and can be useful in complexity studies for identifying clusters in the data [44]. One drawback is an added layer of subjectivity present in the results. Given the exploratory nature of the research questions, this drawback is accepted.

Real-world social influence networks are likely much more sophisticated than those that are simulated in this dissertation. This is true of most models in scientific practice, but regardless, the real-world applicability of the research findings may be limited as a consequence. Fortunately, the agent-based simulations in this work are crafted as extensible frameworks to which additional detail and nuance could easily be added, making it straightforward to modify the simulation to answer new or refined questions.

## 1.5   Summary of articles

The articles are sequenced in chronological order from oldest to newest based on when the research was performed. Table 1.1 compares and contrasts several aspects of the articles to demonstrate coverage of the dissertation topic. The rest of this section contains a brief summary of each one.

**Chapter 2**: "Regression-based social influence networks and the linearity of aggregated belief" by Michael J. Garee, Wai Kin Victor Chan, and Hong Wan, 2018. In *Proceedings of the 2018 Winter Simulation Conference.*

Table 1.1.

Comparison table of research articles presented in Chapters 2-5.

| | Chapter 2 | Chapter 3 | Chapter 4 | Chapter 5 |
|---|---|---|---|---|
| Status | *published* | *submitted* | *preparation* | *preparation* |
| Aggregation target | opinion | entropy (multiple measures) | same as Ch. 3 | complexity (custom measure) |
| Aggregation method | sum | average | average | vectorization |
| Response variables | $R^2$ of meta-regression model across replications | average of entropy measure (multiple responses) | same as Ch. 3 | complexity over time |
| Homogeneous agents | yes | yes | no | yes |
| Creates new sim data | yes | yes | no | no |
| Results summary | trends between design factors and regression model quality | trends between design factors and aggregated entropy measures | same as Ch. 3 but adds nonhomogeneous scenarios to design factors | new complexity measure and trends between design factors and complexity |

In this paper, we simulate opinion values (i.e., beliefs) for a collection of social influence networks, aggregate the opinions across the network in several ways, fit linear regression models to that aggregated opinion data, and analyze the relationships between the quality of those models and the simulation input variables used for generating the opinion values. (Our use of the term *belief* instead of *opinion* in Chapter 2 is based on the usage in the prior work upon which that chapter's research was based [36]. We moved to *opinion* in later chapters as a result of literature review, but we view the two terms as synonyms.)

Opinion values for the social influence networks are generated using an agent-based simulation. The inputs to that simulation are varied through experimental design methods to produce a collection of networks in which each member of a network changes its opinion over time due to influence from its neighbors. Specifically, an individual's opinion is altered using a function of their initial opinion, a weighted average of their neighbors' opinions, and a random error term; this influence model has the functional form of a linear regression equation. Next, the opinions are aggregated to produce terms of a regression model for the entire network: the dependent variable is the sum of all agents' opinion and each independent variable is a sum of a subset of agent opinions. The subsets are based on properties of the agent-neighbor relationships. The $R^2$ and $p$-values for the regression model are then used to classify the model as *Linear* ($R^2 \geq 0.50$, $p < 0.10$) or *Not Linear* (otherwise); however, some experimental trials are *Invalid* if the data yields undefined regression values. Finally, we study the effects on this classification that are caused by changing the simulation input variables (design factors). In short, the paper explores how well a system-level regression model fits data from networks where agents exchange beliefs using an individual-level model.

Some of the key findings are as follows. The likelihood of a trial being classified as *Linear* is highest when one agent at a time updates its opinion and lowest when all agents update simultaneously (Figure 1.1). Similarly, a trial is more likely to be *Linear* if the edge weights used in the influence model are sampled from the standard

uniform distribution $U(0, 1)$ compared to either $U(-1, 1)$ or the standard normal distributions, and normalizing the edge weights per-agent further improves linearity. The classification of a trial is computed at every time step during the simulation, and trials are able to change classification during the simulation, which shows that observing the system at single time steps does not reliably predict future behavior. Lastly, normalizing opinion values across the population nearly doubles the rate at which an otherwise *Invalid* trial produces usable data.

**Chapter 3**: "Social influence network simulation design affects behavior of aggregated entropy" by Michael J. Garee, Hong Wan, and Mario Ventresca, 2020. Under review at *The Journal of Mathematical Sociology*.

For the second article, we change our response variables from ones using regression models to ones involving three different entropy measures, and we study the simulation using a designed experiment with five design factors based on prevalent attributes in the literature. As before, our analysis emphasizes the relationships between design factors and response variables for social influence networks.

Three types of entropy are applied to opinion data from social influence network simulations: relative entropy, mutual information, and transfer entropy. Entropy is a popular information measure, and we can view the exchange of opinions between social agents as an information-generating process. For each agent, its opinion values over a period of time are assigned to a set of states, where the states are either fixed-width bins over the range of possible opinion values or symbols based on the relative values of its opinion at consecutive time steps. After mapping opinion values to states, empirical probability distributions are used for computing the entropy measures for each agent. Relative entropy for an agent is the agent's opinion distribution relative to the uniform distribution (where maximum entropy is found), while mutual information and transfer entropy are functions of joint probabilities between agent and neighbor. Each entropy measure for the agents is aggregated to the system-level at each time step using the mean value across the population. These values are then

Figure 1.1. Dot plot of fraction of trials with each linearity classification per level for each factor from Chapter 2. Some observations from this include: the update sequence setting has little effect on linearity, but moderate effect on trial validity; using only one batch is detrimental to linearity; and normalizing $y_i$ leads to more valid—but not Linear—trials. Reprinted with permission.

further aggregated across multiple independent replications (with different initial randomization states) for an experimental trial to produce the response variables, six in total, for each of 1800 trials.

Overall, the relationships between simulation design factors and entropy response variables are varied. For example, the choice of influence model, which determines how opinions are changed over time, is impactful to relative entropy when opinions are binned but not when mapped to symbols (Figure 1.2). Also, both response variables for transfer entropy (binned and symbolized opinions) are highly positively correlated; the same is true for relative entropy, but *not* for mutual information, which has a subset of trials with negative correlations. Population size has negligible effect on all response variables, perhaps because the aggregation approach averages the entropy values across the population. These results, and the many others presented in the article, contribute to the overall effort of this dissertation by evaluating links between the inputs and outputs of social influence network activity.

**Chapter 4**: "Effects of nonhomogeneous agents in social influence networks on aggregated entropy" by Michael J. Garee, Hong Wan, and Mario Ventresca. In preparation.

This article is an extension of the previous one, using the same response variables and most of the same data generation and analysis techniques, but it differs in the composition of the agents in the social networks. Previously, all agents were homogeneous, varying only in their initial conditions and location in the network, but now, agents belong to one class whose members behave differently from agents in the other classes. The behavioral differences vary between scenarios that are based on heterogeneous social network studies in the existing literature, which enhances the real-world applicability of the study: authentic social networks are rarely homogeneous.

Four independent scenarios are created for this study, and each defines two agent classes. First, agents are either informed or uninformed. An uninformed agent initially has no opinion and does not influence others; it will remain in this class until it has an informed neighbor whose opinion it copies, whereby it becomes informed.

median relative entropy, binning (RE-B), all trials, grouped by influence model



Figure 1.2. All trials from Chapter 3 are grouped by influence model, and the groups' median relative entropy is plotted as a function of time.

Second, agents are either agreement-seeking (Concord-type) or partially antagonistic agents. Concord-type agents always move their opinion toward the average opinion of their neighbors, while partially antagonistic agents move their opinion away from their neighbors if their current opinions are too far apart. Third, agents are either bots or humans. Agents in each class preferentially interact with and trust agents in the same class (an example of homophily), and bots have a higher rate of network activity. Fourth, agents are either stubborn or normal. Stubborn agents never deviate from their initial opinion yet are able to influence other agents. For each scenario, a smaller experimental design was used to vary scenario-specific design factors and those common across all scenarios, as applicable. The entropy response variables from the previous article are applied to the data from these new trials, and the analysis focus is on the impact that the scenario designs have on a similarly designed but homogeneous system.

A key observation from this study is that it may be less important to accurately model nonhomogeneous systems if the purpose is to analyze many systems at once instead of only a single system at a time. This is because the response variable distributions across all trials for the nonhomogeneous scenarios and homogeneous base cases were very similar overall, but the variability between an *individual* scenario trial and its corresponding homogeneous system could be great—up to several orders of magnitude—and that difference may vary as a function of scenario design factors (Figure 1.3). Also, we observe a robust relationship between the response variables and experimental design factors, in that the scenario design showed little impact on the entropy measures.

**Chapter 5**: "A complexity measure for opinion dynamics in social influence networks" by Michael J. Garee, Hong Wan, and Mario Ventresca. In preparation.

In this final article, we continue to explore the relationships between opinion change in social influence networks and the variables that lead to such changes, shifting our focus from entropy in particular to complexity theory in general. We first

Percent change from corresponding base case data, MI-S, Scenario 4

Figure 1.3. The mutual information via the symbolic method (MI-S) on *stubborn agent* opinion data (Scenario 4) from Chapter 4 deviates from the corresponding base case trials (no stubborn agents). As the fraction of stubborn agents in the population increases, the percent difference becomes more negative.

establish a concept of what arrangement of opinion data is considered simple in this context, and then construct a measure of subjective complexity that captures the difference between our baseline simplicity and an observed set of opinion data. As before, we study the different patterns in the results relative to changes in experimental design factors.

Creating a measure of complexity requires an idea of what a simple system is for a given situation (the subjective simplicity), a specification or instance of the simple system (the reference simplicity), and a way to measure how an observed system differs from the reference [56]. Our subjective simplicity for studying opinion dynamics is a network of agents who share the same unchanging opinion. We define the reference simplicity to be a set of opinions equal to zero (the neutral opinion for our defined range of opinions) for all agents at all times. The difference, or dissimilarity, measure is a function of the observed opinions' difference from their reference opinions, and the slope, variance, and entropy of the observed opinion time series. The output of this measure is our subjective complexity measure and the response variable used for this study.

This complexity measure generally behaves as desired when applied to a set of simple, synthetic data sets: higher complexity for opinionated agents with diverse or dynamic opinions, and lower for agents with moderate or concurrent opinions. When applied to more complicated opinion data (borrowed from the simulation in Chapter 3), the complexity measure is sensitive to the influence model and presence of noise between neighbors, producing patterns in time series plots of the complexity that could be used to differentiate the different levels for those design factors (Figure 1.4). Varying the choice of network structure model (used for generating the network itself) had little impact on complexity. In summary, the complexity measure created in this article may be helpful to analysts who are working to infer the properties of social influence networks from the opinion data those systems produce.

Complexity time series for case study 2, influence model factor

population: 100
network model: small world(0.33, 10)
error distribution: $N(0, 0.05)$
activation regime: random

attractive-repulsive

standard model

random adoption

time

Figure 1.4. The new complexity measure from Chapter 5 is sensitive to the choice of influence model, when other design factors are equal.

## 1.6  Overview

The rest of this dissertations is organized as follows. Chapters 2–5 contain the research articles summarized in the previous section. The references for each chapter are included in the combined references section at the end of the dissertation. Supplemental information for the articles appears in Appendices A–D. Chapter 6 contains overall discussion and concluding remarks.

# 2. REGRESSION-BASED SOCIAL INFLUENCE NETWORKS AND THE LINEARITY OF AGGREGATED BELIEF

**Article abstract:** Consider an agent-based social influence (belief adoption) network where agents share beliefs with neighbors using a linear regression model. One relevant question is: can aggregated, system-level belief also be fit by a linear regression model? Earlier work demonstrated several scenarios where system-level linearity of belief holds. This chapter extends that research, varying model and simulation factors through experimental design. When linearity does not hold, we isolate the responsible factors. Finally, we investigate whether system-level linearity acts as an absorbing state, that is, when system-level linearity is present at some time $t$, it continues to hold for all later times.

## 2.1 Introduction

Agent-based simulation is a useful tool for building and analyzing social influence networks—systems in which agents exchange information with their neighbors and influence each other's levels of belief over time. In this study, agents update their beliefs via a linear model, based on a weighted sum of their neighbors' beliefs and modified by internal bias and white noise. Similar update functions are used elsewhere in social network analysis [15] and sensor network consensus modeling [57]. We

---

[1]Supplemental information for this article appears in Appendix A.

build system-level measures of influence by aggregating values from each agent. One reasonable hypothesis is that when agents interact in a linear way, the system-level measures may also respond in a linear fashion. Chan [36] finds that for a particular network configuration, this idea is valid. To examine this hypothesis more completely, we build upon the previous work by varying model and simulation elements through experimental design, identifying factors that impact the linearity of system-level belief, and exploring whether individual observations of system-level linear behavior make good predictors of steady-state activity.

In this research, we choose to focus only on linear models and behaviors. Linear systems are widely used in literature and in practice, so we seek to explore their validity as social influence network models. However, we do not make any claims to the importance of linear systems to this topic, nor wish to imply that linear systems are more or less valid than non-linear ones.

We identify ten factors that affect the structure and properties of the network, the schedule used for agent updates, and the settings that control the agent's initial states. We then build a nearly orthogonal Latin hypercube design to systematically study the behavior of the system under different factor settings. The key constant across all trials is the agents' use of a linear model to update their beliefs. We find that linear agent interactions, in most cases, do not generate linear system-level responses. Given the ten experimental factors in our design, three factors can significantly hinder whether we observe linear responses, and five factors have little to no effect for the range of levels used. However, no single factor on its own is observed to absolutely prevent system-level linear responses. Also, we observe that the degree to which aggregated belief can be fit by a linear model can vary over time (i.e., linearity is not an absolute absorbing state in general), so measuring linear behavior for some single time $t$ does not make for a perfect predictor of future performance.

The rest of this paper is organized as follows. We provide a brief review of the literature related to our topic in Section 2.2. In Section 2.3, we develop the model. In Section 2.4, we describe the experimental design for the study. We present our

analysis methods and results in Section 2.5, and we provide conclusions and discuss future work in Section 2.6.

## 2.2 Background

### 2.2.1 Social influence and learning networks

The terms *contagion* and *social influence* are often used interchangeably to describe the process of altering behavior or belief due to communication and comparison among actors in a social system [58]. Social learning is the process of "learning through observation or interaction with other individuals" [59]. In social network analysis, *learning* tends to be used when agents in influence networks seek an optimal behavior or true belief.

Social learning in network analysis is broadly divided into diffusion models and information aggregation models [60]. Diffusion looks at the spread of information through a population; information aggregation focuses on convergence of opinions. The Bass model is a straightforward diffusion model that describes binary adoption of a belief or behavior without using the network structure [15]. DeGroot [29] developed a simple linear updating model to describe information aggregation. There, agents begin with initial estimated beliefs or opinions, and all agents update simultaneously, replacing their current level of belief with a weighted mean of their neighbors' belief levels and their own. Agents will converge to a consensus value if the network structure meets certain conditions of aperiodicity and communication [14, 29].

A second way of dividing social learning is into Bayesian models and DeGroot models. Bayesian models focus on agents learning by observing the actions of neighbors and the payoffs they receive, while DeGroot models learn myopically from communicating and processing only the current system state [61]. The DeGroot model remains a seminal model of information transmission and social learning analysis [15, 20, 60, 62].

### 2.2.2   Regression analysis

Regression analysis is popular in social influence studies. Here, we comment on several recent examples. Mavrodiev et al. [30] studied indirect social influence in a sequential decision making experiment with humans. Participants had access to the mean of all previous decisions, but did not interact directly with other individuals. The authors found a statistically significant fit for a linear regression model relating the amount individuals changed their decision over time and the distance between their previous decision and the current mean. Similarly, Cheng et al. [63] used logistic regression on opinion data from a Taiwanese online bulletin board. Their results showed that users are more likely to post comments that match the sentiment (approval or disapproval) in recent posts, while users are indifferent to the average sentiment of the entire history of comments. Finally, Chan [36] modeled a social influence network where agents interact using linear regression equations. For the particular network configurations used, he found that the aggregated system-level belief could be well-described using a linear regression model. Those findings are a key motivation for the present paper.

## 2.3   Model

### 2.3.1   Networks and agents

We use agent-based simulation to model a social influence network with $N$ agents (nodes). Each agent is connected to one or more other agents using directed edges; self-loops are not permitted. The degree of each agent and the distribution of degree across the network depend on the network structure model family (e.g., scale-free, random, etc.) for a given trial. We use only static network structures for this study, so the set of agents and their edges do not change during a run of the simulation (Figure 2.1). Particular network instances are a function of the structure family,

network parameters, and randomness, and we use an assortment of network instances in our experimental design (Section 2.4).



Figure 2.1. Example of directed Erdős-Rényi random graph with 15 nodes. Arrows point from an agent to an agent's neighbor. This structure model allows for a different number of neighbors for each agent.

Agents are indexed $i = 1, 2, ..., N$. Agent $i$'s neighbors are the agents that receive out-edges from $i$, so neighbor relationships are not reciprocal. Each agent keeps a list of its neighbor indexes, sorted in ascending order. The list of neighbors is indexed by $j = 1, 2, ..., d_i$, where $d_i$ is the out-degree of agent $i$ (Figure 2.2). Agents also track their own current level of belief $y_i$, internal bias $b_{i0}$, and multipliers $b_{ij}$ for neighbor $j$'s belief, which can be thought of as weights for the network edges. As a practical example, $y_i$ may be a person's current opinion of a political topic, $b_{i0}$ is their intrinsic or baseline opinion that cannot be changed by others, and $b_{ij}$ is the weight the person places on the opinions held by their friends. When we need to explicitly compare these values between different time steps, we use superscript $(t)$ to index them by time step $t$, as in $y^{(t)}$.

Figure 2.2. Data structures inside an agent. This example shows an agent with index $i = 12$.

### 2.3.2  Simulation procedure

First, we explain how updates are scheduled from the system-level perspective, then we describe interactions at the agent level, and finally we provide pseudocode for the simulation algorithm. Code for this study is written in Python 3; noteworthy packages include the agent-based simulation framework Mesa [64], the NetworkX package for graph structures and algorithms [65], and Statsmodels for regression analysis [66]. Complete code and other supporting files for this study are available in our online appendix [67].

Our simulation uses a scheduling approach we call batched simultaneous update. We divide the population into batches of equal size. At each time step, batches are updated one at a time in either a fixed or random sequence. When a batch is updated, all agents in the batch update simultaneously. Simultaneous update takes two steps: first, all agents in an updating batch compute their new belief value $y_i$ but store it in a temporary variable, then all agents in the batch update their $y_i$ with the value of their temporary variable. This two-step approach ensures that the update sequence within a batch does not matter, though the sequence in which batches are visited may affect the outcome. Once all agents in the population update, their belief values may be normalized by dividing them by the sum of all $y_i$ values, if the trial settings call for normalization of $y_i$. This batched update mechanism mimics a real-world situation in which people first exchange ideas within a small group of acquaintances and then extend the idea exchange to other groups.

Agents are assigned to batches $1, 2, ..., n_b$ uniformly at random, independent of the network structure. If $n_b = 1$, all agents are in the same batch, so the full population updates simultaneously (as in the original DeGroot model. If $n_b = N$, all agents are in different batches (of size 1), so the full population updates sequentially.

Some trials create *uninformed agents* that have zero initial belief. If an agent is uninformed, it updates its belief value only if it has one or more informed neighbors,

at which time it becomes informed permanently and updates normally. This concept can appear in information diffusion models.

Agents interact only with their neighbors. Agent $i$ computes its level of belief $y_i$ using the linear regression model

$$y_i^{(t)} = b_{i0} + \sum_{j=1}^{d_i} b_{ij} x_{ij}^{(t)} + \varepsilon_i^{(t)} \tag{2.1}$$

where $\varepsilon_i^{(t)}$ is a random error term generated each time step and $x_{ij}^{(t)}$ is the current belief value for agent $i$'s $j$th neighbor (i.e., if the $j$th neighbor of agent $i$ has index $k$, then $x_{ij} = y_k$), and the other terms are as defined previously in Section 2.3.1. The simulation procedure for a single trial is described in Algorithm 1.

### 2.3.3 Response variables

Our focus in this study is to explore linearity in the system-level belief, achieved with a multiple linear regression model of the form

$$Y^{(t)} = B_0^{(t)} + B_1^{(t)} X_1^{(t)} + B_2^{(t)} X_2^{(t)} + ... + B_{d*}^{(t)} X_{d*}^{(t)} + E^{(t)} \tag{2.2}$$

where $Y^{(t)}$ is the system-level belief; $X_j^{(t)}$ represents the aggregated belief of neighbor $j$ across the network, for $j = 1, 2, ..., d^*$; $d^*$ is the maximum out-degree over all agents in the network; $B_0$ is the system's internal bias; $B_j$ is the multiplier for the aggregation of neighbor $j$'s beliefs; and $E^{(t)}$ is the random error term. These values are captured each time step, so they are indexed by time. These interpretations are based on Chan [36]. The aggregated values $Y^{(t)}$ and $X_j^{(t)}$ are the dependent and independent terms, respectively, in our regression model and are our key response variables obtained from the simulation:

$$Y^{(t)} = \sum_{i=1}^{N} y_i^{(t)} \qquad \text{and} \qquad X_j^{(t)} = \sum_{i=1}^{N} x_{ij}^{(t)} \left[ j \leq d_i \right]. \tag{2.3}$$

(Since $d_i$ can vary between agents, $x_{ij}^{(t)}$ may not be defined for all values of $j$ for some agents. The Iverson bracket $[j \leq d_i]$ resolves this.) The set of all $X_j^{(t)}$ terms

---

**Algorithm 1** Simulation procedure for a single trial. First, the network is created based on trial parameters. Then, values for each agent are initialized, including their initial belief values. Finally, the belief update process is performed for each time step, and the process is reinitialized for each replication.

---

1: Trial Initialization:

2:     Construct network from structure model family and parameters

3:

4: **for** each replication of trial **do**

5:     Replication Initialization:

6:         Generate $b_{i0}$, $b_{ij}$, and $\varepsilon_i^{(0)}$ for each agent

7:         Set $y_i^{(0)} = b_{i0} + \varepsilon_i^{(0)}$ as initial level of belief

8:     **for** each time step $t$ **do**

9:         **for** each batch **do**

10:             **for** each agent $i$ in batch **do**

11:                 Generate $\varepsilon_i^{(t)}$

12:                 Set temporary variable for new belief using Equation 2.1

13:             **end for**

14:             **for** each agent $i$ in batch **do**

15:                 Set $y_i^{(t)}$ equal to temporary variable

16:             **end for**

17:         **end for**

18:         Set $Y^{(t)} = \sum_{i=1}^{N} y_i^{(t)}$

19:         If using $y_i$ normalization, set $y_i^{(t)} = y_i^{(t)}/Y^{(t)}$ for each agent $i$

20:         Collect time-step level data

21:     **end for**

22:     Collect system-level data

23: **end for**

---

for a single time step form the vector $\mathbf{X}^{(t)}$. Together, $Y^{(t)}$ and $\mathbf{X}^{(t)}$ make up a single observation of system-level responses for the time step.

We validated our simulation framework with two tests. First, we built a DeGroot model [29] and observed that the aggregated system-level belief converged as expected. Then, we replicated Chan's [36] model and obtained qualitative agreement with his system-level regression results.

## 2.4  Experimental design

In this paper, a *factor* is an input variable that may have an impact on the responses, *levels* are values a factor may be assigned, a *trial* is a combination of levels for each experimental factor (one row from the design matrix, also known as a design point), and a *replication* is one repetition of a trial using different initial randomization settings. We run a replication for 500 time steps and replicate each trial 100 times. The ten experimental factors we use for this study affect network characteristics, update scheduling, and agent interaction. These factors and their associated levels are:

1. Number of agents $N$. 100, 500, or 1000.

2. Network structure instance. For each level of $N$, we create 14 structure instances, discussed below.

3. Number of batches $n_b$. 1, 5, $N/4$, or $N$. For our chosen levels of $N$, we ensure $N \mod n_b = 0$.

4. Update sequence. Batches are updated each time step in either a fixed or random sequence.

5. Distribution of $b_{ij}$ coefficients. Uniform(0, 1), Uniform(-1, 1), or Normal(0, 1).

6. $b_{i0}$ coefficients. Initialize using the distribution for $b_{ij}$ or set to zero to remove internal agent bias.

7. Normalize $b_{i0}$ and $b_{ij}$ coefficients. Within each agent, across the population, or do not normalize.

8. Variance of error terms. Error terms $\varepsilon_i$ are sampled from the normal distribution with mean zero and variance $\sigma^2$ of 0.5, 1.0, or 2.0.

9. Normalize $y_i$ each time step. Yes or No.

10. Fraction of uninformed agents. 0, 0.05, or 0.25. This fraction of agents have $y_i$ set to zero and are flagged as *uninformed* at the start of the run.

Each factor affects one or more high-level features of the simulation that may influence the linearity of system-level responses. Factors 1 and 2 affect the size and shape of the network, while Factors 3 and 4 control agent update scheduling. Factors 5 to 8 govern the regression terms for the update equation (Equation 2.1) within each agent. Factor 9 alters the scale of belief values, and Factor 10 lets us see how diffusion of initial belief affects the results.

A network structure instance is a particular realization of a network model for a given structure family, input parameters for that family (including network size $N$), and a randomization seed, when applicable. For a given family and set of parameters, changes to randomness can yield different network shapes with different metrics. The network structure families we use are scale-free, directed random tree, Erdős-Rényi random graph, and random $k$-out. Two to four instances from each family are made using several sets of input parameters and randomization seeds. We select inputs that produce a satisfactory range of values for the standard network measures of mean out-degree, assortativity, reciprocity, and efficiency.

A full factorial crossed design for this experiment is written as $2^3 \times 3^5 \times 4^1 \times 14^1$ and requires a costly 108K trials. Instead, we adopt a data farming view and use a nearly orthogonal Latin Hypercube (NOLH) design; this choice is motivated by Sanchez and Wan [50]. The NOLH design tool we use [68] reduces our experiment to 255 trials. The final experimental design matrix uses all ten factors and their associated levels defined earlier and is available with our online appendix [67].

## 2.5 Analysis and results

### 2.5.1 Data processing

The first stage of analysis is to convert simulation outputs into a suitable analysis database. Figure 2.3 gives a schematic overview of our data processing activity. To process a single trial, we group the simulation outputs $\left(Y^{(t)}, \mathbf{X}^{(t)}\right)$ by time step from across all replications of that trial. Within each time step, data cleaning removes outliers (e.g., due to floating point error or very extreme values) based on the standard 1.5-IQR (interquartile range) rule and drops observations with null $Y$ values (e.g., caused by $Y$ growing to infinity). Using the remaining observations, we fit a main-effects only linear regression model of the form described in Section 2.3.3, record the model's adjusted $R^2$ $\left(R^2_{adj}\right)$, and assess $R^2_{adj}$ as *Significant* if the model's $p$-value is below 0.10 or as *Not Significant* otherwise. This is done for each time step of the trial.

$R^2_{adj}$ and the significance rating can move unpredictably between time steps due to randomness, so we apply to both parameters a smoothing function and the prefix "MA" for *moving average*. For $R^2_{adj}$, we use a five-time step simple moving average on $R^2_{adj}$ (the arithmetic mean of the $R^2_{adj}$ values for the last five time steps) and call the result MA-$R^2_{adj}$. For the smoothed significance rating, which we call MA-Sig., we assess the rating at each time step as the rating of the most recent sequence of length three or greater: Over time, as we observe three Significant time steps in a row (based on $R^2_{adj}$), we begin assigning MA-Sig. as Significant, until we observe three Not Significant time steps in a row and switch to assigning MA-Sig. as Not Significant, and so on, switching back and forth as needed.

$R^2_{adj}$ and MA-Sig. drive a function that classifies each time step as *Linear*, *Not Linear*, or *Invalid*. The Invalid classification is applied if either MA-$R^2_{adj}$ or MA-Sig. is undefined, which seems to occur when $Y$ or elements of $\mathbf{X}$ grow too large or have infinite variance, or when the residual degrees of freedom for the model is too low. (The maximum degree of the network structure governs the number of elements

Figure 2.3. Schematic for data processing and analysis. Each independent replication of a trial yields one observation of the system $\left(Y^{(t)}, \mathbf{X}^{(t)}\right)$ for every time step $t$. For each time step, the set of observations are cleaned and put into a linear regression model of the form defined in Section 2.3.3 to produce an $R^2_{adj}$ value and Significance rating based on the $p$-value. Due to randomness, these values can move erratically between time steps, so we apply a smoothing function and the prefix "MA" for moving average. The smoothed data classifies each time step as Linear, Not Linear, or Invalid, ending data processing and creating the Analysis Database. This processed data is sliced per trial for time series studies, or per time step to support regression analysis on our experimental design and finding classification rates for each factor-level in the design matrix.

in $\mathbf{X}$ and number of factors in the regression model, so trials with large networks and highly-connected agents require more replications to potentially be valid.) We also classify cases where MA-$R^2_{adj}$ is exactly equal to unity as Invalid; rather than indicating a perfect fit for the regression model on the simulation output, this occurs for some trials with very extreme $Y$ values, many outliers, or otherwise problematic data. A valid time step is Linear if MA-$R^2_{adj} \geq 0.50$ and MA-Sig. is Significant. Otherwise, it is classified as Not Linear. (Note: Linear and Not Linear are defined only with respect to our goal of fitting time step data with a model having the same functional form as the update equation used by the agents. We acknowledge that there may exist linear models that strongly fit data that we classify as Not Linear, but they must contain explanatory factors outside the current scope.) The values for MA-$R^2_{adj}$, MA-Sig., and Classification for each time step, and for each trial, populate the analysis database, which we use in concert with the experimental design matrix for all subsequent analysis.

### 2.5.2 Analysis

Our first two analysis products, factor-level classification and experimental design regression, consider the system at a single time step ($t = 500$), while the third is a time series study. First, for each factor-level in the experimental design, we filter the analysis database by trials containing that factor-level and calculate the fraction of trials with each of the three classifications (Linear, Not Linear, and Invalid). This provides an intuitive way of making qualitative assessments of how the factors affect system-level linearity. A dot plot of this data (Figure 2.4) lets us make several observations:

- Updating the population in a single batch is significantly worse for linearity; $N$ batches are best.

- Randomizing the update sequence of batches has little effect on linearity.

- Generating $b_{ij}$ values from the Uniform(0, 1) distribution produces linear results significantly more often than when we use Uniform(-1, 1) or Normal(0, 1), which both have mean of zero.

- Normalizing all $b_{i0}$ and $b_{ij}$ values per agent yields the most linear trials.

- Normalizing $y_i$ may be beneficial because it reduces the chance of a trial being classified as invalid.

We do not include the network structure in this part of the analysis, because each structure instance is a direct function of $N$ and also has a very low sample size (10-20).

Next, we build regression models at $t = 500$ on the experimental design using MA-$R^2_{adj}$ as the dependent variable and omitting any Invalid data points. (The $R^2_{adj}$ of these models is only indirectly related to the $R^2_{adj}$ from the raw simulation output that underpins MA-$R^2_{adj}$.) The initial main effects model with the ten original design factors performs adequately ($R^2_{adj} = 0.513$). We slightly improve on this baseline model by proxying the network structure factor, a categorical variable, with the maximum degree $d_i$ of the network, a discrete variable. Surprisingly, if we replace the categorical levels 1, 5, $N/4$, and $N$ for number of batches with the discrete number of batches $n_b$, we find a significantly worse model fit. The model we select as best contains only five factors: the categorical batch quantity level, the distribution function for $b_{ij}$, whether we normalize $y_i$, the way we normalize $b_{i0}$ and $b_{ij}$, and the maximum degree $d_i$ of the network (a proxy factor for network structure instance). This model's $R^2_{adj}$ is 0.532 and is statistically significant. All factors are significant and have low variance inflation factors. Table 2.1 summarizes the model evolution process, and Figure 2.5 contains diagnostic plots of the best regression model. Full model results are available in the online appendix, but we provide a brief interpretation here:

- Increasing the number of batches causes MA-$R^2_{adj}$ to monotonically increase, but the rate of increase falls off rapidly above $n_b = 5$.

- Selecting Uniform(0, 1) for the $b_{i0}$ and $b_{ij}$ distribution is associated with higher values of MA-$R^2_{adj}$.

Figure 2.4. Dot plot of fraction of trials with each classification per level for each factor at $t = 500$. Some observations from this include: the update sequence setting has little effect on linearity, but moderate effect on trial validity; using only one batch is detrimental to linearity; and normalizing $y_i$ leads to more valid—but not Linear—trials. The network structure factor is omitted from this plot.

- Normalizing $b_{i0}$ and $b_{ij}$ values per agent, instead of population-wide or not at all, is linked to higher MA-$R^2_{adj}$.

- Normalizing $y_i$ and the maximum network degree are statistically significant but have very minor coefficients. Removing these factors from the model lowers $R^2_{adj}$ to 0.507 from 0.532, so we instead retain them.

These observations agree strongly with our earlier comments on factor-level classification percentages.



Figure 2.5. Diagnostic plots of main effects-only model of experimental design regressed on MA-$R^2_{adj}$. This model uses five of the ten experimental factors and achieves $R^2_{adj} = 0.532$.

We finish our regression analysis by constructing a model with main effects and two-way interaction terms. One of the highest-quality interaction models we found uses a slightly different set of design factors, created from the main effects model by removing maximum degree $d_i$ and adding whether the update sequence is randomized, the variance of the error distribution, and whether $b_{i0}$ is set equal to zero. Backward elimination is used to selectively remove low $p$-value interaction terms, yielding a statistically significant model with $R^2_{adj} = 0.754$. Diagnostic plots of this model are similar to Figure 2.5 but with tighter distributions.

Table 2.1.

Summary of regression models fit to experimental design with dependent variable MA-$R^2_{adj}$ at time step $t = 500$. All results in the table are statistically significant ($p$-value $< 0.001$).

| Regression Model on Experimental Design ($t = 500$) | $R^2_{adj}$ |
|---|---|
| *Main Effects Only* | |
| Baseline: All 10 original factors | 0.513 |
| Proxy network structure (categorical) with max $d_i$ (discrete) | 0.533 |
| Proxy batch quantity level (categorical) with batch quantity (discrete) | 0.388 |
| Best fit: baseline, omit $N$ and randomize update sequence, use max $d_i$ (8 factors) | 0.538 |
| Best model: batch quantity level, $b_{ij}$ dist., normalize $b_{i0}$ & $b_{ij}$, normalize $y_i$, max $d_i$ (5 factors) | 0.532 |
| *Main Effects and Two-Way Interactions* | |
| Baseline: All 10 original factors & all interactions—input rank exceeds observation count | n/a |
| All 10 factors, proxy network structure (categorical) with max $d_i$ (discrete) | 0.739 |
| Best fit with all interactions: baseline, omit $N$ and uninformed rate, proxy network structure | 0.761 |
| Best model: 7 factors, proxy network structure, backward eliminate interaction terms | 0.793 |

Forward or step-wise regression is a more typical approach when using large-scale design of experiments.[2] The design used in this research was not created explicitly for simultaneous estimation of all second-order effects and not fully support backwards elimination.

Finally, we consider how trial classification (Linear, Not Linear, or Invalid) changes over time, and whether linearity continues in the future once it appears. Some trials have identical classification patterns and can be merged, reducing the data from 255 trials to 94 patterns. This data reveals that the classifications of a small number of time steps do not help predict long-run activity (Figure 2.6). A single Linear time step can be part of a trial where the future is always Linear, where the system moves between classifications frequently, or as a random Linear time step as part of a mostly Not Linear trial.

## 2.6   Conclusions and future work

In this paper, we use experimental design to build a collection of agent-based social influence networks populated by agents that exchange belief via a linear regression model. We then investigate the behavior of the aggregated, system-level belief using state transition probabilities, regression analysis, and a custom classifier of trial linearity. The hypothesis is that linear agent-level interactions would lead to linear system-level responses, but we identify several model features that challenge this assumption.

Features that negatively affect system-level linearity include updating the full agent population simultaneously as a single batch, generating agent's $b_{ij}$ coefficients from a zero-mean probability distribution, and choosing not to normalize $b_{i0}$ and $b_{ij}$ values per agent. In real-world influence networks, people generally exchange ideas within small groups of acquaintances over time, so multiple update batches are more similar to reality. Also, the trust that real people have in their acquaintances (i.e. $b_{ij}$) is unlikely to be zero, so a zero-mean distribution may be inappropriate—if people

---

[2]This paragraph was not present in the original publication of this article. Thanks to Susan M. Sanchez for this correction.

Figure 2.6. Trials are classified Linear, Not Linear, or Invalid at each time step based on MA-$R^2_{adj}$ and MA-Sig., and trials with identical classification patterns are merged. This figure shows the 94 distinct classification patterns (rows), sorted left to right by time step (columns). The solid green bar at the bottom of the figure represents trials that are Linear for every time step (40 trials). Single time step observations do not predict future performance.

trust their neighbors, the whole system may be more well-behaved. Therefore, our experimental results suggest that actual influence networks may tend more toward linearity.

We observe no factor-level that completely prevents linearity on its own. Half of our experimental factors have no real effect on classification for the range of levels we selected, namely the population size, randomizing the batch update sequence, fixing (or not) agents' internal bias $b_{i0}$ to zero, error variance, and the initial uninformed rate of agents. Testing these factors over a greater range of levels could justify omitting them from later experiments. Normalizing $y_i$ each time step is valuable not for affecting linearity but rather for reducing trial invalidation. These findings may help analysts design more effective simulations in the future. Lastly, if we wish to predict long-run behavior with respect to classifying a system as linear, evaluating single time steps is inadequate. Further study is required to understand how long-run behavior relates to the experimental design factors.

Many rich research areas remain that can build upon this work through small changes to the existing model, as the structure of our simulation allows us to easily test more intricate scenarios. We assess system-level linearity based strictly on regression models of the same functional form as the agent's interaction equation, but relaxing this definition and bringing new model features into the analysis may reveal new insights. Allowing self-loops in the network structure would see agents factoring their current belief value into the update process; applying an autoregressive-type model may be interesting. It could be worthwhile to explore the effect that linearity has on model behavior metrics such as speed of belief adoption. Instead of updating the full population each time step, agents could update stochastically or subject to conditions about their neighbors (e.g. homophily/thresholding, where agents ignore opinions too different from their own). A more substantial change to our simulation would be to have dynamic network structures, creating and destroying links over time. Finally, additional network metrics in the experimental design regression analysis could shed light on the influence of network structure on system-level linearity of belief.

# 3. SOCIAL INFLUENCE NETWORK SIMULATION DESIGN
# AFFECTS BEHAVIOR OF AGGREGATED ENTROPY

**Article abstract:** As agents interact and influence one another in a social network, the opinions they hold about some common topic can change over time. These changes may enable us to infer mechanisms of the network that control how interactions lead to opinion change. Inferring such mechanisms from opinion data could enable analysis of social influence in data-sparse scenarios. However, limited work has focused on this problem, despite its clear value. In order to address this gap, we create opinion data using agent-based simulation and experimental design. By viewing opinion changes as an information-generating process, opinion dynamics can be studied using entropy. This work explores the relationships between aggregated entropy and five simulation design factors. Three entropy measures are calculated on continuous-valued opinions and are analyzed using a main effects model and cluster analysis. Overall, the choices of influence model and error distribution are most important to the entropy measures, activation regime is important to some measures, and population size is unimportant. Also, design variation can be detected using time series cluster analysis. These findings may support work in inferring properties about real-world social influence networks using opinion data collected from their members.

---

[1]Supplemental information for this article appears in Appendices B and C.

## 3.1 Introduction

As agents interact and influence one another in a social network, the opinions they hold about some common topic can change over time. To understand the relationships between opinion changes and the underlying mechanisms of the network (e.g., influence process, interaction frequency), we can simulate social influence networks and observe the effects that varying simulation inputs can have on agent opinions. In this paper, we adopt the reverse perspective: given a collection of agent opinion data, make inferences about the underlying mechanisms of the influence network. This approach could be useful in scenarios where there is insufficient data about the network to construct a credible simulation, yet ample data is available on how agent opinions evolve.

Social influence networks are "social cognition structures assembled by individuals who are dealing with a common issue" [69]. Its members, herein called *agents*, have opinions about the "common issue" and the ability to modify those opinions based on influence from other agents. The *model* for a social influence network defines how an agent's opinion changes due to interaction with its neighbors or other information sources [70]. Following Flache et al. [53], *opinion* is the "agent's property that is affected by social influence in a model" and generically represents social terms such as belief, behavior, and attitude. In this work, we define opinion on a continuous scale (e.g., relative degree of support/opposition for an issue), which is consistent with much of the existing work in opinion dynamics (e.g., [2, 14, 29, 53]).

The evolution of opinions among agents in a social network can be viewed as an information-generating process. Entropy is a popular concept for information measurement in general, and it has been successfully used to study opinion dynamics for individual agents in particular. For instance, transfer entropy can be used to infer causal relationships between peers in a social network, on the basis that influence is detectable by changes in information distribution [31], and relative entropy can characterize the rate of learning in a population based on injecting information at

different locations in the network structure [71]; for other examples, see Xie et al. [72] or Zhao et al. [73].

Here, we wish to extend the scope of analysis for social influence and entropy of the individual to the entire system, defining *system* as the union of the agents, their properties, their network, and the methods by which agents interact (i.e., the system is everything required for creating opinion change). The entropy measures we have selected are computed on individual agents or pairs of agents, but we wish to produce a combined measure for the network as a whole. So, we take entropy measured for individual agents and aggregate it across the population to look for insights about the system. (In this work, we aggregate by averaging, but other methods can be considered.) For example, can we use entropy measurements on population opinion data to infer the influence model using data from previously-studied networks? We focus on three entropy measures from the literature: relative entropy [71], mutual information [74], and transfer entropy [47], selected for their relevance to information movement within networks and utility in the social network analysis performed in the cited works. We calculate entropy on social influence networks using the time series of opinion values held by the agents in the network: opinions vary over time, and this leads to a distribution of opinion values.

To illustrate this approach, consider one agent in a social influence network. The relative entropy for the agent is a function of the distribution of its opinion values over some time period, relative to a uniform distribution across the maximum range of opinions. If the agent's opinion changed at random, relative entropy would be near zero, while if its opinion was unchanging, the relative entropy would be maximized. When aggregated across the population, the order (or disorder) of the relative entropy may reveal details about the system that we can tie to its properties, discussed later.

Similarly, mutual information and transfer entropy are functions of the joint probability distributions for the opinions of a pair of agents. Mutual information is zero if there is no apparent relationship between the joint and marginal distributions of their opinions, and it increases as a relationship emerges. Transfer entropy is similar but

accounts for changes over time in greater detail (Section 3.2). There are many pairs of agents in a network, but we assign an agent a single value for mutual information (transfer entropy), taken as the mean of the mutual information (transfer entropy) between that agent and each of its neighbors. We use the phrase *aggregated entropy* to refer to the result of aggregating the entropy values for individual agents across the social influence network. The inputs to this aggregation depend on the choices of entropy measure and discretization method, but the aggregation process is the same for each of the response variables used in this study.

Broadly, the objective of this research is to study the relationships between the characteristics of a social influence network and the entropy measures computed on network using agent opinion data as it changes over time. We use agent-based simulation to generate opinion data, so these network characteristics take the form of simulation input variables, or design factors, such as the size and shape of the network and the influence model controlling interactions between agents. The design of a social network simulation should affect the observed opinion dynamics. More specifically, our objective is to study the relationships between the *system design factors* (input) of a social influence network simulation and the *aggregated entropy measures* on the simulated opinion data (output).

In this study, we address the relationship between system design and aggregated entropy through the following research questions:

1. Which system design factors contribute most to aggregated entropy? (This can reveal which factors are good candidates for inference.)

2. How is system design related to the distribution and scale of entropy time series data? (Clusters in the outputs may be linked to commonalities in the inputs.)

3. How do different entropy measures respond to changes in system design? (Some measures may be more sensitive, and useful for inference, than others.)

For the first question, we focus on the one-way sensitivity of aggregated entropy to changes in the system design factors by identifying patterns in response variable plots

and performing statistical comparison tests between the experimental levels of design factors. We address the second question through cluster analysis on time series data from each trial and inspect cluster composition with respect to the design factors. The third research question relies on qualitative analysis of spatial patterns in the time series plots of the entropy response variables and correlations within trials for different response variables.

**Summary of main contributions**

The results from this study identify relationships, or lack thereof, between the design factors for our social influence network simulations and the aggregated entropy measures applied to opinion data created by the simulation. For one, relative entropy can be sensitive to the influence model and the distribution of communication errors, in that different choices for those factors produce different patterns in plots of relative entropy, but it is insensitive to population size, network structure, and timing of agent interactions. The strength of such relationships with the design factors vary for the different entropy measures and even the different discretization methods (e.g., influence model is more impactful to relative entropy on binned data than on symbolized data). These findings may support work in inferring properties about real-world social influence networks using opinion data collected from their members.

Also, the results add support to the idea that the presence of noise in communication between agents and the timing of agent interactions can affect the evolution of opinions. Those design factors are challenging to capture in purely analytical models and are often omitted from social network analysis in the literature, but they are straightforward to build into an agent-based simulation, like the one created for this study.

The remainder of this paper is organized as follows. In Section 3.2, we review entropy metrics in the context of social networks. In Section 3.3, we explore design factors of social influence networks that allow simulating a diverse set of system

designs. Section 3.4 details the methods for designing the simulations and generating data. We present the results of our different analysis techniques in Section 3.5. Findings and conclusions are discussed in Section 3.6.

## 3.2 Entropy in social influence networks

Social influence networks are systems in which agents exchange information with their neighbors to influence opinion values over time [28]. Agent activity in a social influence network can be viewed as a kind of information-generating process, so entropy can characterize the information and uncertainty associated with the exchange of opinions. Entropy measures in social network analysis can be predominantly divided into activity-based measures and structural measures. Activity-based measures calculate entropy about message exchange between agents [47, 75], agent decisions, or output signal distributions [71]. Structural measures, such as connectivity entropy and centrality entropy [76], rely on properties of the underlying graph, like node degree and path length. Since our focus is on changes in agent opinions caused by interactions with their neighbors, rather than structural properties of the social networks, activity-based measures are used herein.

Social network activity generates information through the exchange of messages among agents, so Shannon entropy is relevant [74]. Indeed, the fundamental ideas of Shannon entropy appear in the measures featured in this section. In this work, an agent's opinion values are discretized using a fixed set of states (via methods in Section 3.2.1), and we use the empirical probability distribution associated with those states to compute the following entropy measures.

*Relative entropy* (or Kullback-Leibler divergence) measures an error between two distributions, usually between an observed and an assumed distribution [77]. Relative entropy for a process $X$, representing a discrete random variable (such as discretized opinions of an agent in a social influence network), is defined as

$$D_X(p \parallel q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)}, \tag{3.1}$$

where $p$ and $q$ are discrete probability distributions with identical support and $x$ represents possible values of $X$.

*Mutual information* is an entropy measure for interacting processes. Considering two processes as discrete random variables, mutual information measures the uncertainty reduction about one variable given knowledge of the other [78]. For information-generating processes $X$ and $Y$, it is expressed as

$$M_{XY} = \sum_{x,y} p\left(x, y\right) \log_2 \frac{p\left(x, y\right)}{p(x)p(y)}. \tag{3.2}$$

The uncertainty reduction increases with the value of $M_{XY}$, while $M_{XY} = 0$ indicates $X$ and $Y$ are independent.

*Transfer entropy* improves on mutual information, such as by preserving the directionality of influence [74]. For a process $X$, let $x_t$ give the state of $X$ at time $t$ so that $x_t^{(k)} = (x_t, \ldots, x_{t-k+1})$ represents the $k$-length time series of states of $X$ ending at time $t$ (and similarly for process $Y$). Then, transfer entropy is defined as

$$T_{Y \to X} = \sum_{x,y} p\left(x_{t+1}, x_t^{(k)}, y_t^{(l)}\right) \log_2 \frac{p\left(x_{t+1} \mid x_t^{(k)}, y_t^{(l)}\right)}{p\left(x_{t+1} \mid x_t^{(k)}\right)}. \tag{3.3}$$

Here, $T_{Y \to X}$ measures the degree to which $X$ depends on $Y$, and, unlike mutual information, $T_{Y \to X}$ is not necessarily equal to $T_{X \to Y}$. The values $k$ and $l$ control how much of the history of $X$ and $Y$ are considered.

### 3.2.1 Techniques for entropy of continuous variables

Entropy is often defined on discrete probability distributions. However, we will define opinion as a continuous variable. The entropy of continuous variables can be evaluated using binning or the symbolic approach.

*Binning* discretizes the continuous state-space of opinion values into discrete bins of some width (i.e., the density histogram of the data). Building discrete probability distributions in this way allows the use of discrete entropy measures on continuous data without altering the equations (e.g., Reinagel et al. [79]).

The *symbolic approach* transforms the input data by first mapping the elements of a time series into patterns of relative orderings between values (e.g., increase-increase, increase-decrease, etc.). Each pattern is identified by a distinct symbol, which allows the data to be described with a discrete probability distribution. An additional parameter for this approach is the pattern length, which varies the complexity of the analysis, as the number of patterns increases with the factorial of the pattern length. Symbolic transfer entropy is one example of this symbolic approach, applied to transfer entropy for frequency analysis of online messages surrounding specific, rare events [75].

## 3.3  Social influence network simulation design

In this section, experimental design is outlined by discussing factors that are potentially fundamental to the design of a social influence network. These factors involve structural aspects of the network and mechanisms by which members of the network influence one another.

This section is organized as follows. After discussing several definitions and notations used throughout this study, we present mathematical network structure models, agent activation schedules, social influence models, and techniques for incorporating stochastic error terms into otherwise deterministic influence models. These items are used as design factors in our study and are selected based on their presence (often singly) in existing social influence network studies and literature.

### 3.3.1  General definitions and notations

The structure of the social influence network is defined on an edge-weighted directed graph $G = (V, E)$, where $V$ is the set of agents, $E$ is the set of directed edges among these agents, $e(i, j) \in E$ is an edge directed from agent $i$ to agent $j$, and $w_{ij}$ is the edge weight for $e(i, j)$. For $e(i, j)$, we say $j$ is a neighbor of $i$, meaning agent $i$ can observe the opinion of agent $j$ and thereby be influenced by $j$. Network structures

are static during simulation runs. Edge weights model the fact that social influence between two agents need not be equal—nor even possible—in both directions. Here, the weights are set by sampling from the standard uniform distribution $U(0,1)$ for each edge and then normalizing the weights so that for each agent with outgoing edges, the weights for those edges sum to one; using $U(0,1)$ for creating edge weights is based on Garee et al. [80].

### 3.3.2 Network structure models

Network structure can affect opinion dynamics in social networks. Several structural models are prevalent in the social network literature, including regular lattices, Erdős-Rényi random networks, small-world, scale-free, and spatial networks [33, 51]. In this paper, we use the Erdős-Rényi random, small-world $\beta$, and scale-free network structure models; parameterization is detailed in Appendix B.

### 3.3.3 Activation regime

Activating an agent causes it to perform its assigned action, such as updating its opinion. The model's activation regime describes the order and frequency of agent activation, sometimes called the "scheduling" approach [52, 81]. Many models use only a single—often simplistic—activation regime [52], even though different approaches to the timing of agent interactions can lead to different outcomes [53, 82]. Choosing the correct activation regime remains an open problem [81]. In this work we consider three activation regimes featured in the literature: synchronous, uniform, and random [52, 81].

### 3.3.4 Influence models

The model for a social influence network defines how an agent's opinion changes due to interaction with its neighbors or other information sources (e.g., mass media in

a real-world network) [70]. This research only considers influence sources within the network by an agent's neighbors. Below, we survey several prevalent models involving continuous opinions. The seminal model of DeGroot [29] updates agent $i$'s opinion $o_i$ to the value of a simple weighted mean of the opinions from the agent's neighbors as time increases from $t$ to $t + 1$. The *standard model* of Friedkin and Johnsen [2] augments DeGroot's and is formulated as

$$o_i(t + 1) = a_{ii} \sum_j w_{ij} o_j(t) + (1 - a_{ii}) o_i(1), \qquad (3.4)$$

where $o_j$ is the opinion of $j$, $w_{ij}$ is the weight assigned to $j$'s opinion such that $w_{ij} \in [0, 1]$, $a_{ii} \in [0, 1]$ is agent $i$'s susceptibility to external influence, and $1 - a_{ii} = w_{ii}$ is the weight $i$ places on its own initial belief $o_i(1)$.

Flache et al. [53] describe an influence model containing a *similarity bias* that causes agents to ignore opinions too dissimilar from their own, acting as a type of homophily. Agent $i$ interacts with a single neighbor, $j$, if their opinions differ by less than a "confidence threshold" that is interpreted as a measure of the uncertainty that agents feel about their own beliefs. Agents with high uncertainty are willing to communicate with neighbors with larger differences in opinion from their own, compared to agents with low uncertainty. Then, if $|o_i(t) - o_j(t)|$ is less than $i$'s confidence threshold,

$$o_i(t + 1) = o_i(t) + \mu \left[ o_j(t) - o_i(t) \right], \qquad (3.5)$$

where $\mu$ is the rate of opinion convergence and other terms are as defined previously.

The *attractive-repulsive* model [53] allows for both attractive and repulsive effects between agents as a function of the degree of opinion difference: small differences between $i$ and $j$ cause their opinions to move toward one another, while large differences push them apart. Unlike the previous models where the new opinion was always within the range of the original opinions, this latest model can push opinions to new extremes. We define opinion to lie in the interval $[-1, 1]$ and use a simple truncating rule: if an agent would update their opinion to a value outside the interval $[-1, 1]$, they instead set it to the nearest boundary of that interval.

Models discussed thus far include some sort of weighting of the opinion of one or more neighbors. Alvarez-Galvez [83] takes a different approach in their study of social contagion called *individual random contagion* (or adoption): each time an agent would update its opinion, it adopts the opinion value of only one of its neighbors, selected at random.

### 3.3.5  Error terms in influence models

Real-world agents "may make errors in perceiving each others traits or similarity" [53]. This can be extended to agents misjudging each other's opinions, for example, due to noise in a communication channel or cognitive biases affecting all interactions. A stochastic error term is appended to each of the interaction equations from Section 3.3.4, representing the naturally imprecise nature of communication between social agents. Literature reviewed to date does not suggest a preferred error distribution for noisy social influence networks, so we assume the noise is created with the zero-mean normal distribution $N(0, \sigma)$, treating its standard deviation $\sigma$ as an experimental design factor.

### 3.4  Methods

We generated opinion data using a social influence network simulation. To create data from a wide array of scenarios, we used experimental design to vary the five simulation design factors (population size, network structure model, agent activation regime, influence model, and influence error term) and created 1800 trials. For each trial, we collected a time series of opinion data for each agent and used that to calculate the six entropy-based response variables that support our analysis in Section 3.5.

### 3.4.1 Experimental design

We used a full factorial design of 1800 trials, based on five design factors: population size, network structure model, agent activation regime, influence model, and influence error term. (Additional details appear in Appendix D.) Each design factor was varied over multiple *levels* (experimental settings). Population size $N$ was varied over the set $\{100, 1000, 10000\}$. The design had ten levels for network structure model: one level for Erdős-Rényi random graphs, six levels for small-world, and three levels for networks created via preferential attachment. These three models are prevalent in the existing social networking literature [51], and the parameters for small-world and preferential attachment were selected based on typical use cases in previous research [36, 84, 85]. Levels for the agent activation regime were synchronous, uniform, and random. The four influence models from Section 3.3.4—the standard model, similarity bias, attractive-repulsive, and random adoption—plus a nonlinear model motivated by the standard model formed the five levels for this design factor. The influence error term was distributed according to a zero-mean normal distribution with standard deviation $\sigma \in \{0.05, 0.1, 0.2\}$, or the error term was omitted entirely.

### 3.4.2 Simulation design

The simulation described in this section was implemented in the Python programming language using an agent-based modeling approach. Agents independently stored their own state information and updated their own opinion values when activated. Also, an agent did not have global awareness of the network, instead knowing only the identities of the agents it follows (its network neighbors) and their opinion values.

Every trial in the experimental design was replicated 100 times. Each replication took place over a series of 500 time steps. At each time step, agents activated according to the activation regime for the current trial, and when an agent activated, it updated its opinion according to the trial's influence model.

### 3.4.3 Entropy calculations

The three entropy measures discussed in Section 3.2, relative entropy, mutual information, and transfer entropy, rely on discrete data. As a first processing step, the continuous opinion data was discretized using the approaches from Section 3.2.1. For binning, an opinion value was mapped to a fixed-width bin; for the symbolic method, an opinion value was mapped to a symbol from a fixed alphabet based on its relation to the agent's prior and later opinions (the two approaches are used for different response variables, discussed in Section 3.4.4.). As a result, each opinion value was mapped to a discrete state. Probability distributions associated with these states were used for computing the entropy measures.

At each time step, entropy was calculated for each agent, based on its own states and those of its neighbors (as necessary), for that and all previous time steps $t$. Relative entropy uses the empirical distribution directly (Equation 3.1), while mutual information and transfer entropy also use joint distributions on pairs of agents or multiple time steps (Equations 3.2-3.3). We now detail the calculation of mutual information; the other measures follow the same process using their respective distributions and equations. Consider two agents, $i$ and $j$, whose opinions have been mapped to the set of discrete states $X$ and $Y$, respectively. Let $p(x, t)$ be the empirical probability distribution of $X$ based on agent $i$'s opinion data for all time steps up to and including time $t$ and define $p(y, t)$ similarly for agent $j$. Then, let $p(x, y, t)$ be the empirical joint distribution of the agents' states up to and including time $t$. Mutual information for agent $i$ at time $t$, following from Equation 3.2, is

$$M_{XY}(t) = \sum_{x,y} p\left(x, y, t\right) \log_2 \frac{p\left(x, y, t\right)}{p(x, t)p(y, t)}, \tag{3.6}$$

with the sum taking place over all pairs of states in $X$ and $Y$.

Each entropy measure required us to decide on certain parameters and approaches. For relative entropy, the agent's data was assessed relative to the uniform distribution, because it has the maximum entropy for any given state space and fits with the intuitive notion of how (un)informative a data set might be. Mutual information and

transfer entropy were computed only between neighbors. For these two measures, the value for a single agent was defined as the average value of the measure across all of the agent's neighbors. Further, transfer entropy includes parameters $k$ and $l$, the number of previous time steps considered for each of the agents. Here, we followed Schreiber [74] and de Assis and de Assis [86] and let $k = l = 1$.

### 3.4.4 Response variables

During each replication of the simulation, an $N \times t_{max}$ matrix was populated with agent opinion values, where $N$ is the size of the agent population and $t_{max}$ is the total number of time steps in the run. One element in the data matrix, $o_i(t)$, is the opinion for agent $i$ at time $t$. Each row of the matrix is a time series of opinions for one agent. We paired each of the two continuous data techniques with each of the three entropy measures to compute six entropy values per agent, for each time step within a replication of the simulation. This produced six entropy time series per agent, which were then aggregated by averaging across the population, yielding six aggregated entropy time series for a replication. Once all 100 replications for a trial were complete, the aggregated entropy data was averaged across replications. These final *response variables* were per-trial time-series data for each of the following:

1. relative entropy, binning (RE-B);

2. mutual information, binning (MI-B);

3. transfer entropy, binning (TE-B);

4. relative entropy, symbolic approach (RE-S);

5. mutual information, symbolic approach (MI-S); and

6. transfer entropy, symbolic approach (TE-S).

### 3.5 Results

The agent-based simulation experiment produced response variable time series data for 1800 independent trials. For each research question, we provide an analysis process overview for a single response variable (RV) before commenting on noteworthy differences among the other RVs.

In Figure 3.1, we show the general distributions of the RVs to provide an overview of the trial data. The distributions for RE-B and RE-S are similar in shape to each other, as are the four MI and TE RVs.



Figure 3.1. The time series of RE-B (left) and MI-S (right) for all trials are plotted with an adjacent kernel density estimate of the values at the final time step. These plots exhibit the two characteristic shapes of the data observed for all six response variables.

Levels are encoded as a single letter in many graphics in this section for better readability and are identified in the text as needed; a full listing appears in Table B.2.

### 3.5.1 Research question 1: Which system design factors contribute most to aggregated entropy?

For this research question, we explore the one-way sensitivity of each entropy response variable to changes in the levels of experimental design factors. This exploration includes qualitative comparisons of RV distributions when trial data is grouped

by experimental levels and statistical tests for differences between levels. These methods support a subjective evaluation of whether an RV is sensitive to changes in each design factor. We remark that all the one-way sensitivity analyses are based on the full set of data, and so illustrate the factors' main effects. Plots such as those used below could give misleading insights about main effects if the data only reflected one-at-a-time variation of each factor from a base case, rather than a designed experiment.

Analysis results for the relative entropy, binning (RE-B) response variable for the current research question are summarized in Table 3.1. For the overall evaluation (final table row), we find that RE-B is not sensitive to population size $N$ and activation regime, while it is sensitive to changes in the levels for the other three design factors. This evaluation is based on an experimental design main effect plot of data at the final time step (Figure 3.2) and plots for each design factor where each trial's RE-B time series is grouped by level (e.g., Figure 3.3), as well as the Kruskal-Wallace test on each factor and the Mann-Whitney U test on each pair of levels for each factor (Table B.3).

The distribution of the data is not unimodal for some RVs (Figure 3.1, left), so the full distribution of the data is considered rather than simply the mean. Each half-violin in Figure 3.2 shows the distribution of RE-B at the final time step of the simulation for each level. Differences between distributions among the levels for a factor qualitatively show the effect that the levels have on the response. For example, the distributions for population size $N$ are almost identical, so we infer that $N$ is not important (i.e., does not have a significant effect on the RV), at least over the levels used in the experimental design; the same holds for the agent activation regime. On the other hand, strong differences are visible for the influence models and error distributions, making these important to RE-B. Structure models appear to fall into at least two similar groups, and these groups resemble the higher/lower density networks (Appendix B.3). These observations are summarized in Table 3.1, item i.

The main effect plot uses data for only a single time step to focus on the final state of the network. Median lines of the grouped data over the full run of the simulation

Table 3.1.

The findings for research question 1, design factors that contribute most to aggregated entropy, for the response variable RE-B are summarized to support the overall evaluation of each experimental design factor (final table row). The parenthetical before items i-iv refer to the graphic or test that supports the item.

| | Factor (number of levels) | | | | |
|---|---|---|---|---|---|
| | $N$ (3) | structure (10) | influence model (5) | error (4) | activation (3) |
| i. | (Main effect plot) What differences are present between the response variable distributions for each level at the final time step? | | | | |
| | negligible | 2 or 3 patterns | 3 or 4 patterns | 3 patterns | negligible |
| ii. | (Grouped time series) What differences are present between the median response values over the duration of the simulation? | | | | |
| | negligible | overall similar shapes; small divide affected by density | 3 patterns; nonlinear+standard, similarity+random paired | 3 patterns; high+medium variance identical | negligible |
| iii. | (K-W test) Does the Kruskal-Wallace test indicate statistical differences in the response variable between each level at the final time step? (i.e., is the p-value <0.05?) | | | | |
| | no | yes | yes | yes | no |
| iv. | (M-W U test) How many pairs of levels are statistically different (p-value <0.05) according to the Mann-Whitney U test? | | | | |
| | 0/3 | 28/45 | 10/10 | 5/6 | 0/3 |
| * | (Evaluation) Is the response variable sensitive to changes in the level for the factor? | | | | |
| | no | yes | yes | yes | no |

Figure 3.2. Each half-violin of this design of experiments (DoE) main effect plot represents the distribution of relative entropy, binning (RE-B) at the final time step for all trials, with the corresponding level on the horizontal axis and its median indicated with a dash; the grand median is shown for reference. This plot suggests, in part, that $N$ and activation are unimportant to RE-B, while zero error (level a) leads to markedly different outcomes from the other error distributions.

reveal the impact of time on the RV (influence model is shown in Figure 3.3). On the plot for influence model, there is distinct separation between some levels and pairing between others. Trials using the similarity bias and random adoption influence models show similar behavior for RE-B, but these similarities vanish for other RVs. Specifically, for TE-B and MI-B, results for the two influence models behave differently from each other, while for the symbolic method RVs, they are similar to each other and the remaining influence models. For population size and activation regime, median lines heavily overlap. Overall, these figures reinforce what was observed in the main effect plot. They also show that for RE-B, the median response values are stable over time (after an initial transient), so observations made at the end of the simulation run should be informative about the system over a longer period of time. These findings are summarized in Table 3.1, item ii.

Grouped time series plots for the other RVs are similar in nature to those for RE-B (initial transient, then stability), though for mutual information and transfer entropy, many lines appear to gradually approach zero rather than some nonzero steady state value.

With respect to the current research question, population size $N$ is unimportant to all six RVs used in this experiment. Network structure, influence model, and influence error distribution are important for the RVs (and different levels of these design factors often lead to very different results). Finally, agent activation regime is important for mutual information and transfer entropy but not for relative entropy.

### 3.5.2 Research question 2: How is system design related to the distribution and scale of entropy time series data?

The preceding analysis shows that some system design factors have greater effects on trial response values than others. We now look for relationships in the distribution and scale using time series cluster analysis. For example, does the setting for the influence error term strongly drive response values into distinct clusters for each

median relative entropy, binning (RE-B), all trials, grouped by influence model



Figure 3.3. All trials are grouped by factor-level as in Figure 3.2 and the groups' median response over time is plotted (solid line). Shaded regions enclose the 25th to 75th percentiles of the data associated with the median line. Some grouping is apparent based on the influence model. One explanation is that the nonlinear and standard models use weighted averages of neighbor opinion, while similarity bias and random adoption interact with (at most) one neighbor at a time.

level? Here, a cluster is *meaningful* if it is heavily composed of one or more levels or can be well-described by some characteristic of the design factors. This is admittedly a subjective judgment, but it supports the notion of external cluster validity with respect to the experimental design factors. Numerical techniques do exist for external validity, but they presuppose a benchmark set of clusters, which do not exist for the current experiment.

In the following analysis, we produce two sets of clusters for each RV using different distance measures and assess cluster quality by visual inspection of the RV space and the concentration of each level within a cluster. Based on this analysis, variation in all design factors except population size can contribute to the presence of meaningful clusters when using dynamic time warping (DTW) as the distance measure. Error distribution and influence model contribute most strongly—often a cluster contains only a single level of one of these factors. For each RV, at least half of the total clusters are meaningful when using the DTW distance measure, while no clusters are meaningful when using Pearson's correlation coefficient.

Cluster analysis has three elements: distance (or dissimilarity) measure, clustering algorithm, and evaluation criteria of the resulting clusters [87]. Few guidelines exist for designing a cluster analysis *a priori*, so we use an assortment of options to search for meaningful clusters in our entropy time-series data. Distance is computed using both DTW and Pearson's correlation coefficient between each pair of trial time series for an RV. DTW is a robust technique for time series cluster analysis and Pearson's correlation is a well-understood way of comparing data sequences of any type. Then, each of fourteen clustering methods provided by the R library *NbClust* [88] suggested an optimal number of clusters, and the number with the most "votes" was selected as the consensus number of clusters. An agglomerative hierarchical clustering method was then used to assign each trial to a cluster, based on the consensus number. This process was applied to both distance measures.

We then conduct a census of the trials assigned to each cluster, with respect to the experimental design factors (Figure 3.4). For DTW, cluster 1 contains only (but

not all) trials with no influence error term (level a). Cluster 4 is strong in similarity bias and random adoption influence models (levels b and d). Interestingly, cluster 4 is also very weak in lower-density network structures (levels b, d, f, and h). While cluster 4 is meaningful, the contribution of both structure and influence model makes the cluster's precise nature unclear. The levels for population size N and activation regime are uniformly distributed within each cluster, further reinforcing the evidence from Section 3.5.1 that those factors are not important to RE-B. Using Pearson's correlation coefficient as a distance measure produced entirely undifferentiated clusters (Appendix B).



Figure 3.4. For each cluster produced with DTW, the trials assigned to the cluster are grouped by factor-level in order to find the percentage of a cluster associated with each factor-level. Within a cluster (row), the percentages for a *single* factor sum to one. For example, trials assigned to cluster 1 mainly use error level a (no error), and trials assigned to cluster 4 mainly use either influence model b or d. Within a factor-level (column), there are no such constraints.

In summary, variation in system design can produce clusters in the response variable space for RE-B that are meaningful with respect to the experimental design factors. This is achieved when using DTW as the distance measure, but not when using Pearson's correlation coefficient. When considering the other five RVs, DTW continues to produce one or more meaningful clusters and Pearson's correlation pro-

duces highly or entirely homogeneous clusters that lack any meaning. Levels that contribute to meaningful clusters most commonly include no influence error (level a) and the similarity bias, attractive-repulsive, and random adoption influence models (levels b–d). Activation regime and network structure density contribute to a lower degree. Population size $N$ is irrelevant in this cluster analysis. Overall, having no influence error (level a) may be the most influential: four of the six RVs have a DTW cluster composed entirely of that level. Only clusters for MI-B and TE-B are relatively unaffected by influence error. Instead, these each have a cluster composed entirely of the random adoption influence model (level d).

### 3.5.3 Research question 3: How do different entropy measures respond to changes in system design?

We next consider a brief comparison of all response variables together. Overall, two general behaviors are exhibited by the RV time-series data for the experiment (Figure 3.1). Mutual information and transfer entropy RVs are characterized by a rapid initial increase followed by a gradual decrease; lines for mutual information have sharper peaks, signifying a more rapid decrease than transfer entropy. Relative entropy, on the other hand, shows a more varied response and generally converges to an extreme value or one of several intermediate values. These two behaviors can be seen in other analysis products from the previous sections, as well. The main effect violin plots for the final time step of each RV can be qualitatively grouped into those with wide distributions and those with compact distributions and long, thin tails (Appendix B).

With respect to sensitivity to one-way variation in design factors (Section 3.5.1), all RVs are similarly affected by differences in network structure density and error term variance. All RVs are also insensitive to changes in population size. Both RE-B and RE-S are insensitive to the agent activation regime, unlike the MI and TE RVs. The influence model factor has the least consistent median responses across RVs.

For every trial, the Pearson correlation coefficient[2] between each pair of RV time series (15 correlation values per trial) is used to construct histograms of the correlations for each RV pair (Figure 3.5). Several pairs of RVs are positively correlated for most trials, while several other pairs have a more even spread of correlations. For relative entropy and transfer entropy, the binning and symbolic methods produce time series that are highly positively correlated overall; mutual information, however, has a wide distribution of correlation values with a positive peak.

Distributions of pairwise correlation per trial response variable



Figure 3.5. Broadly, MI and TE are positively correlated with each other and not correlated with RE. For each trial, we compute the Pearson correlation coefficient of the time series for each pair of RVs. These histograms reveal the correlation patterns for the responses. For example, the upper left distribution (MI-B, RE-B) displays a histogram of the correlation values between the MI-B and RE-B time series for all trials. The binning and symbolic methods for both RE and TE show generally high correlation, while those for MI show a wider spread.

---

[2]This use of correlation differs from that in the cluster analysis; there, it was a distance measure of an RV between trials, while here it compares RVs within a trial.

Despite differences between the response variable data for RE, MI, and TE, Figure 3.6 reveals some common themes in DTW cluster membership across the RVs. Perhaps most interesting is that four of the RVs have one cluster containing only trials with no influence error term (level a). For the other two RVs, MI-B and TE-B, they show little distinction between levels for the error term and instead present one cluster entirely composed of trials using the random adoption influence model (level d). The network structure levels do show separation between higher and lower density networks.

Overall, the six response variables react somewhat differently to the trials in our experimental design. The RVs can be grouped by the shape of their time series plots (i.e., their raw values) or main effect plots. They have a similar grouping with respect to their correlation values within each trial. A somewhat different pattern appears when considering their cluster analysis results. These aspects of our analysis will be further explored in the discussion.

Figure 3.6. This figure combines the DTW cluster membership plots (e.g., Figure 3.4) for all RVs, revealing similarities and differences between the RVs. For example, four of six RVs have a cluster made wholly with no error term (level a), and MI and TE are sensitive to the activation regime while RE is not.

### 3.5.4 Partition tree regression modeling

This final analysis step combines aspects of the previous results for the three research questions and sets a path to future analysis of design factor interactions. For each RV measured at $t = 490$, near the end of the simulation, a partition tree metamodel is constructed. Each tree is built using ten splits to standardize trees across the six responses. Then, we inspect the contribution of each design factor to the $R^2$ of the tree for each RV (Table 3.2), as inspired by work from Marlow et al. [89]. This data complements the evaluations for research question 1 (Tables 3.1 and B.4), though network structure is more impactful in those analyses than the current one. These results also align well with the DTW cluster analysis: factors that are impactful to a cluster show noteworthy contributions to the partition tree results.

Table 3.2.
Factors are classified by their contributions to $R^2$ for partition tree meta-models constructed separately for each RV, as measured at $t = 490$ (near the end of the simulation run). Not all factors appear in every tree. Trees are constructed with 10 splits to standardize across RVs. Classifications in the table below come from [89]: VVS, VS, S, M, w, and vw denote factor contributions to $R^2$ that are very, very strong ($> 0.50$), very strong ($> 0.25$), strong ($> 0.125$), moderate ($> 0.0625$), weak ($> 0.03125$), and very weak ($> 0.015625$), respectively.

| Factor | RE-B | MI-B | TE-B | RE-S | MI-S | TE-S |
|---|---|---|---|---|---|---|
| | | | Response variable | | | |
| Population size | | | | | | |
| Network structure model | w | S | | w | | S |
| Influence model | VS | VVS | VVS | M | VS | VS |
| Error distribution | VVS | M | vw | VVS | M | VS |
| Agent activation regime | | M | S | w | VS | S |
| $R^2$ | 0.880 | 0.852 | 0.806 | 0.824 | 0.367 | 0.593 |

This analysis approach will support future exploration of interaction effects between design factors. For example, a simpler partition tree on RE-B using five splits (which still achieves a comparable $R^2$) lets us observe an interaction between the error distribution and influence model. For trials using no influence error (level 'none'), the random adoption and similarity bias influence models yield a mean RE-B that is 1.21 units less than trials with the other influence models; however, for trials using influence error (the remaining three levels), the same split in influence model reveals a difference of 2.05 units. These values correspond to a 41% change in the impact of choice of influence model, depending on the use or omission of the influence error term in the social influence network simulation.

## 3.6    Discussion and conclusions

In this study, agent-based simulation and experimental design are used to explore the degree to which social influence network simulation design affects aggregated entropy. Based on this research, variation in system design can affect the distribution of entropy time series (Section 3.5.1); the strength of this effect depends on the particular system design factor and response variable (RV) in question. Changes in system design can induce changes in time series of entropy measures that lead to meaningful clusters of experimental trials (Section 3.5.2). Different entropy measures respond differently to changes in system design (Section 3.5.3). As a result, the aggregated entropy measures presented in this study may be useful tools for inferring properties about real-world social influence networks from opinion data of its members.

### 3.6.1    Major findings

In addition to the high-level observations for the research questions posed in this study, we present the following findings. For the six RVs in this study, varying the population size $N$ had no significant effect. This may be because each RV is computed as the mean of the entropy values for individual agents, or it may reflect something

more fundamental about entropy in networks. However, the RV distributions for each level of $N$ are not identical, only very similar, so it is possible that networks with much smaller or larger populations would show more diversity in the responses.

The selection of influence model is responsible for the greatest amount of variation among RVs. The random adoption model is naturally the most disordered and presents the most extreme differences from the other influence models, so it is surprising that it strongly attracts a cluster for MI-B and TE-B yet is unimportant to the clusters for RE-S and TE-S (Figure 3.6). In the cases where the standard and nonlinear models show very similar behavior (as for the three binning RVs), this may be due to how their equations are structurally similar, involving a weighted mean of neighbor opinions.

Existing social influence literature rarely considers error or noise in the influence process, yet real-world social systems are often noisy. The trials with error standard deviation 0.1 and 0.2 (levels c and d of the error term design factor) showed no significant difference. There are often differences between those trials and trials with smaller standard deviation or no error term. The grouped median responses for error distribution do not intersect for RE and TE; for MI, the "none" line does cross one or more of the non-zero error levels. Present work does not explain this difference, but it contributes to a larger point that care must be taken in applying findings from error-free models to noisy real-world systems.

The level of activation regime leads to the greatest variation for MI and TE, little variation for RE-S, and no variation for RE-B. This is reasonable, because RE-S and RE-B measure entropy relative to a uniform distribution and do not directly consider opinion differences between neighbors, so the update sequence need not affect the outcome. MI and TE, on the other hand, are functions of information transfer between agent pairs, so the timing of opinion updates should matter. These findings support conclusions from existing work (e.g., [52]) that careful selection of activation regime is important for agent-based modeling.

All RVs for most trials show an initial period of rapid change before stabilizing. This could allow for shorter simulation runs or for reasonable predictions to be made about the systems well beyond $t = 500$. Some of this may be due to how RVs are calculated: at each time step, the entropy is based on the agent opinions for all previous time steps, so the number of observations per data point increases with $t$. Entropy measures do tend to be calculated using all available information, but it may be informative to compare these results to responses using a fixed-width rolling window.

Pearson's correlation coefficient may not typically be used as a distance measure for cluster analysis, but it is common way of comparing two sets of observations, so we expected it to have some utility in this study. Unfortunately, it proved entirely unsuitable as a distance measure between response variable time series. Every set of clusters created using correlation as the distance measure was almost completely homogeneous, in contrast to clustering via dynamic time warping, which produced differentiated clusters.

### 3.6.2 Limitations

Despite the breadth of the experimental design, response variables, and analysis techniques, several factors may limit the detail or applicability of the results. An apparent divide between network structures of higher and lower densities occurs throughout our analysis, leading to grouped patterns in the main effect plots, stratification in the grouped time series plots, and some visible effects on cluster assignments. Density differences align with the variation in observed results more strongly than degree distribution or network shape (e.g., tree, lattice, etc.). At this time, however, we do not claim density to be a driving factor due to the limited range of density values present in the network structures in our experiment (e.g., densities in the interval $[0.01, 0.07]$ for $N = 100$ networks). A new experiment using a wide range of densities as a design factor, rather than network structure model, could shed light on this idea.

Real-world communication errors may not be well-described by zero-mean normal distributions but could actually be biased. If so, then the error distributions in this experiment may distort social influence processes.

With respect to the analysis on the experimental design, we principally studied the main effects. Investigation of two-way (or more) interaction terms may be a direction for later work, as motivated by the partition tree analysis in Section 3.5.4.

The cluster analysis in Section 3.5.2 produced interesting results, but they are best treated as a starting point for future investigation. Selection of an appropriate distance measure is perhaps the most important step of a successful cluster analysis [90]; we considered merely two distance measures, and only dynamic time warping showed any success. Also, our consensus method used many clustering algorithms to suggest an ideal number of clusters, but we studied a single set of cluster assignments (per RV, per distance measure). Finally, the definition of "meaningful" clusters in this study is subjective. Future work could improve on this by classifying trials in advance, based on their experimental design settings and then measuring cluster quality with respect to those classifications.

### 3.6.3  Future work

The major findings can help refine future work on social influence networks, system design, and entropy analysis. In addition to studying interaction effects of design factors and making improvements to cluster analysis, there exist several avenues for future work.

Thus far, we have considered networks inhabited by homogeneous populations of agents: in a single network, agents may differ from one another in the specific values of their attributes, but they share a common set of attributes, follow the same interaction rules, and are created using a common process. This is not an accurate representation of most real-world social networks. A nonhomogeneous population, in contrast, has disjoint sub-populations (or classes) of agents that follow different

rules or are initialized in different ways. Future work is planned to incorporate non-homogeneous social network models from existing literature into our simulation and analysis design.

Networks in this study are static in their structure, while real-world networks are often dynamic. A suitable network evolution model could allow the experimental design factors and entropy response variables to be applied to a changing network. One obstacle is that real-world network dynamics depend on the interaction model between agents, which we take as a design factor, so trials with different interaction models may yield markedly different networks.

Given the growing interest in social network analysis and complexity studies, the current results and potential outcomes of future research may be of much practical value.

# 4. EFFECTS OF NONHOMOGENEOUS AGENTS IN SOCIAL INFLUENCE NETWORKS ON SYSTEM-LEVEL ENTROPY

Article in preparation for submission. Authors: M. Garee, H. Wan, and M. Ventresca.

**Article abstract:** Analysis of social influence networks typically involves agents that act according to a shared set of rules. This simplifies analysis but sacrifices the diversity present in real-world social networks, potentially limiting the relevance of any findings. Here, we design influence networks that have nonhomogeneous agent populations to create four distinct scenarios to compare against a homogeneous baseline system, based on examples from current literature. Agents are assigned to classes that alter their influence model, network participation, homophily with same-class neighbors, and other initial conditions. Agents influence their neighbors and opinions change over time by way of an agent-based simulation. System-level entropy measures, based on relative entropy, mutual information, and transfer entropy, are calculated on opinion time series data. All systems produce similar output distributions when aggregated over many independent trials, and changes to system design factors have similar impacts on all scenarios. At the individual trial level, however, nonhomogeneity can create drastic changes in system-level entropy as compared to a baseline system, and these changes vary in magnitude across the four scenarios. The results suggest that accurate modeling of nonhomogeneous populations may be vital when studying singular systems but could be discounted in studies with high replication counts.

## 4.1 Introduction

Studies in social network analysis are dominated by populations of functionally identical agents. Within these homogeneous populations, agents interact using a common set of rules and often differ only in their position in the network and some numerical initial conditions. Such populations are straightforward to work with (e.g., a single model to consider), especially in analytic models, but fail to capture the diverse behaviors exhibited in real-world social groups [91, 92]. In this study, we feature simulated nonhomogeneous agent populations and explore whether the literature's focus on homogeneous populations degrades the ability of social network research to properly address real-world issues.

A social network is nonhomogeneous (or heterogeneous) if some property of the network is not uniform across the population, such as attributes or functions of the nodes or the relationship represented by the edges. Node-based variation, especially with respect to interaction behaviors, is the focus of this study. Perhaps the most basic form of node-based nonhomogeneity in networks appears when agents represent different types of objects, such as authors and documents in citation networks [93]. This can lead to different interaction modes between each type of agent, which in turn can promote community development around shared interests or attributes [94].

Nonhomogeneous social networks have value beyond academic interest. In epidemiology, accurate models of disease spread are essential for selecting effective mitigation strategies, yet many studies use homogeneous mixing models that fail to account for differences in infectivity and vulnerability among members of the population, requiring researchers to act on intuition rather than data [95]. National security also relies on nonhomogeneous networks being modeled properly: dynamic network analysis is used to support vulnerability assessments of terrorist networks, which feature a variety of entity classes, such as actors, resources, and organizations [22].

In this study, agent classification and behavior within a network are the properties treated as nonhomogeneous. However, some authors use the term to describe variation

in the edges between agents or numerical values within each agent. Several examples should help to disambiguate the term from the current experimental context. Santos et al. [92] simply apply *nonhomogeneous* to networks where agents have different edge degrees (which covers all networks except regular lattices). In Cai et al. [96], multi-relational networks are considered, where edges represent disparate types of links between agents (e.g., friendships versus business relationships). Galeotti et al. [97] model network formation processes where the costs of creating a link between two agents, and the benefits derived from having that link, depend on properties of both agents that can vary across the population. Finally, Delre et al. [98] observe that in diffusion models, diffusion speed increases with the heterogeneity of adoption thresholds held by agents; this type of nonhomogeneity does not change the rules for behavior, only the conditions under which a common action takes place.

The purpose of this study is to explore how introducing nonhomogeneity into social influence network populations alters system-level entropy. This serves as an extension of work by Garee [91], which examined the relationship between system design and entropy in social networks composed of homogeneous agents. Specifically, that study considered one-way sensitivity of several entropy measures to each of five system design factors, cluster analysis with respect to the experimental design, and relative differences in behavior of the six response variables. There, they found, in part, that the influence model (interaction equation) contributed most to the variation seen among response variables, the presence or absence of noise in the interactions did affect the entropy measures, and there was no significant relationship between population size and entropy. Here, we expand their findings by first reducing the system design space and then creating four scenarios with nonhomogeneous agent populations based on compatible examples from the literature [60, 72, 99, 100]. Differences among social network participants can generally be viewed as nonhomogeneous features, but in the current work, nonhomogeneity is considered only with respect to agent classification and class-associated behaviors. A subset of the data and analy-

sis procedures from the original study [91] is used to compare the nonhomogeneous population results to the homogeneous base cases.

In each scenario, initial conditions and interaction rules between agents are varied in some way across the population. Scenario 1 features *uninformed* agents who do not participate in the exchange of opinions until communication with an informed neighbor occurs. Scenario 2 uses two distinct influence models (friendly and partly antagonistic) and skews the starting opinions either positively or negatively to represent bias toward an extreme position. In Scenario 3, an online social network composed of human users and *bots* is modeled, and agents show a level of homophily toward others of the same type. Finally, Scenario 4 designates a subset of the population as *stubborn* agents that never change their own opinions but can still influence others.

This exploration uses a combination of analysis techniques: direct comparison of entropy time series data between homogeneous and nonhomogeneous simulation trials, response variable sensitivity to changes in individual design factors, scenario identification through cluster analysis, and correlation patterns between entropy measures. Broadly, response variables for the nonhomogeneous scenarios as a whole have similar distributions as the homogeneous base cases, but individual trial results often change dramatically when scenario rules are applied. System design factors have consistent effects across the four scenarios, and not all scenario-specific factors are important to the response variables. Cluster analysis using dynamic time warping on system-level entropy is ineffective at assigning trials to scenarios. Finally, the response variables show similar trial-level correlation patterns across all scenarios and the base case.

The rest of this paper is organized as follows. Section 4.2 details the experimental design and scenario construction. In Section 4.3, we present the results of the analysis, and Section 4.4 concludes the work.

## 4.2 Methods

The methods for this work expand on those from Garee [91], which considered only homogeneous agent populations in social influence networks. By creating non-homogeneous populations, we can explore whether they are better suited for studying real-world social networks (which are rarely homogeneous). Here, we use a subset of their experimental design (Section 4.2.1) and all of their system-level entropy response variables (Section 4.2.3), in addition to the general design approach, described presently.

Influence occurs between agents in a social network. Edges connecting agents are directed; the edge $e(i, j)$ represents the relationship where agent $i$ can be influenced by $j$ (i.e., $j$ is a neighbor of $i$). Agents have an *opinion* (about some topic), bounded to the interval $[-1, 1]$. This interval supports positive, negative, and neutral opinions, with the assumption that it is feasible to feel positive and negative opinions with equal intensity. When an agent is activated, it may revise its opinion according to an influence model using its neighbors' opinions, and the revised value may be affected by gaussian noise that represents the imprecise nature of communication. Agents activate over a series of time steps, which forms one run of the agent-based simulation. System parameters including the network structure, influence model, distribution of noise, and agent activation timing are controlled by the experimental design.

### 4.2.1 Experimental design

Population size $N$ is fixed at $N = 1000$ for this experiment, because $N$ was found to be unimportant to entropy analysis in [91] and is the median value used there. Seven of the original ten network structure models show differentiated response variable behavior and are used here. Four influence models, three influence error distributions, and two activation regimes are kept from the original experimental design for the same reason. This reduced experimental design forms the *base*

*case* as summarized in Table 4.1; details appear in the Supplemental Information, Section D.1.

Table 4.1.
The experimental design from Garee and Ventresca [101] is reduced to four factors and 168 total trials based on insights from their results. This design is the *base case* upon which the nonhomogeneous scenarios are constructed.

| structure | influence model | error | activation |
|---|---|---|---|
| Erdős-Rényi random | standard model | none | synchronous |
| small-world(0.0, 3) | similarity bias | $N(0, 0.05)$ | random |
| small-world(0.0, 10) | attractive-repulsive | $N(0, 0.10)$ | |
| small-world(0.66, 3) | random adoption | | |
| small-world(0.66, 10) | | | |
| scale-free(1) | | | |
| scale-free(5) | | | |

### 4.2.2   Nonhomogeneous agent populations

In a nonhomogeneous population, each agent is a member of one *class* of several defined for the system. The class associates certain properties and behaviors to its members that make them distinct from members in other classes; anything not explicitly defined for a class is assumed to be common for members of all classes. Simple agent classes may have different opinion update frequencies, while more novel, complex classes may use different interaction models and treat agents of the same class preferentially.

There may be countless ways to create sets of classes on social influence networks. In this work, we define four scenarios using nonhomogeneous agent classes, motivated

by existing literature. These scenarios provide diversity in the influence models, initial opinion bias, activation regimes, and edge weights. Class definitions are adjusted from the source material to align with model considerations and constraints in the current experiment. Each scenario is detailed in this section, and new experimental design factors for these scenarios are summarized in Table 4.2. The new design factors augment the default design (Table 4.1), except where certain factor-levels are excluded for a particular scenario.

Table 4.2.
New experimental design factors are introduced and existing factors are omitted for several of the scenarios described below. These new factors augment the default factors (Table 4.1), which are modified as needed for each scenario. In total, 1974 trials make up this experiment.

| Factor | Levels |
| --- | --- |
| *1. Informed/Uninformed, based on [60] (840 trials)* | |
| fraction uninformed agents | $\{.25, .33, .50, .66, .75\}$ |
| | |
| *2. Concord/Partial Antagonism, based on [99] (378 trials)* | |
| fraction Concord-type agents | $\{.25, .50, .75\}$ |
| fraction left-oriented agents | $\{0, .25, .50\}$ |
| (omit influence model as factor) | |
| | |
| *3. Bots/Humans, based on [100] (84 trials)* | |
| no additional factors | |
| (omit activation regime as factor) | |
| | |
| *4. Stubborn/Normal, based on [72] (672 trials)* | |
| fraction stubborn agents | $\{.05, .15, .33, .50\}$ |

**1. Informed/uninformed**. Banerjee et al. [60] focus on a standard DeGroot learning model, generalized to use uninformed agents. An agent is either *informed* (has an opinion) or *uninformed* (has no opinion). If an uninformed agent activates and

has at least one informed neighbor, it becomes informed permanently and activates normally; otherwise, it does nothing. Informed agents ignore uninformed neighbors when updating their own opinions. If all agents become informed, the simulation proceeds with a homogeneous population. While an agent is uninformed, its opinion is treated as zero for the purpose of entropy calculation.

For trials using these agent classes, the initial fraction of uninformed agents is varied as a design factor to span the range while avoiding homogeneous networks at 0 and 1. This new design factor is combined with all base case factors (Table 4.1).

**2. Concord/partial antagonism**. In Kurmyshev et al. [99], they build upon an existing homogeneous model of bounded confidence by introducing a "mixed model that takes into account two psychological types of individuals." Concord agents (C-type) are friendly and always move their opinion closer to their neighbor's when interacting, while partial antagonism agents (PA-type) move toward or away from their neighbor's depending on the degree of difference in their opinions. These C-type and PA-type agents align with the similarity bias and attractive-repulsive influence models, respectively.

This scenario has a second level of diversity from adding a bias to the distribution of initial agent opinions. Normally, the agents would have their initial opinions drawn uniformly at random from the interval $[-1, 1]$. With the bias factor, which [99] describes as "political preference," agents designated as "left-oriented" draw their initial opinion from the interval $[-1, 0]$, while "right-oriented" agents draw theirs from $[0, 1]$.

For the trials using these agent classes, we vary the fraction of C-type agents and the fraction of left-oriented agents as design factors. (Due to left-right symmetry, we need only consider half of the latter's range.) Agent type and orientation are assigned independently. These two new design factors are combined with all base case factors except influence model.

**3. Bots/humans**. Gilani et al. [100] classified real-world Twitter user accounts as either humans or automated agents ("bots"), and then they conducted feature

analysis on network activity. From their observations on the behaviors of the two classes, they found that bots post more content on average than humans: bots generated roughly 52% of content in their data set while comprising 43% of the user accounts. Also, users show a type of homophily by forwarding messages from members of their own class more frequently than the other class (twice as often for humans, three times as often for bots).

We adapt these results for the current experiment to set the fraction of the population assigned to each class and to bias edge weights to favor neighbors in an agent's own class (representing greater forwarding frequency). We also alter the random activation regime to select bots more often, since they generate more content; this aspect of the scenario is not compatible with the synchronous and uniform activation regimes, so those levels are excluded. There are no new design factors for this scenario, so the scenario design is applied to all base case factors except activation regime.

**4. Stubborn/normal**. In Xie et al. [72], pairs of agents in an undirected network influence each other depending on the classes of the two agents. Specifically, two normal-class agents change their opinions with a similarity bias model, a normal agent and a stubborn-class agent interact with some low probability based on their current opinion difference, and two stubborn agents do nothing. In effect, a stubborn agent blocks the flow of information.

We modify this concept for directional networks: stubborn agents never change their opinion, and normal agents always use their assigned influence model. For the trials using these agent classes, we vary the fraction of stubborn agents to span half the possible range while avoiding homogeneous networks at 0 and overly-static networks with more than half the population set as stubborn. This new design factor is combined with all base case factors.

### 4.2.3    Response variables

The simulation response variables are based on three entropy measures and two approaches for discretizing continuous data, as used in [91]. Relative entropy (measured with respect to the uniform distribution), mutual information, and transfer entropy are applied to agent opinion data over time. Relative entropy of an agent's opinion $X$ is defined here as

$$D_X(p \parallel q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)}, \tag{4.1}$$

where $p$ is the discrete probability distribution of the agent's opinion over time and $q$ is a uniform distribution with the same support as $p$. Mutual information is measured between two agents ($X$ and $Y$) as

$$M_{XY} = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}. \tag{4.2}$$

Finally, transfer entropy is also measured between two agents while considering the history of the agents' opinions. In general, any amount of history can be included in the calculation, but we follow the convention of using a single time step in the histories of $X$ and $Y$, such that

$$T_{Y \to X} = \sum_{x,y} p(x_{t+1}, x_t, y_t) \log_2 \frac{p(x_{t+1} \mid x_t, y_t)}{p(x_{t+1} \mid x_t)}. \tag{4.3}$$

Here, $T_{Y \to X}$ measures the degree to which $Y$ influences $X$, and unlike mutual information, $T_{Y \to X}$ might not be equal to $T_{X \to Y}$.

These measures operate on discrete state spaces, so agent opinions are discretized by binning (as with histograms) and by the symbolic approach, where patterns based on increasing or decreasing values are mapped to symbols (e.g., increase-increase maps to $A$, increase-decrease maps to $B$). Each simulation trial is processed to create a time series for the following response variables:

1. relative entropy, binning (RE-B);

2. mutual information, binning (MI-B);

3. transfer entropy, binning (TE-B);

4. relative entropy, symbolic approach (RE-S);

5. mutual information, symbolic approach (MI-S); and

6. transfer entropy, symbolic approach (TE-S).

To compute a response variable, the corresponding entropy equation and continuous data approach (binning or symbolic approach) is applied to the agents' opinion values at each time step, creating an entropy time series for each agent. These time series are averaged across the population to create system-level results, and then these are averaged across all replications of a single trial to create the trial-level response variable.

## 4.3    Results

The four scenarios led to 1,974 independent simulation trials, each composed of 100 replications lasting 500 time steps. Further, a subset of the trial data from the original study on homogeneous populations is included as the *base case*, using the 168 trials that share the reduced experimental design.

### 4.3.1    Comparison between homogeneous and nonhomogeneous trials

As a first analysis step, the scenarios are visually compared to each other and the homogeneous base case by using their raw response variable (RV) time series (Appendix D Figure D.1). Each scenario shows qualitatively similar behavior in the time series plots for every RV, with respect to the profile and visual distribution of the data. Profiles for relative entropy show a rather uniform distribution of trials that rapidly decrease from a maximum (that varies between scenarios) to near zero, trials that decrease slightly and remain near the maximum, and trials that fall somewhere in-between. On the other hand, the profiles for mutual information and transfer

entropy are dominated by a rapid increase to a maximum followed by a gradual but continuous decrease, except for a few trials that appear to constantly increase in entropy.

Instead of using raw RV time series data, we can consider the scenario trial results with respect to the homogeneous base case, which captures the effects of transforming a homogeneous system using the scenario rules from Section 4.2.2. Each trial can be paired with a single base case trial[1] that uses the same design factor-levels but does not use any scenario-specific rules. In this way, the effect of applying scenario rules to a homogeneous system can be directly observed. The percent change between the a scenario trial RV and the corresponding base case data is used to account for the different ranges of each RV (Figure 4.1). Of the 11,844 percent-change time series (from all trials across the four scenarios and six response variables), 3,252 (27%) showed changes of less than 10% for each time step, while 952 (8%) differed from their respective base case trial by over 100% at one or more time steps.

For individual RVs, no patterns emerge across the four scenarios.

For individual scenarios, however, some patterns can be found across the RVs. Scenario 4's stubborn agents induce changes from the base case unlike any other aspect of the nonhomogeneous scenarios. In Figure 4.2, we observe four bundles of lines plotting the percent change in MI-S between Scenario 4 and the corresponding base case trials. Each bundle corresponds to one level of the design factor for the fraction of stubborn agents in the population, and as this fraction increases, the core of the bundle becomes more negative, which indicates that as the population becomes more stubborn, the raw entropy values approach zero. This effect occurs for both MI and TE response variables, likely because there is no flow of information into stubborn agents. For RE-B and RE-S, the effect is not present. The large increases for RE may be explained by how stubborn agents' fixed opinions maximally deviate from the uniform distribution on the full range of opinion $[-1, 1]$.

---

[1]Scenario 2 fits with two base case trials since its agents use both the similarity bias and attractive-repulsive influence models. We compare this scenario's results to a weighted average of the two base case trials, where the weights correspond to the fraction of agents using each influence model.

Percent change from corresponding base case data



Figure 4.1. For each scenario trial, RV data from a homogeneous base case system is selected (or constructed, for Scenario 2). The percent change between the base case and scenario trials illustrates the effects of applying nonhomogeneous scenario rules to the original system. (All response variables are nonnegative.)

Percent change from corresponding base case data, MI-S, Scenario 4

Figure 4.2. The MI and TE response variables decrease relative to the corresponding base case trials as the fraction of stubborn agents in the population increases. MI-S is shown here, but MI-B and both TE RVs behave similarly.

For Scenario 2, as the left-biased fraction is increased from 0 to 0.5, the distribution of the agents' initial opinions approaches the uniform on $[-1, 1]$, and therefore the MI and TE results approach those of the homogeneous base case (not pictured). Scenario 1's uninformed agents do not have a well-structured effect on the base case. Scenario 3 introduced subtle changes in the agents to make a nonhomogeneous population, making humans and bots somewhat homophilic to neighbors of the same class; the differences between these trials and the base case are likewise subtle.

### 4.3.2   Importance of design factors on system-level entropy

We assess the importance of individual design factors to the response variables by grouping trials by their level for a given factor and plotting the group median response value over time (e.g., Figure 4.3). The grouped medians plots are used to visualize the impact of each design factor on the RVs, based on attributes such as grouping, relative separation, line shape, etc. Detailed results appear in Table D.2.

A design factor is rated as *important* or not for each scenario-RV pairing; importance is based on qualitative analysis of the grouped medians plots as was done in [91] to identify factors "having a significant effect on the RV." Quantitative approaches do exist; for example, we computed the covariance of the group median lines, which reinforced the visual analysis but failed to capture any clustering of those lines. Other numerical methods may capture both. Network structure, influence model, and error distribution are rated as important in more than 83% of the evaluations and are important for all RVs except MI-S; activation regime and scenario-specific factors are important in 54% and 56% of the evaluations, respectively. The scenario-specific design factors are of rather mixed importance across the RVs, but each factor is important to at least one RV.

Several observations from the grouped medians plots of the scenario-specific design factors serve to illustrate the impact of those factors on the RVs (Appendix Figures D.2-D.4). For RE-B, the median lines grouped by the fraction of initially

median mutual information, binning (MI-B), Scenario 2, grouped by structure



Figure 4.3. In each grouped medians plot, a solid line plots the median response value for all trials with the specified factor-level, and shaded regions enclose the 25th to 75th percentiles of the data associated with the median line. Here, the network structure design factor has a noteworthy impact on MI-B in Scenario 2 (concord/partial antagonism). Network density partially explains the observed grouping of levels.

uninformed agents (Scenario 1) for values below 0.5 rapidly reach their minimum and then gradually increase over time, whereas those for 0.5 or greater do not increase over time. The separation between lines and the presence of different behaviors suggest this design factor to be important to RE-B. Increasing the fraction of Concord-type agents (Scenario 2) decreases the median RE-B. Increasing the fraction of left-biased agents increases median RE-B. Median MI-B decreases as the fraction of stubborn agents increases (Scenario 4), since stubborn agents inhibit information flow between agents. This design factor is important to MI-B.

### 4.3.3  Additional analyses

The nonhomogeneity of the agent population may cause the response variables to behave in a distinctive way (e.g., Scenario 2 in Figure D.1). Specifically, the scenario used to generate the data may be identifiable using cluster analysis. As a test of this idea, we apply dynamic time warping (DTW) to the time series data for one response variable (RE-B) and use hierarchical clustering to assign trials to clusters. This clustering approach performs poorly overall, achieving a best-case identification accuracy of 0.32. Given the visual similarity between scenarios in each of the raw RV time series plots, the low accuracy is unsurprising. Details are available in the appendix (Section D.3).

Finally, we compute the correlation between each pair of RV time series for each trial and construct histograms on this data for each RV pair (Figure D.6). The correlation histograms have similar profiles for each scenario, so population homogeneity (due to scenario design) has no clear effect on the relationship between RVs for a single trial.

## 4.4  Conclusions

There is tension in the overall results. On one hand, the distributions of RVs for nonhomogeneous scenarios are qualitatively very similar to those of the homogeneous

base case trials. On the other hand, the RVs of individual trials from each scenario can differ radically from their base case counterparts (Figure 4.1). So, while the nonhomogeneous populations can introduce changes in the selected response variables for *single* trials, there is not enough deviation from homogeneity to cause radical changes in the response variable profiles of many trials taken together. This suggests the importance of accurately modeling nonhomogeneous populations may be reduced if the research focus is on aggregate behavior of many systems rather than detailed behavior of a single system.

Some types of population nonhomogeneity are more impactful to system-level entropy than others. Agents who do not participate in opinion exchange act as information bottlenecks and can have a large effect on the system, especially if they exert influence over others (i.e., stubborn agents). The presence of multiple influence models in a network has less impact on the response variables, so it is not a particularly significant type of nonhomogeneity. Low levels of homophily applied to edge weights is not impactful.

Scenario design has little effect on the importance of individual system design factors on the system-level entropy response variables, over the range of scenarios used in this experiment. This suggests a relationship between system design and entropy that is robust to variation in population nonhomogeneity. Similarly, the correlation between RVs for individual trials is robust to variation in scenario design.

Scenario-specific elements can impact entropy measurements. Stubbornness increases relative entropy and decreases mutual information and transfer entropy. Uninformed and stubborn agents have much in common, but Scenarios 1 and 4 have vastly different effects on the homogeneous case (Figure 4.1). Uninformed agents eventually become informed and their opinions can change, but stubborn agents never change. Also, uninformed agents are ignored by the rest of the network, while stubborn agents continue to influence others. These aspects of the two scenarios may explain the observed differences.

Cluster analysis using dynamic time warping and hierarchical agglomeration on the response variable time series does not achieve high accuracy in matching trial data to scenarios. Other clustering methods may give higher accuracy, as could using equal sample sizes across the scenarios. However, the findings here demonstrate that different types of nonhomogeneous population design can manifest in the output data in distinctive ways: Scenarios 2 and 4 each had over forty percent of their trials grouped together, while the other scenarios had thirty percent or less.

A key limitation of the work is that the scenarios use only two agent classes each. Real-world social networks are not limited in the diversity of their populations, and while not all types of diversity have an effect on system-level entropy, it is likely that there are ways to vary the populations not included in this study or the surveyed literature. Future work can include networks with greater nonhomogeneity, such as by using more than two classes at once. An additional limitation is the reliance on visual, qualitative analysis. Although it is an accepted technique for time series analysis, quantitative methods exist, and they could yield better results.

While this study may not cover the full range of possibilities in social influence network design, these findings should prove informative to researchers seeking to add greater realism and diversity to their social network analysis.

# 5. A COMPLEXITY MEASURE FOR OPINION DYNAMICS
# IN SOCIAL INFLUENCE NETWORKS

Article in preparation for submission. Authors: M. Garee, H. Wan, and M. Ventresca.

**Article abstract:** In the age of big data, it has become routine to observe opinion dynamics on social systems. Their study has facilitated research of political polarization, technology adoption, and more. In this paper, we propose a new method for using opinion dynamics to measure complexity in social influence networks, which may be useful for inferring properties of real-world social systems via the dynamics of individual opinions or behaviors. A framework built on a context-dependent, subjective notion of complexity supports the development of a dissimilarity measure based on functions of the location, velocity, variance, and entropy of network-wide opinion changes over time. Leveraging simulated social network data as a case study, we find that complexity is sensitive in varying degrees to different simulation design factors, such as influence model and population size. These findings suggest that the new complexity measure could be useful for differentiating social network designs via the population's opinion data, perhaps further acting as a foundation for future studies that combine social influence and complexity theory.

## 5.1 Introduction

Social influence networks are systems wherein social agents interact with one another in order to exchange information and influence each other's opinions over time. These systems can be used to study political polarization [19], purchasing decisions [102], the spread of innovation [23], and many other topics. Multiple disciplines take

interest in the opinion dynamics that result from interactions among network members over time [103], with special interest on the conditions that lead to group consensus versus fragmentation [104] and the roles that society's most influential members play in affecting public opinion [105]. Recent applications also include the influence of bots and "fake news" in social networks on the outcome of political elections [106, 107].

Characteristics of individual agents can be inferred from the time series of their opinion values; we refer to this time series as the agent's *opinion history*. For example, an agent with a fixed opinion may be described as stubborn or may simply not be receiving new information [72], while an agent that attracts others to its opinion may be unusually influential and a good target for word-of-mouth marketing [102]. Can the collective opinion histories of the network population be used in a similar way to inform us about the social influence network as a whole? Rather than comparing agents within a single population, we wish to compare entire networks with one another. For instance, similarities in the opinion histories of two groups of consumers may help a product recommendation system make better selections.

The dynamics of opinion histories for a network population, taken as a whole, may reveal insights about the underlying system (structure, influence model, etc.). Several possible scenarios and system-level interpretations include:

- all agents begin with different opinions but converge to a shared opinion: the population has communicated and reached a consensus;

- all agents hold different and unchanging opinions: communication is not taking place, or agents do not trust their neighbors;

- opinions fluctuate, apparently at random: communication is taking place in a noisy environment, perhaps with rapidly changing external factors; and

- opinions converge to several different values across the network: members of smaller communities are reaching consensus with each other but not with neighboring communities.

An analyst might study the opinion histories of a social influence network and arrive at these kinds of insights and inferences [108], but such a manual process relies on the experience of the individual, making it difficult to reproduce or transfer to other situations. Accordingly, a framework for obtaining these results is desired, with some sort of measurement at its core.

We choose complexity measurement as the basis for such a framework. Specifically, we wish to use a measure of complexity on the opinion dynamics from social influence networks in order to capture insights about them, like those discussed previously. According to Sporns [109], "Differences in complexity among such related systems may reveal features of their organization that promote complexity." Social influence networks can display behavior that may be casually viewed as "complex," but this differs from the formal discipline of complexity studies. *Complexity* takes on various and often unrelated forms across different contexts, and unfortunately, the term is often used without definition or qualification, stripping it of much of its meaning [37, 44]. Nonetheless, complexity broadly captures elements of both randomness and order in a system's structure or behavior [44] and the related difficulty of reproducing or describing the system [110]. This tension between order and chaos is present in many models of social influence (and likely all real-world social networks) where noise or randomness exists, so complexity measurement is appropriate for the systems in our research context.

### 5.1.1   Objective and contributions

A complexity measure for the aggregated opinion dynamics of social influence networks has not been identified in the existing literature. The research objective of this paper is to develop one.

We approach complexity measurement through an objective-subjective framework [56] and begin by identifying a context-dependent notion of simplicity (opposed to complexity) for a set of opinion histories from a social influence network. This

leads to a baseline "reference simplicity" in the form of specific opinion histories. Next, we construct a function to measure the degree to which an observed data set deviates from the baseline. That function is our subjective complexity measure and is the main contribution of this paper. It is based on the location, velocity, variance, and entropy of the observed opinion data with respect to the baseline, as detailed in Section 5.3.3.

The subjective complexity measure is applied to two case studies. The first demonstrates the measure on simple, synthetic data sets. The second uses opinion data from previous work with social influence network simulation and explores the use of the complexity measure for inferring simulation inputs. Overall, the behavior of the subjective complexity measure broadly agrees with our notion of subjective complexity. The measure is sensitive to several of the simulation design parameters tested, suggesting that it may be useful for distinguishing different social network designs based on the agents' opinion data.

### 5.1.2 Overview

The remainder of this paper is organized as follows. In Section 5.2, we review relevant literature on the topic. Section 5.3 details our methods for developing a complexity measure for social influence networks. We evaluate the complexity measure through two case studies in Section 5.4 and discuss our findings in Section 5.5.

## 5.2 Background

In this section, we give a brief overview of complexity, remark on how noise in communication can affect the complexity of social influence data, and share a complexity measurement framework to combine these items together.

### 5.2.1 Complexity

Measuring complexity is difficult [44]. Some of that difficulty arises from the various ways in which complexity is defined and categorized in different contexts. Complexity measures can be categorized as either deterministic or statistical [37]. Deterministic complexity measures, such as Kolmogorov complexity, must "account for every bit—random or not—in an object" and become dominated by randomness, while statistical complexities (often based on entropy), "discount for randomness" and better capture the regularities of a system. We consider only statistical complexity in the current work, because the social influence networks under study feature stochastic elements and the research focus is on the behavior of systems rather than their precise reproduction.

Adding to the challenges of studying complexity is disagreement in what characterizes maximum complexity. There are two general approaches for defining the range of a complexity measure. One has complexity monotonically increase with increasing randomness or disorder of a system, while the other has complexity reach its minimum for both fully ordered and fully random systems and is maximized in-between [111]. In other words, the first approach treats complexity as a measure of disorder, while the second treats it as a measure of the mixture of order and randomness. For example, Shannon entropy is maximized for the maximally-random uniform distribution [43], but the physical complexity measure of Huberman and Hogg [112] is maximized for systems that are "intermediate between perfect order and complete disorder." The approach one chooses largely depends on whether one views complexity as a measure of randomness or as a measure of structure and information [109].

We adopt the former position, treating pure randomness as maximum complexity, in alignment with our focus on statistical complexity measures. With this approach, complexity monotonically increases with increasing randomness. This provides unambiguous comparisons between two systems—if system $x$ has a greater measured complexity than system $y$, then $x$ is more random. Also, this approach differentiates

purely random and purely ordered systems. In the current research context, this is appropriate, because different insights can be gleaned about systems at the two extremes. Based on this discussion, associating maximum complexity with purely random systems is a better fit for measuring the complexity of aggregated social influence network opinion histories.

### 5.2.2 Noise in communication

Communication is a noisy process. Since "the effects of noise, communication errors, and in general random fluctuations, should not be disregarded in any realistic representation of information spreading" [54], we must consider the presence of noise when studying opinion dynamics in social influence networks. Indeed, in complex social systems, the line separating signal and noise is "likely to be blurred" [44]. Techniques do exist to filter noise from input signals, but with noise filtering, there is a risk of removing genuine signal from the inputs. For example, white noise applied to the links between agents is an artifact of communication errors in the system and should be considered noise, while the random adoption influence model has agents engage in stochastic behavior yet represents actual behavior in the system [83]. In other words, a signal can appear noisy due to actual noise, chaotic inputs, or a combination of both.

For the analysis in Section 5.4, we primarily focus on unfiltered data (which may contain random noise) to avoid confounding the effects of noise filtering with the actual performance of our complexity measure. We illustrate these effects in Figure 5.4 by applying a 5-period moving average filter [113] to opinion history data prior to computing the complexity. Such a filter is able to produce good noise reduction results in general [113]; identifying an optimal filtering technique for opinion history data is outside the scope of this research.

### 5.2.3   Framework for complexity measurement

Network complexity measures typically center around structural properties of the network [46]. While these can address complexity associated with information channels, they do not account for actual information flow in the context of social influence. This requires an approach that can be sensitive to interactions taking place in the network.

Efatmaneshnik and Ryan [56] present a general framework for complexity built on the foundation that complexity has both objective elements and subjective elements. Objective complexity is based on the size or a description of the system and is independent of the observer. Subjective complexity is constructed relative to the observer and the observer's "reference simplicity." For example, a complexity measure for the structure of a network could relate the objective element to the number of nodes and edges in the network, while it could relate the subjective element to the network's departure from a complete graph (as the reference simplicity). This framework does assume that the observer can confidently evaluate what is "simple" in a given situation.

A criticism of this framework is its dependence on the observer's selection of the subjective element, which is focused on a specific application and may not be meaningful in different contexts. Another minor weakness is that for certain situations, the objective and subjective elements could be defined to be the same, in which case this framework provides no benefit over other approaches. However, the objective-subjective framework is sufficiently valuable in our current study of social influence networks and opinion dynamics that we accept these drawbacks and leverage the framework to construct a complexity measure in the next section.

We focus on the subjective elements of the framework in the current study. There are options for the objective complexity of networks, such as the number of nodes and edges, but measuring a network's objective complexity remains an open question in the field [114, 115], is highly sensitive to changes in network structure [109], and is

not part of the current research focus. Therefore, we omit objective complexity from the remainder of this paper.

## 5.3 Methodology

This section defines our methods for measuring the complexity of dynamic opinion data from social influence networks. It proceeds in three parts. First, we present general definitions used throughout this section. Next, we construct the reference system against which complexity is to be measured; this is based on our subjective interpretation of simple and complex systems as they relate to opinion dynamics in social influence networks. Finally, we define the dissimilarity measure to quantify how an observed set of opinion data deviates from the reference system, which becomes the measure of complexity used in Section 5.4.

### 5.3.1 General

A social influence network is a collection of individual agents connected with and able to influence one another via a network. The network is defined on the graph $G = (V, E)$ with each agent $i \in V$ connected to its neighbors via weighted directional edges in the set $E$. The agent population is of size $|V| = N$. Each agent has an opinion, the "agent's property that is affected by social influence in a model" [53], which encodes their attitude or position on some topic at a given time. Agent $i$'s opinion at time $t$ is expressed as $o(i, t)$ and is defined on the continuous interval $o(i, t) \in [-1, 1]$ to allow for two extreme positions, a neutral opinion at zero, and a spectrum of intermediate opinions in between. Opinions can change over time due to interactions among agents and random factors (e.g., noise), but external influence (e.g., mass media) is not a part of this study. Interactions are governed by an influence model and take place only along network edges. These interactions can be affected by random noise to model the imperfect nature of communication.

We call the sequence of an agent's opinion as it changes over time its *opinion history*. The term *system* refers to the entire set of processes and structures involved in creating opinion data: network, influence model, agent characteristics, etc. As an analogy, if we consider the opinion histories to be a set of signals, then the system is a signal generator.

### 5.3.2 Reference simplicity

Complexity analysis must be done "within the context of an appropriate framework" or else risk the production of "misleading or useless results" [44]. The objective-subjective framework from Efatmaneshnik and Ryan [56], detailed in Section 5.2.3, splits complexity into objective and subjective components; the objective component measures the size of the system or its minimum description, and the subjective component captures the system's deviation from a context-dependent reference model. The reference model is called the *reference simplicity* and is based on a *subjective simplicity*: "an idea of what is simple in a given context."

Our research context is opinion dynamics, and we posit that social influence networks that have agents with unchanging opinions are very simple. Simpler yet are networks where all agents share the *same* unchanging opinion. Lastly, a population sharing the same neutral opinion is less complex than one sharing an extreme opinion (a neutral opinion is effectively the same as no opinion, which intuitively has no complexity at all). From this subjective simplicity, the reference simplicity is defined as a set of opinion histories with values equal to zero for all agents at all times; zero is chosen as a neutral opinion because we define opinion to lie in the interval $[-1, 1]$. Formally, the reference simplicity contains a reference opinion for agent $i$ at time $t$ of

$$o_{ref}(i, t) = 0 \quad \forall i \in V, t \in \mathbb{Z}^+. \tag{5.1}$$

This approach does ignore the connections between agents (i.e., network structure), but we argue that the *effects* of those connections are present in the opinion histories: opinions change over time due to agent interactions along network edges, so

the time series of an agent's opinion depends (in part) on the network structure. If the network structure were changed, agent opinions could change. Alternative reference systems are discussed in Section 5.5.

Using this reference simplicity as a baseline, we expect the measured complexity to depend on the value of observed individual opinions (e.g., opinions far from neutral are more complex than those near neutral), the velocity of an agent's opinion (e.g., rapidly changing opinions are more complex than stable ones), and the distribution of opinions (e.g., populations with high variance in their opinions contain more complexity than those that are closely grouped). In the next section, we present a dissimilarity measure based on these expectations.

### 5.3.3   Dissimilarity measure

The reference simplicity serves as the baseline or zero-point of the complexity measure. Next, we must define the dissimilarity measure for quantifying how much a target system (observation) deviates from the simplicity. Given our construction of the reference simplicity as a set of equal, straight lines with value zero, we identify four key characteristics describing how an observed system deviates from the reference simplicity: location, dynamics, density, and disorder (Figure 5.1). Location refers to the distance that opinion histories lie from their reference lines (which need not be measured with a true distance metric), dynamics describes the rate of change of opinions at a point in time, density refers to the spread of opinions across the whole population, and disorder captures the uniformity of their distribution.

We measure location using the $L^1$-norm between observed opinion histories and the reference simplicity.[1] The dynamics term is a function of the change in opinion between consecutive, discrete time steps. Density is captured with the variance of opinion values. Disorder is a function of the Shannon entropy of discretized opinion values. These are straightforward approaches that philosophically capture our

---

[1] The $L^1$-norm is chosen instead of the more common $L^2$-norm based on results from Aggarwal et al. [116], which found that in high-dimensional spaces, the $L^1$-norm is more well-behaved.

Figure 5.1. The four characteristics that are the basis of our dissimilarity measure are shown here at lower and higher values in each half of the subplots. Each line segment represents an individual opinion history, showing opinion on the vertical axis and time on the horizontal; zero opinion is the reference simplicity. For example, the slowly changing opinion on the left of the dynamics plot represents a low value of the dynamics element, while the right side shows rapid change and a higher dynamics value. Most line segments are shown with constant opinion to emphasize a single dissimilarity characteristic.

identified characteristics of dissimilarity. It is reasonable to consider both variance and entropy, as they capture slightly different aspects of the distribution of data: "Variance is a measure of spread from the mean of a distribution, whereas entropy measures the spread across the distribution" [44]. The dissimilarity measure is defined as the sum of the four elements, detailed below. Each element is normalized, either by its theoretical maximum value for the problem context or by the maximum observed value for a given data set, in order to make the term dimensionless and equally weighted in the sum. This does cause each element to be bounded to the interval $[0, 1]$ as a coincidence, not a deliberate choice of scaling.

*1. Location.* To measure the location at time $t$, we treat both the observed and reference opinions as vectors, defining $\overrightarrow{o}(t) = (o(1, t), \ldots, o(N, t))$ and $\overrightarrow{o_{ref}}(t) = \overrightarrow{0}_N$, respectively.[2] Then we compute the $L^1$-norm on the difference of the two vectors:

$$|\overrightarrow{o}(t) - \overrightarrow{o_{ref}}(t)|_1 = \sum_{i=1}^{N} |o(i, t) - 0|. \tag{5.2}$$

Subjective complexity should not simply scale with the size of the network—that is the role of *objective* complexity—but Equation 5.2 does (at least for our choice of reference simplicity). To correct this, we divide Equation 5.2 by the maximum possible $L^1$-norm for a given population size $N$, based on the reference simplicity and the possible range of opinion values. The maximum $L^1$-norm for the current situation happens to be $N$, since opinion is bounded to $[-1, 1]$ and the reference value is zero, making the maximum distance per agent equal to 1. Therefore, the location element for our dissimilarity measure for time step $t$ is

$$\text{Location}(t) = \frac{\sum_{i=1}^{N} |o(i, t)|}{N}. \tag{5.3}$$

*2. Dynamics.* This element of the dissimilarity measure addresses the velocity of opinions. In agent-based systems, complexity measures should "reflect increasing complexity with increasing activity of agents" [38]. For each agent, we compute

---

[2]While we chose zero as the reference simplicity, this approach generalizes for any choice of reference, as $\overrightarrow{o_{ref}}(t) = (o_{ref}(1, t), \ldots, o_{ref}(N, t))$.

the absolute difference between two consecutive time steps and normalize by the maximum observed absolute difference across the population for a given time:

$$\Delta(i,t) = \frac{|o(i,t+1) - o(i,t)|}{\max_j |o(j,t+1) - o(j,t)|}. \tag{5.4}$$

If the maximum is zero, then $\Delta(i,t)$ is set to zero. The absolute value is used because we do not consider the sign of the difference to be relevant to the reference simplicity—in our perspective, a negative opinion is no more complex than a positive opinion. Then, $\Delta(i,t)$ is averaged across the population to yield

$$\text{Dynamics}(t) = \frac{1}{N} \sum_{i=1}^{N} \Delta(i,t). \tag{5.5}$$

*3. Density.* The density element is the variance of the opinion history data for some time $t$ divided by the maximum possible variance, based on the allowed bounds of opinion values. In general, for data $X$ with upper bound $b$ and lower bound $a$, the population variance $\text{Var}(X)$ is bounded by

$$0 \leq \text{Var}(X) \leq \left(\frac{b-a}{2}\right)^2. \tag{5.6}$$

For data on the interval $[-1, 1]$, variance has an upper bound of 1. Thus, for the set of opinion values at a single time step, $\overrightarrow{o}(t)$, the density element of the dissimilarity measure is

$$\text{Density}(t) = \frac{\text{Var}\left(\overrightarrow{o}(t)\right)}{1}. \tag{5.7}$$

*4. Disorder.* The disorder element is measured by normalizing the Shannon entropy of the discretized opinion values. Shannon entropy can be regarded as a measure of the information content of a distribution and is defined by Shannon [43] as

$$H(X) = - \sum_s p(x_s) \log_2 p(x_s), \tag{5.8}$$

where $p(x_s)$ is the probability of finding variable $X$ in state $s$. Adapting this to the current research context, $p(x_s)$ is the probability of an agent's opinion occupying state $s$, but states are discrete and opinion is continuous. So, we map each of the

continuous opinion values into one of twenty discrete bins of equal width, which form the set of states.[3] Then, the entropy $H(X)$ is computed for each time step based on the empirical probabilities associated with each bin: $p(x_s)$ is the fraction of agents with opinions that map to bin $s$. To normalize entropy (as is done for the other elements), we divide $H(X)$ by the maximum possible entropy. Entropy is maximized when data is uniformly distributed (i.e., each of the 20 bins has equal probability), so

$$0 \leq H(X) \leq -\sum_{s=1}^{20} \frac{1}{20} \log_2 \frac{1}{20} = \log_2 20. \tag{5.9}$$

Therefore, the normalized entropy that is the disorder element of our subjective complexity is

$$\text{Disorder}(t) = \frac{H(X)}{\log_2 20}. \tag{5.10}$$

The sum of Equations 5.3, 5.5, 5.7, and 5.10 defines the dissimilarity measure for an observed set of opinion histories, relative to a reference simplicity:

$$\begin{aligned} \text{Dissimilarity}(t) &= \text{Location}(t) + \text{Dynamics}(t) + \text{Density}(t) + \text{Disorder}(t) \\ &= \frac{\sum_{i=1}^{N} |o(i,t)|}{N} + \frac{\sum_{i=1}^{N} \Delta(i,t)}{N} + \text{Var}\left(\vec{o}(t)\right) + \frac{H(X)}{\log_2 20}. \end{aligned} \tag{5.11}$$

Dissimilarity is bounded to the interval $[0, 4]$ by construction, but the achievable maximum may be lower due to competition between variance and entropy (e.g., maximum variance occurs for data evenly split into two extremes, which yields very low entropy due to two occupied states). The dissimilarity value is dimensionless and can be scaled as desired for comparing with other complexity measures. This equation becomes the measure of subjective complexity applied to opinion history data in Section 5.4.

## 5.4    Case studies

The subjective complexity measure developed in Section 5.3.3 was applied to opinion history data from two case studies. (All uses of "complexity measure" in this

---

[3]The specification of bins can affect the outcome, but many approaches for "optimal" binning exist. Different sets of bins are not explored in the current study.

section refer only to the subjective complexity; objective complexity is not included in this analysis, as discussed in Section 5.2.3.) In the first case study, notional data was created based on the illustrative examples in Section 5.1 to show different values along a subjective range of complex systems. This demonstrated that the complexity measure shows the expected behavior in simple data sets. In the second case study, simulation data from prior work representing 39 instances of social influence networks was repurposed for use with the new measure. This exercises the measure on more realistic opinion history data where the simulation inputs could be used to anticipate changes in complexity. Complexity was computed at multiple time steps over the length of the opinion histories, producing a new time series of complexity values. The results showed that the performance of the complexity measure generally aligns with our notion of subjective complexity and that the measure may have potential as a tool for differentiating different social network designs.

## 5.4.1 Case study 1: notional systems for illustration

For this case study, we constructed six notional data sets that represent opinion histories inspired by scenarios from the introduction (Figure 5.2). These scenarios help demonstrate that the complexity measure performs as expected for several simple systems and are described as follows:

1. static and equal: agents have equal, unchanging opinions;

2. static but unequal: agents have different and unchanging opinions;

3. uniform random: opinions are set uniformly at random for all time steps, with no dependence on the previous time step;

4. single convergent: agents begin with different opinions but converge to a shared opinion;

5. multiple convergent: opinions converge to several different values across the population; and

6. divergent: agents begin with one of several opinions before diverging into many different opinions (the multiple convergent scenario, reversed).
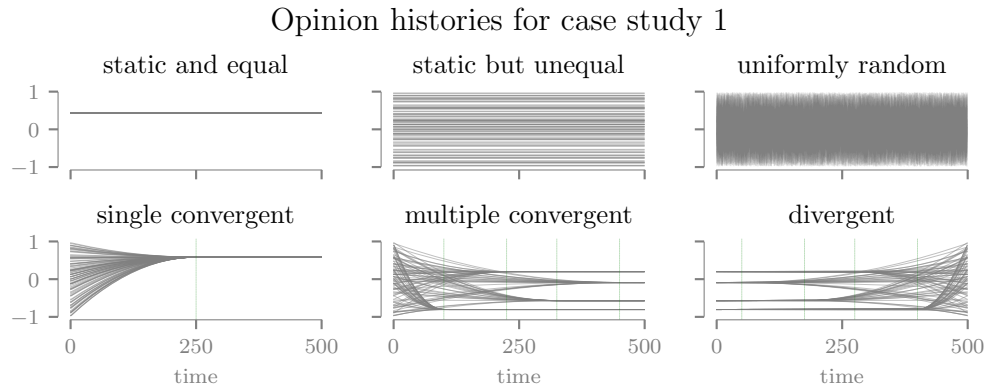
Opinion histories for case study 1



Figure 5.2. The opinion histories represented by the six notional data sets in case study 1. For each set of axes, time is on the horizontal axis and opinion is on the vertical. For the bottom three scenarios, vertical grid lines mark convergence/divergence times. The measured subjective complexity values for each of these data sets are presented in Figure 5.3.

As a basis for assessing the complexity measure, we made predictions for each of the simple scenarios (Table 5.1). The static and equal scenario should have constant complexity equal to the fixed opinion value of the population, with only the location term contributing to complexity: the dynamics term is zero (all opinion histories have zero slope), the density term is zero (each observation is equal to the mean), and the disorder term is zero (the entropy of all data in a single bin is zero). For the static but unequal case, the complexity should be constant at $\frac{1}{2} + 0 + \frac{1}{3} + 1 = 1.83$: with the opinion data distributed uniformly in the interval $[-1, 1]$, the mean of the absolute value of the data (location) is approximately $1/2$, all slopes are zero (dynamics), the variance is $\frac{1}{12}(1 - -1)^2 = \frac{1}{3}$ (density), and the entropy is maximized (disorder equals 1).

The uniform random data is the same as the previous case, except each time step is independent, so the slopes between time steps may not be zero, leading to a

small positive expected value for the dynamics term. For this data set, the slopes are distributed according to a triangular distribution[4] with expected value 2/3, and from a Monte Carlo simulation, the expected maximum slope for such opinion data is approximately 1.82. This gives an average dynamics term around 0.36 and a predicted complexity of $1.83 + 0.36 = 2.2$. The data in the remaining scenarios is too diverse to afford simple proofs and numerical predictions, but our qualitative estimates appear in Table 5.1.

Table 5.1.

Predictions about the complexity of each system for case study 1 are compared to the result of computing the complexity measure for the data. For most scenarios, the predicted and observed behaviors closely agree (✓). The disagreements (X) were likely due to failure to predict the impact of random number generation variability on the small sample size ($N = 100$).

| Scenario | Predicted | Observed |
|---|---|---|
| static and equal | constant at fixed opinion (0.44) | constant at 0.44 ✓ |
| static but unequal | constant near 1.83 | constant at 1.72 X |
| uniform random | constant near 2.2 | noisy but flat near 2.14 X |
| single convergent | moderate, trending lower | matches ✓ |
| multiple convergent | moderate, trending slightly lower | matches ✓ |
| divergent | reverse of multiple convergent | matches ✓ |

**Results**

The measured complexity for each scenario is graphed in Figure 5.3. Expected behavior of the complexity measure time series for each of the scenarios is compared to the observed behavior in Table 5.1. In most cases, the desired and observed behaviors

---

[4]The distribution of the absolute difference of two uniformly distributed random variables is triangular. For $U(-1, 1)$ variables, the triangular distribution has minimum 0, maximum 2, and mode 0.

closely agreed. For example, we predicted the complexity for the single convergent scenario to be "moderate, trending lower," and the result was a complexity time series that began at a moderate level (near 2), gradually dropped to a low value (near 0.5) as the opinions converged (at time $t = 250$), then remained constant.[5] We did expect a greater range of complexity for the multiple convergent scenario, in the form of larger decreases (relative to the other fluctuations in complexity) as each group converged in opinion.



Figure 5.3. The complexity measure defined in Section 5.3 is computed for on the opinion history data for the six scenarios in the first case study. This measure is bounded to the interval $[0, 3]$. For the last three scenarios, the vertical grid lines correspond to the convergence/divergence times as in Figure 5.2.

The disagreements between expected and observed behavior for this case study were the static but unequal and the uniform random scenarios. For the former, we correctly predicted a constant value but erred by 0.1 units in the magnitude. This difference is likely due to uneven distribution of the opinion data (Figure 5.2) and small sample size ($N = 100$). For the latter, the magnitude differed from prediction

[5]The complexity for the single convergent scenario reached its minimum slightly before the opinion values actually converged, which is likely due to the bin size for the entropy calculation—once all opinions fit in one bin, they are indistinct with respect to entropy.

by about 0.06 units and the complexity time series was noisy instead of smooth, a difference likely caused by variability from the random number generation.

In Section 5.2.2, we discussed the impact of noise on social influence networks, but the results presented so far have used unfiltered opinion history data. We analyzed the raw data so as not to confound the effect of noise filtering with the genuine behavior of our complexity measure. For Figure 5.4, we measure the complexity of filtered data to highlight the impact of noise filtering. When using a noise filter, the uniformly random data has lower complexity than the static but unequal scenario, which at first appears to violate our intuition of subjective complexity. On further consideration, however, this contrast reinforces our understanding: purely random data contains no real signal, so any apparent complexity that is a result of noise should be discounted.



Figure 5.4. The complexity measurements for this cases study were repeated after filtering the opinion history data using a 5-period moving average. The complexity measure is designed with the goal of being maximized for random data, but filtering makes the data more similar to static data with lower variance, driving down the complexity.

In this first case study, the complexity measure was shown to perform mostly as expected for a set of simple scenarios (four of six predictions matched). This helps to build confidence in and intuition about the measure, supporting the analysis of more realistic opinion data in the next case study.

### 5.4.2   Case study 2: simulation data from prior work

The second case study evaluated opinion history data from Garee and Ventresca [101], which created opinion histories for a wide range of simulated social influence networks by varying five different simulation and network design factors from the literature (population size, network structure model, agent activation regime, influence model, and influence error distribution) for each simulation trial. While still artificial, that data is more representative of real-world influence networks than the simpler data sets in the first case study. We selected 13 trials from their experimental design, chosen to allow comparisons between trials that intuitively should have different degrees of complexity.[6] For example, given two trials that differ in their design only by the presence of random noise in the links between agents, it is reasonable to expect the noisy trial to have greater measured complexity than the noise-free trial. Single replications are used in this analysis for illustrative purposes; a more comprehensive analysis of the complexity measure would use all available data.

For each of the five original design factors listed previously, we chose a distinct set of trials that differ only by the setting (level) of that design factor (i.e., one-way variation). Specifically, for each design factor, we randomly selected a trial from among all trials in the reduced data set. Then, we (non-randomly) selected the trials that differed from the randomly selected trial in only the given design factor. For example, if the randomly selected trial for population size has a population size of 100, we then manually select the trial with the same settings except with population size 10,000. A different random trial was chosen for each factor. The selected settings for each design factor and our predictions about the relative complexity measurements (within each factor) are described as follows:

---

[6]In Garee and Ventresca [101], each trial is replicated 100 times using different random number seeds and then aggregated by that study's response variables. Here, we select single replications for the chosen trials.

- network structure model—Erdős-Rényi random graph, small world, or preferential attachment: these are ordered from greatest to least predicted complexity, based on the relative degrees of structure versus randomness in the models;

- agent activation regime—random, uniform, or synchronous: ordered from greatest to least predicted complexity, based on the degree of randomness in scheduling within each regime;

- influence model—random adoption, attractive-repulsive, or standard model: these are ordered from greatest to least predicted complexity, since the first uses a random selection among neighbors, the second may move opinion closer to *or* further from neighboring opinions, and the third averages the opinion values among neighbors;

- influence error term—no error or normally-distributed zero-mean error with variance 0.1: trials with no error term should have lower complexity than those with an error term; and

- population size—100 or 10,000 agents: negligible difference is predicted, since network size is an element of the objective, rather than subjective, complexity.

We classify as *impactful* the factors where we expect to see distinct behavior in the complexity time series for the different settings. In this case, we predict all factors except population size to be impactful (Table 5.2).

**Results**

The complexity for each trial was calculated on its (unfiltered) opinion history using the measure defined in the previous section, and the time series for several of the factor-level comparisons are presented in Figures 5.5-5.9. We performed a visual assessment to classify each factor as impactful or not, which we compare with our predictions in Table 5.2 and detail below. Overall, two of five factors were impactful,

Table 5.2.
For three of the network and simulation design factors, our predicted and observed impact assessments closely agreed (✓). The other two factors showed disagreement (X).

| | Impact on complexity | |
| --- | --- | --- |
| Factor | Predicted | Observed |
| network structure model | impactful | not impactful X |
| agent activation regime | impactful | not impactful X |
| influence model | impactful | impactful ✓ |
| influence error term | impactful | impactful ✓ |
| population size | not impactful | not impactful ✓ |

and the results agreed with our predictions for three factors. Note that the results from Table 5.2 do not necessarily mean that network structure, activation regime, and population size have *no impact* on the complexity measure, merely that they have no apparent impact as main effects as measured using our qualitative visual analysis. Interaction effects may exist and are an avenue for future exploration.

The three trials for network structure model produced erratic and frequently intersecting time series for the complexity measurement (Figure 5.5), although the small-world complexity has about half the range and less average complexity than the other two. But with such little visual distinction between the three trials, the network model is not impactful; this disagrees with our prediction. For agent activation regime, all three trials showed noisy but rather flat complexity plots with small vertical ranges of approximately 0.15 units (Figure 5.6). The lines intersect in many places and lack any visual distinction, so we conclude activation regime to be not impactful, also in disagreement with our prediction.

On the other hand, we correctly predicted influence model to be impactful to complexity (Figure 5.7). The three influence models had clear differences in complexity: the plot for random adoption moves erratically, while the other two models were flat with minimal noise and large separation (approximately 0.8 units). However, the attractive-repulsive model was more complex than the random adoption, likely due to the effect of polarizing opinions on the density (variance) element of the complexity measure. The influence error distribution factor behaved as expected and is impactful, showing clear separation (0.2 units) between the two settings (Figure 5.8). Finally, changes in population led to little difference in complexity (Figure 5.9). The plot for the larger population is smooth, while that for the smaller population is very erratic, but both are flat and within about 0.05 units of each other on average. Population's classification as not impactful agrees with our prediction.

Thus far, we have considered only a single set of trials for each factor. To expand our study, the complexity measurement process was repeated using data from a different replication of the same trials (i.e., simulation data via a different random

Complexity time series for case study 2, network model factor



Figure 5.5. The high degree of overlap of complexity time series for the three network structure models suggests this factor is not impactful to the complexity measure.

Complexity time series for case study 2, activation regime factor



Figure 5.6. The three agent activation regimes show little distinction in complexity, leading to an assessment of not impactful for this factor.

Complexity time series for case study 2, error distribution factor



Figure 5.8. Unlike the random adoption influence model, the influence error distribution directly adds noise to the opinion histories. Thus the complexity of the trial with no added noise is lower than the trial with added noise, as predicted.

Figure 5.9. The two trials that differed only in the population size, 100 versus 10,000, produced complexity time series with similar average values. This result agrees with our prediction that population size is not impactful for the complexity measure.

number seed), and then again using a different set of trials, selected to support the same comparisons between design factor settings. Results for these two additional data sets differed from the original results in terms of their measured complexity values but not with respect to their qualitative behavior and our impact assessments.

In this case study, we applied our complexity measure to opinion history data from a number of simulated social influence networks. Only the choice of influence model and influence error distribution were found to be impactful to the complexity. We expected population size to not be impactful based on our notion of subjective complexity. The lack of impact for the network structure model and agent activation regime was unexpected; this is discussed further in the next section.

## 5.5   Conclusion

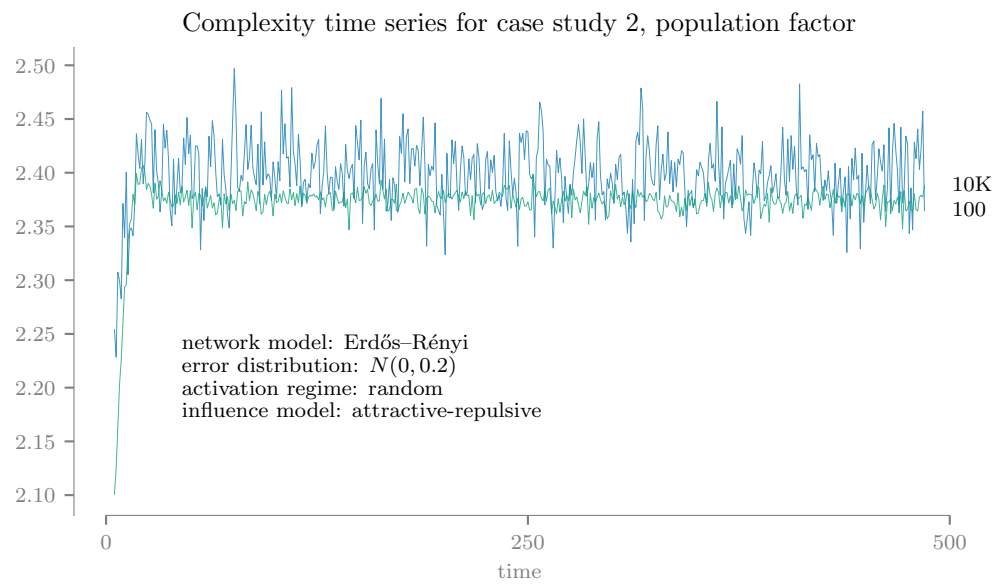We conclude this paper with a discussion of the major findings, before covering alternative reference systems, outlining some limitations of our results, and remarking on opportunities for future work.

In this paper, we developed a subjective complexity measure for analyzing opinion time series data from social influence networks. The measure performed well in the two case studies. Overall, the results showed that its performance generally aligns with our notion of subjective complexity and that the measure may be a viable tool for distinguishing different social network designs. We observed the complexity measure to be sensitive to the selection of influence model and influence error term, in that the different settings of those parameters produced visually distinct behavior in time series plots of the complexity. This in particular indicates that the measure could be useful to studies that try to infer those parameters from opinion data.

The objective-subjective complexity framework [56] at the heart of development process was valuable. Its notion of subjective simplicity made creating a reference system an intuitive process and helped scope the complexity measure. However, it

does not aid in the actual definition of the measure, which remains an exploratory activity.

The results agreed with our predictions in four of six scenarios in case study 1 and three of five in case study 2. The discrepancies in the first case study are due to minor errors in our prediction process. In the second case study, it was surprising to observe such little impact on complexity from the selection of network structure model, which may weaken our claim that the effect of the network structure is present in the opinion histories. The actual network structures from the different trials may have been quite similar to begin with, or the settings for the other experimental factors may be dominating the complexity measure.

### 5.5.1   Alternative reference systems

In this study, we selected a constant, neutral opinion as the reference simplicity, where zero is the neutral value for opinions bounded to $[-1, 1]$. Below, we identify several alternative approaches.

First, reference values could be set to the average opinion across the network. In such a case, populations with a narrow range of opinions have low complexity, regardless of their magnitude: a population with opinions concentrated at an extreme is no more complex than one with neutral opinions. This does not match our subjective notion of complexity but is a feasible approach.

We have assumed in this work that network structure is implicit in opinion histories, but if incorrect, an alternative reference system could explicitly include the network structure with the reference opinions. This does limit use of the complexity measure to social influence networks where the full structure is known, which can be difficult in real-world situations (such as an application where users record their preferences but not their sources of influence or their interactions with other users). The dissimilarity measure presented in this paper does not make explicit use of the

network structure, so the measure would need to be revised or replaced for it to operate directly on structural information.

A third alternative is a set of opinion histories generated at random. This exists at the opposite end of the order-chaos spectrum from a set of static opinions, so it could be a reasonable reference simplicity. However, it carries with it questions about what sampling distribution to use. It also causes the complexity measurement to be stochastic, since different initial randomization conditions would lead to different reference systems. For these reasons, a randomly generated reference system is undesirable.

### 5.5.2   Limitations and future work

Although the objective-subjective framework proved valuable here, its use leads to complexity measures that are highly context-dependent. Such measures cannot be viewed as general-purpose tools, so our complexity measure may only be valid in our defined scope.

Objective complexity for a network is often based on its size (e.g., number of nodes and edges). Subjective complexity ignores the network's size. Since our analysis did not use objective complexity, large and small populations had similar complexity. This may conflict with more common notions of larger systems as being naturally more complex. Also, different configurations of edges for the same population can have different objective complexities (sparse versus dense graphs), but our measure was not significantly impacted by changes in network structure.

The time over which the system is observed has a significant impact on the measured complexity. Referring to Figure 5.3 for example, a complexity value of approximately 1 could be associated with the single convergent system at time 200 (bottom left), the multiple convergent system at time 500 (bottom center), or the divergent system near time 0 (bottom right). If these systems were only observed near the specified times, we could conclude the systems are similar in complexity, despite the

different ways the systems evolve. Changes to the dissimilarity measure that could compensate for this effect would be desirable.

Interaction between the various system design factors in case study 2 may be confounding our results on which factors impact our complexity measure. A full experimental design using all simulation data from Garee and Ventresca [101] could clarify the relationships between the five design factors and the complexity of the opinion histories.

Our subjective complexity measure is the sum of four terms (location, dynamics, density, and disorder), weighted equally. The measure could be modified to weight the terms unequally to prioritize the complexity due to some characteristics over others. Further, Thompson and Young [44] observe that some researchers require multiple measures of complexity to study a system, while others focus on single summary measures. We developed a single complexity measure, but its four terms could be studied separately. A detailed analysis of the relationships among those terms may provide insights for refining the measure or for weighting the terms unequally.

# 6. CONCLUSION

The objective of this dissertation was to explore the degree to which opinion change in social influence networks could be used to determine characteristics of those networks. Complexity theory was the lens through which this exploration was conducted, so opinion data from simulated social influence networks was analyzed using several complexity measures (regression, several types of entropy, and a new measure). Overall, changing the inputs and specifications for a social influence network simulation can change how opinion evolves over time in the network; this in itself is not a particularly surprising result. What is noteworthy is that for some design factors, relationships exist that may allow researchers to infer some details of the inputs by analyzing only the outputs. These relationships are diverse: not all design factors affect opinion equally. The influence model and the distribution of communication errors—a factor often omitted from the literature—are consistently impactful, with their various settings producing distinct profiles in time series plots of the measurements. Activation regime is impactful to some entropy measures. Network structure has little impact on the new complexity measure developed in Chapter 5, and population size has little impact in general. These relationships, as well as the measures and problem-solving approaches used in this work, may be helpful to analysts working to infer the properties of real-world social influence networks from the opinion data those systems generate.

Results presented in this dissertation are not without limitations. Many of the detailed findings in this work are qualitative in nature, using the shapes or profiles of time series plots of the different response variables, and may not appear as rigorous as numerical methods (although comparison of profiles is an accepted approach in time series studies). This methodology adds a layer of subjectivity to the results and may invite disagreement from experts in the field. Also, all opinion data in this dissertation

was generated via agent-based simulation to ensure direct control of system design factors. Real-world social influence networks are likely much more sophisticated than those in the simulations, however, and this difference in sophistication may limit how well the broader findings can be applied to authentic social networks. Nonetheless, the conclusions presented throughout this document can serve as a foundation for further studies that expand the scope of the work performed herein.

## 6.1 Future perspectives

Several findings from this dissertation may help focus continued study of social influence networks. Since the overall findings demonstrate the potential for inferring network properties from dynamic opinion data, future efforts would benefit from a larger database of relationships between network inputs and outputs. Rather than waiting for the next research question to arise, we can set to work now developing large-scale social network databases to analyze for relationships. Then, as new questions emerge, researchers could rapidly leverage that database to help them understand their particular social networking context.

With respect to individual system design factors, activation regime and communication error should receive more consideration in the literature. Some theoretical results find that synchronous and random asynchronous activation regimes lead to little difference in qualitative behavior for some DeGroot-like models[1] and should therefore not be treated as an important factor. However, experimental findings here show differentiation among activation regimes in many situations, especially when using the more stochastic design elements (e.g., communication error or the random adoption influence model). DeGroot-like and other theoretical models of influence rarely include communication noise between agents, so it is understandable that activation regime is also often ignored, but both factors can be impactful to opinion

---

[1]Based on private correspondence from an anonymous journal reviewer.

dynamics and should be included (or omitted only with careful justification) in future research.

Related to the activation regime is the choice of discrete-event versus discrete-time. The discrete-time approach is used for all simulations in this dissertation, as is nearly universal for agent-based simulation. However, work by Buss and Al Rowaei [117] questions this approach, observing excessive penalties in output accuracy of discrete-time simulations, relative to their discrete-event counterparts. Future studies could model social influence networks using discrete-event methodology, which may give more accurate results. That would also better align with the intuitive understanding of social influence as a stochastic process: in this work, influence effectively flows at a constant rate through our simulated networks,[2] but the rate of real-world social interactions vary.

Additionally, nonhomogeneity among network populations is sometimes important (Chapter 4) and deserves closer scrutiny in future studies. At the level of individual trials, nonhomogeneity created drastic changes in system-level entropy as compared to a baseline homogeneous trial. Dividing interacting agents into distinct classes is a challenge for analytical models, but real-world social networks are not as homogeneous as such analytical models would imply. Skillful use of agent-based simulation can overcome this challenge and improve the degree to which analysis results can be applied to real-world scenarios.

Throughout this dissertation, we have observed diversity between system design factors and their importance to different response variables in a variety of circumstances. The diversity of these relationships emphasizes the value of using experimental design techniques for simulation-based investigations of social influence networks. We focused on the impact of individual design factors on the simulation outputs—this is a *main effects* approach. Using the same data, we can explore the impact of multiple factors at a time, measuring the *interaction effects*. These effects may

---

[2]Trials using the random activation regime have at least some agents update their opinions each time step, while all other trials have the entire population update at each time step.

provide new and interesting insights into the relationships between opinion dynamics and the inner workings of social influence networks.

Also, more emphasis should be placed on defining complexity measures within a rigorous *framework*, as was done with the new complexity measure in Chapter 5. Too often, existing measures are applied to new situations with limited consideration by the authors about what exactly that measure is describing. Entropy, for example, is widely adapted for use as a complexity measure, but authors rarely articulate what it actually describes about their research context, leaving open questions such as, "what does high/low entropy tell us about the system?" (This critique could reasonably be made about Chapter 3's entropy response variables.) Independent of the primary findings from applying the new complexity measure to opinion data, developing the measure within an established framework (i.e., measuring the difference between an observed complex system and a notional simple reference system) produced the most cogent interpretations of the three approaches to complexity measurement featured in this dissertation. Not only would more explicit use of frameworks enhance the understanding of individual complexity measures, it would benefit complexity theory at large by better communicating the aims of the work.

Each of us is a member of multiple social influence networks: friends, family, professional relationships, and more, that are all responsible for influencing our opinions on a multitude of topics. Improved understanding of the mechanisms that shape our opinions in these networks could reveal insights about ourselves and the world around us. This dissertation has endeavored to support that understanding by adding to our knowledge of the relationships between influence and opinion change in these complex systems, which could help us to be open to, or guarded against, influence from our network peers in the future.

REFERENCES

REFERENCES

[1] Francesco C. Billari, Alexia Prskawetz, Belinda Aparicio Diaz, and Thomas Fent. The "wedding-ring": An agent-based marriage model based on social interaction. *Demographic Research*, 17(3):59–82, 2007.

[2] Noah E. Friedkin and Eugene C. Johnsen. *Social Influence Network Theory*. Cambridge University Press, Cambridge, 2011.

[3] Peter-Paul van Maanen and Bob van der Vecht. An agent-based approach to modeling online social influence. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM '13*, pages 600–607, New York, New York, USA, 2013. ACM Press.

[4] Seth Lloyd. Measures of complexity: A nonexhaustive list. *IEEE Control Systems*, 21(4):7–8, 2001.

[5] R.E. Kalman. On the general theory of control systems. *IFAC Proceedings Volumes*, 1(1):491 – 502, 1960. ISSN 1474-6670. doi: https://doi. org/10.1016/S1474-6670(17)70094-8. URL http://www.sciencedirect.com/science/article/pii/S1474667017700948. 1st International IFAC Congress on Automatic and Remote Control, Moscow, USSR, 1960.

[6] Ted Meyer, Stephen C. Upton, Mary L. McDonald, and Christina Bouwens. Investigating social network analysis methods for identifying emergent behaviors in agent-based models, 2018.

[7] Barry Wellman and R. Hiscott. From social support to social network. *Research Paper - University of Toronto, Centre for Urban & Community Studies*, 146, 1983.

[8] John Scott and Peter Carrington. *The SAGE Handbook of Social Network Analysis*. SAGE, 2014.

[9] Maira Gatti, Paulo Cavalin, Samuel Barbosa Neto, Claudio Pinhanez, Cicero Dos Santos, Daniel Gribel, and Ana Paula Appel. Large-scale multi-agent-based modeling and simulation of microblogging-based online social network, 16113349.

[10] Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. Statistical properties of community structure in large social and information networks. In *Proceeding of the 17th international conference on World Wide Web - WWW '08*, page 695, 2008.

[11] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group formation in large social networks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*, page 44, New York, New York, USA, 2006. ACM Press.

[12] Okreša urić Bogdan, Igor Tomičić, and Markus Schatten. Towards agent-based simulation of emerging and large-scale social networks. examples of the migrant crisis and mmorpgs. *European Quarterly of Political Attitudes and Mentalities EQPAM*, 5(4):1–19, 2016.

[13] Patrick Wagstrom, Jim Herbsleb, and Kathleen Carley. A social network approach to free/open source software simulation. In *First International Conference on Open Source Systems*, pages 16–23, 2005.

[14] Ben Golub and Matthew O. Jackson. Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1): 112–149, 2010.

[15] Matthew O. Jackson. *Social and economic networks*. Princeton, NJ : Princeton University Press, Princeton, NJ, 2008.

[16] Tao Yang and Mario Ventresca. Constructing representative social networks for disease simulation, 2014.

[17] Lisa F. Berkman and Aditi Krishna. Social network epidemiology. In Lisa F. Berkman, Ichiro Kawachi, and M. Maria Glymour, editors, *Social Epidemiology*, chapter 7, pages 234–289. Oxford University Press, 2nd edition, 2014.

[18] Roberto C.S.N.P. Souza, Renato M. Assunção, Derick M. Oliveira, Daniel B. Neill, and Wagner Meira. Where did I get dengue? detecting spatial clusters of infection risk with social network data. *Spatial and Spatio-temporal Epidemiology*, 29:163–175, June 2019. ISSN 18775845.

[19] Peter Duggins. A psychologically-motivated model of opinion change with applications to american politics. *Jasss*, 20(1), 2017.

[20] Ben Golub and Evan Sadler. Learning in social networks. In Yann Bramoullé, Andrea Galeotti, and Brian Rogers, editors, *Oxford Handbook on the Economics of Networks*, volume 1. Oxford University Press, 2016.

[21] Jure Leskovec, Lada A Adamic, and Bernardo A Huberman. The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1), 2007.

[22] Kathleen M Carley. Dynamic network analysis for counter-terrorism, 2005.

[23] Sebastiano A. Delre, Wander Jager, Tammo H.A. Bijmolt, and Marco A Janssen. Will it spread or not? the effects of social influences and network topology on innovation diffusion. *Journal of Product Innovation Management*, 27(2):267–282, 2010. ISSN 07376782.

[24] H. Peyton Young. Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review*, 99(5): 1899–1924, 2009.

[25] Mei Li, Xiang Wang, Kai Gao, and Shanshan Zhang. A survey on information diffusion in online social networks: Models and methods. *Information*, 8(4): 118, 2017. ISSN 2078-2489.

[26] Lilian Weng, Filippo Menczer, and Yong Yeol Ahn. Virality prediction and community structure in social networks. *Scientific Reports*, 3(1):2522, December 2013.

[27] Karthik Subbian, B. Aditya Prakash, and Lada A. Adamic. Detecting large reshare cascades in social networks. *Proceedings of the 26th International Conference on World Wide Web - WWW '17*, pages 597–605, 2017.

[28] Noah E. Friedkin. *A Structural Theory of Social Influence*. Structural Analysis in the Social Sciences. Cambridge University Press, 1998.

[29] Morris H DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974. ISSN 0162-1459.

[30] Pavlin Mavrodiev, Claudio J. Tessone, and Frank Schweitzer. Quantifying the effects of social influence. *Scientific Reports*, 3(1):1360, December 2013. ISSN 2045-2322.

[31] Saike He, Xiaolong Zheng, Daniel Zeng, Kainan Cui, Zhu Zhang, and Chuan Luo. Identifying peer influence in online social networks using transfer entropy. In *Lecture Notes in Computer Science*, volume 8039, pages 47–61. Springer, Berlin, Heidelberg, 2013.

[32] Elizabeth Bruch and Jon Atwell. Agent-based models in empirical social research. *Sociological Methods and Research*, 44(2):186–221, May 2015. ISSN 15528294.

[33] Angelico Giovanni Fetta. *Investigating Social Networks with Agent Based Simulation and Link Prediction Methods*. PhD thesis, Cardiff University, 2014.

[34] E Franchi and M Tomaiuolo. A unified framework for traditional and agent-based social network modeling. *Interdisciplinary Applications of Agent-Based Social Simulation and Modeling*, 2014.

[35] Julie Dugdale, Narjès Bellamine Ben Saoud, Fedia Zouai, and Bernard Pavard. Coupling agent based simulation with dynamic networks analysis to study the emergence of mutual knowledge as a percolation phenomenon. *Journal of Systems Science and Complexity*, 29(5):1358–1381, 2016.

[36] Wai Kin Victor Chan. Agent-based and regression models of social influence. In *2017 Winter Simulation Conference (WSC)*, pages 1395–1406. IEEE, 2017.

[37] David P. Feldman and James P. Crutchfield. Measures of statistical complexity: Why? *Physics Letters A*, 238(4-5):244–252, February 1998.

[38] Jason B Clark and David R Jacques. Practical measurement of complexity in dynamic systems. In *Procedia Computer Science*, volume 8, pages 14–21, 2012.

[39] James Ladyman, James Lambert, and Karoline Wiesner. What is a complex system? *European Journal for Philosophy of Science*, 3(1):33–67, January 2013.

[40] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *International Journal of Computer Mathematics*, 2(1-4):157–168, 1968.

[41] Donald E Knuth. Big omicron and big omega and big theta. *ACM SIGACT News*, 8(2):18–24, 1976. ISSN 01635700.

[42] David H. Wolpert and William Macready. Using self-dissimilarity to quantify complexity. *Complexity*, 12(3):77–85, 2007. ISSN 10762787.

[43] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 1948. ISSN 15387305.

[44] Michael Thompson and Louise Young. The complexities of measuring complexity, 2012.

[45] Carter T. Butts. The complexity of social networks: Theoretical and empirical findings. *Social Networks*, 23(1):31–71, January 2001.

[46] Abbe Mowshowitz and Matthias Dehmer. Entropy and the complexity of graphs revisited. *Entropy*, 14(3):559–570, March 2012. ISSN 10994300.

[47] Greg Ver Steeg and Aram Galstyan. Information-theoretic measures of influence based on content dynamics. *Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13*, page 3, August 2012.

[48] Marcin Kulisiewicz, Przemysław Kazienko, Boleslaw K. Szymanski, and Radosław Michalski. Entropy measures of human communication dynamics. *Scientific Reports*, 8(1):15697, 2018.

[49] Wai Kin Victor Chan, Young Jun Son, and Charles M. Macal. Agent-based simulation tutorial - simulation of emergent behavior and differences between agent-based simulation and discrete-event simulation. In *Proceedings - Winter Simulation Conference*, pages 135–150, 2010.

[50] Susan M. Sanchez and Hong Wan. Work smarter, not harder: A tutorial on designing and conducting simulation experiments. In *2015 Winter Simulation Conference (WSC)*, pages 1795–1809. IEEE, 2015.

[51] Frederic Amblard, Audren Bouadjio-Boulic, Carlos Sureda Gutierrez, and Benoit Gaudou. Which models are used in social simulation to generate social networks? a review of 17 years of publications in jasss. In *2015 Winter Simulation Conference (WSC)*, pages 4021–4032. IEEE, 2015.

[52] Kenneth W. Comer and Andrew G. Loerch. The impact of agent activation on population behavior in an agent-based model of civil revolt. In *Procedia Computer Science*, volume 20, pages 183–188. Elsevier, January 2013.

[53] Andreas Flache, Michael Mäs, Thomas Feliciani, Edmund Chattoe-Brown, Guillaume Deffuant, Sylvie Huet, and Jan Lorenz. Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4):2, October 2017. ISSN 1460-7425.

[54] M. F. Laguna, Guillermo Abramson, and Damián H. Zanette. Vector opinion dynamics in a model for social influence. *Physica A: Statistical Mechanics and its Applications*, 329(3-4):459–472, 2003.

[55] Haili Liang, Housheng Su, Ying Wang, Chen Peng, Minrui Fei, and Xiaofan Wang. Continuous-time opinion dynamics with stochastic multiplicative noises. *IEEE Transactions on Circuits and Systems II: Express Briefs*, pages 1–1, 2018. ISSN 1549-7747.

[56] Mahmoud Efatmaneshnik and Michael J. Ryan. A general framework for measuring system complexity. *Complexity*, 21:533–546, September 2016. ISSN 10990526.

[57] B Touri and A Nedic. Distributed consensus over network with noisy links. In *12th International Conference on Information Fusion*, pages 146–154, Seattle, WA, 2009.

[58] Roger Th.A.J. Leenders. Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24(1):21–47, 2002. ISSN 0378-8733.

[59] L. Rendell, R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M. W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap, and K. N. Laland. Why copy others? insights from the social learning strategies tournament, 2010.

[60] Abhijit Banerjee, Emily Breza, Arun G. Chandrasekhar, and Markus Mobius. Naive learning with uninformed agents. *Working paper*, page 27, 2016.

[61] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011. ISSN 0034-6527.

[62] Arun G. Chandrasekhar, Horacio Larreguy, and Juan Pablo Xandri. Testing models of social learning on networks: Evidence from a lab experiment in the field, 2015.

[63] Shu-li Cheng, Wen-hsien Lin, Frederick Kin Hing Phoa, Jing-shiang Hwang, and Wei-chung Liu. Analysing the unequal effects of positive and negative information on the behaviour of users of a taiwanese on-line bulletin board. *PLOS ONE*, 10(9):e0137842, 2015.

[64] David Masad and Jacqueline Kazil. Mesa: An agent-based modeling framework index termsagent-based modeling, multi-agent systems. In *14th PYTHON IN SCIENCE CONFERENCE*, pages 53–60, 2015.

[65] Aric A Hagberg, Daniel A Schult, and Pieter J Swart. Exploring network structure, dynamics, and function using networkx. In *7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA, 2008.

[66] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *9th PYTHON IN SCIENCE CONFERENCE*, pages 57–61, 2010.

[67] Michael J. Garee. Github repository for online appendix, 2018. URL `https://github.com/mgaree/wsc2018`.

[68] Susan M. Sanchez. Nolhdesigns spreadsheet, 2011. URL `https://nps.edu/web/seed/software-downloads`.

[69] Peng Jia, Anahita MirTabatabaei, Noah E. Friedkin, and Francesco Bullo. Opinion dynamics and the evolution of social power in influence networks. *SIAM Review*, 57(3):367–397, 2015. ISSN 0036-1445.

[70] Alina Sîrbu, Vittorio Loreto, Vito D. P. Servedio, and Francesca Tria. Opinion dynamics: models, extensions and external effects. *Participatory Sensing, Opinions and Collective Awareness*, pages 363–401, 2017.

[71] Ali Jadbabaie, Pooya Molavi, and Alireza Tahbaz-Salehi. Information heterogeneity and the speed of learning in social networks. *SSRN Electronic Journal*, May 2013. ISSN 1556-5068.

[72] Hui Xie, Guangjian Li, Yongjie Yan, and Sihui Shu. Evolution of bounded confidence opinion in social networks. *Discrete Dynamics in Nature and Society*, 2017:1–5, 2017. ISSN 1026-0226.

[73] Kun Zhao, Márton Karsai, and Ginestra Bianconi. Entropy of dynamical social networks. *PLoS ONE*, 6(12):e28116, 2011.

[74] Thomas Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461–464, 2000.

[75] Javier Borge-Holthoefer, Nicola Perra, Bruno Gonçalves, Sandra González-Bailón, Alex Arenas, Yamir Moreno, and Alessandro Vespignani. The dynamics of information-driven coordination phenomena: A transfer entropy analysis. *Science Advances*, 2(4):e1501158, 2016.

[76] Daniel Ortiz-Arroyo and Dm Hussain. An information theory approach to identify sets of key players. *Proceedings of the 1st European Conference on Intelligence and Security Informatics*, pages 15–26, 2008.

[77] Jacques Calmet. From entropy to ontology. Technical report, Karlsruhe Institute of Technology, 2004.

[78] P E Latham and Y Roudi. Mutual information. *Scholarpedia*, 4:1658, 2009.

[79] Pamela Reinagel, Dwayne Godwin, S Murray Sherman, and Christof Koch. Encoding of visual information by lgn bursts. *Journal of Neurophysiology*, 81 (5):2558–2569, 1999. ISSN 0022-3077.

[80] Michael J. Garee, Hong Wan, and Wai Kin Victor Chan. Regression-based social influence networks and the linearity of aggregated belief. In *2018 Winter Simulation Conference (WSC)*, pages 941–952. IEEE, 2018.

[81] Meysam Alizadeh and Claudio Cioffi-Revilla. Activation regimes in opinion dynamics: Comparing asynchronous updating schemes. *JASSS*, 18(3), 2015. ISSN 14607425.

[82] Robert L Axtell. Effects of interaction topology and activation regime in several multi-agent systems. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*, pages 33–48. Springer, 2000.

[83] Javier Alvarez-Galvez. Network models of minority opinion spreading: Using agent-based modeling to study possible scenarios of social contagion. *Social Science Computer Review*, 34(5):567–581, 2016. ISSN 15528286.

[84] Albert-László Barabasi and Réka Albert. Emergence of scaling in random networks. *Science (New York, N.Y.)*, 286(5439):509–12, 1999. ISSN 1095-9203.

[85] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of small-world' networks. *Nature*, 393(6684):440–442, 1998.

[86] Juliana Martins de Assis and Francisco Marcos de Assis. Estimation of transfer entropy between discrete and continuous random processes. *Journal of Communication and Information Systems*, 33(1):1–11, 2018. ISSN 19806604.

[87] T. Warren Liao. Clustering of time series data - a survey. *Pattern Recognition*, 38(11):1857–1874, 2005. ISSN 00313203.

[88] Malika Charrad, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, 61(6):1–36, 2014.

[89] David Marlow, Susan M. Sanchez, and Paul J. Sanchez. Testing policies and key influences on long-term aircraft fleet management using designed simulation experiments. *Military Operations Research*, 24(3):5–26, 2019. ISSN 10825983, 21632758. URL https://www.jstor.org/stable/26796825.

[90] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York, New York, NY, 2009.

[91] Michael J. Garee. Github repository for supplemental material, 2019. URL https://github.com/mgaree/entropy-social-networks.

[92] F C Santos, J M Pacheco, and Tom Lenaerts. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proceedings of the National Academy of Sciences*, 103(9):3490–3494, 2006. ISSN 0027-8424.

[93] Lu Liu, Jie Tang, Jiawei Han, Meng Jiang, and Shiqiang Yang. Mining topic-level influence in heterogeneous networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM '10*, page 199, 2010.

[94] Lei Tang, Huan Liu, Jianping Zhang, and Zohreh Nazeri. Community evolution in dynamic multi-mode networks. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, page 677, New York, New York, USA, 2008. ACM Press.

[95] Dionne M Aleman, Theo Wibisono, and Brian Schwartz. A nonhomogeneous agent-based simulation approach to modeling the spread of disease in a pandemic outbreak. *Interfaces*, 41(3):301–315, 2011.

[96] Deng Cai, Zheng Shao, Xiaofei He, Xifeng Yan, and Jiawei Han. Mining hidden community in heterogeneous social networks. In *Proceedings of the 3rd international workshop on Link discovery - LinkKDD '05*, pages 58–65, New York, New York, USA, 2005. ACM Press.

[97] Andrea Galeotti, Sanjeev Goyal, and Jurjen Kamphorst. Network formation with heterogeneous players. *Games and Economic Behavior*, 54(2):353–372, 2006. ISSN 08998256.

[98] Sebastiano A. Delre, Wander Jager, and Marco A. Janssen. Diffusion dynamics in small-world networks with heterogeneous consumers. *Computational and Mathematical Organization Theory*, 13(2):185–202, January 2007. ISSN 1381-298X.

[99] Evguenii Kurmyshev, Héctor A. Juárez, and Ricardo A. González-Silva. Dynamics of bounded confidence opinion in heterogeneous social networks: Concord against partial antagonism. *Physica A: Statistical Mechanics and its Applications*, 390(16):2945–2955, 2011.

[100] Zafar Gilani, Reza Farahbakhsh, Gareth Tyson, Liang Wang, and Jon Crowcroft. Of bots and humans (on Twitter). In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 - ASONAM '17*, pages 349–354, New York, New York, USA, 2017. ACM Press.

[101] Michael J. Garee and Mario Ventresca. Social influence network simulation design affects behavior of aggregated entropy. *In preparation*, 2019.

[102] David Kempe, Jon Kleinberg, and Éva Tardos. Influential nodes in a diffusion model for social networks. In *Lecture Notes in Computer Science*, pages 1127–1138, 2005.

[103] Jan Lorenz. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12):1819–1838, December 2007. ISSN 0129-1831.

[104] Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence: Models, analysis and simulation. *JASSS*, 5(3), 2002.

[105] Mohammad Afshar and Masoud Asadpour. Opinion formation by informed agents. *JASSS*, 13(4), 2010. ISSN 14607425.

[106] Dhiraj Murthy, Alison B. Powell, Ramine Tinati, Nick Anstead, Leslie Carr, Susan J. Halford, and Mark Weal. Bots and political influence: A sociotechnical investigation of social network capital. *International Journal of Communication*, 10:4952–4971, 2016. ISSN 19328036.

[107] Alexandre Bovet and Hernán A. Makse. Influence of fake news in Twitter during the 2016 us presidential election. *Nature Communications*, 10(1):7, December 2019.

[108] S. L. Gonzalez Andino, R. Grave de Peralta Menendez, G. Thut, L. Spinelli, O. Blanke, C. M. Michel, and T. Landis. Measuring the complexity of time series: An application to neurophysiological signals. *Human Brain Mapping*, 11 (1):46–57, 2000. ISSN 10659471.

[109] Olaf Sporns. Complexity. *Scholarpedia*, 2(10):1623, 2007. ISSN 1941-6016.

[110] Seth Lloyd and Heinz Pagels. Complexity as thermodynamic depth. *Annals of Physics*, 1988. ISSN 1096035X.

[111] Xiaofeng Liu, Aimin Jiang, Ning Xu, and Jianru Xue. Increment entropy as a measure of complexity for time series. *Entropy*, 18(1):22, January 2016. ISSN 10994300.

[112] B.A. Huberman and T. Hogg. Complexity and adaptation. *Physica D: Nonlinear Phenomena*, 22(1-3):376–384, 1986. ISSN 01672789.

[113] Torsten Köhler and Dirk Lorenz. A comparison of denoising methods for one dimensional time series. *Technical Report DFG SPP1114, University of Bremen,Bremen, Germany,*, 1:1–15, 2005.

[114] M E J Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2005.

[115] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242, 2007.

[116] Charu C Aggarwal, Alexander Hinneburg, and Daniel A Keim. On the surprising behavior of distance metrics in high dimensional space. In *Lecture Notes in Computer Science*, volume 1973, pages 420–434, 2001.

[117] A. Buss and A. Al Rowaei. A comparison of the accuracy of discrete event and discrete time. In *Proceedings of the 2010 Winter Simulation Conference*, pages 1468–1477, Dec 2010. doi: 10.1109/WSC.2010.5679045.

[118] Michael J. Garee and Mario Ventresca. Effects of nonhomogeneous agents in social influence networks on system-level entropy. *In preparation*, 2019.

[119] Michael J. Garee and Mario Ventresca. A complexity measure for opinion dynamics in social influence networks. *In preparation*, 2020.

[120] Kevin Sheppard. Randomgen randomgen v1.14.4 documentation, 2018.

[121] Margarita Sordo and Qing Zeng. On sample size and classification accuracy: A performance comparison. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 3745 LNBI, pages 193–201. Springer, Berlin, Heidelberg, November 2005.

APPENDICES

# A. SUPPLEMENTAL MATERIAL FOR CHAPTER 2

This appendix provides supplemental information for the article, "Regression-based social influence networks and the linearity of aggregated belief."

Table A.1.
Network structure characteristics. We use network metrics not to rigorously analyze each graph, but only to show that there exist some sorts of structural differences between them. Networks are slightly modified to ensure all agents have neighbors and no self-loops exist. The randomization seed parameter is omitted in the table.

| Family | Parameters | Mean out-degree centrality | Assortativity | Reciprocity | Efficiency |
|---|---|---|---|---|---|
| Directed random tree | $N$, edges point toward/away from root | Low | High | Low | Low |
| Scale-free | $N$, $\alpha$, $\beta$, $\gamma$, $\delta_{in}$, $\delta_{out}$ | Low | High | Low to Moderate | Low |
| Erdős-Rényi | $N$, connection probability $p$ | High | Low | Low or High (based on $p$) | High |
| Random $k$-out | $N$, $k$ | Exactly $k$ | Undefined | Various | Moderate |

## A.1 Experimental design matrix

Table A.2 contains an excerpt of the experimental design matrix used for this simulation experiment.

Table A.2.

Excerpt of experimental design matrix.

| Trial | N | Struct. | g | Randomize | b distro | $b_0 = 0$ | normalize bs | Var. | normalize yi | uninformed |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 500 | 12 | 125.0 | False | U-11 | False | network | 2 | True | 0 |
| 2 | 100 | 6 | 100 | False | U-11 | False | network | 2 | False | 0.1 |
| 3 | 100 | 9 | 1 | False | N01 | False | network | 2 | False | 0.1 |
| 4 | 500 | 3 | 5 | False | U-11 | False | network | 2 | True | 0.1 |
| 5 | 500 | 8 | 125.0 | True | N01 | False | network | 1 | True | 0.25 |
| 6 | 500 | 6 | 500 | True | N01 | False | agent | 2 | False | 0.1 |
| 7 | 100 | 9 | 5 | True | U-11 | False | network | 1 | False | 0.1 |
| 8 | 500 | 2 | 5 | True | N01 | False | agent | 1 | True | 0 |
| 9 | 100 | 8 | 100 | False | U-11 | False | agent | 2 | True | 0.1 |
| 10 | 500 | 2 | 500 | False | U01 | False | agent | 1 | False | 0.25 |

## A.2 Source code

Complete source code is available online at https://github.com/mgaree/wsc2018.
Below is an excerpt.

```python
class Agent:
    """
    Args:
        unique_id (int): Unique id number associated with instance.
        model (obj:Model): Reference to parent Model.

    Attributes:
        unique_id (int): Unique id number associated with this agent
            .
        model (obj:Model): Reference to parent Model.
        y_i (float): Belief value.
        b_0 (float): Internal bias.
        b_j (np.array of floats): Coefficients for neighbors belief.
        neighbors (list of Agents): Neighbors of self, based on
            edges in
          self.model.G of the form (self, other).
        x_ij (np.array of floats): Most recently cached values of
            neighbors
          belief.

    """
    def __init__(self, unique_id, model):
        """ Create a new agent. """
        self.unique_id = unique_id
        self.model = model

        # Meet the neighbors
        self.neighbor_ids = list(self.model.G.neighbors(self.
            unique_id))
        self.d_i = len(self.neighbor_ids)

        # Set b_* coefficients
        if self.model.b_0_is_zero:
            self.b_0 = 0
        else:
            self.b_0 = self.model.b_distro()

        self.b_j = np.array([self.model.b_distro() for _ in self.
            neighbor_ids])

        self.y_i = self.b_0 + self.model.error_distro()  # last term
             is eps_i
        self._next_y_i = 0
```

```python
        self.informed = True    # Model will decide if an agent is
            uninformed

<...>

    def step(self):
        """Update belief level using linear regression equation."""
        self.x_ij = self.get_neighbor_beliefs()

        if not self.informed:
            # Check up on neighbors
            neighbor_status = [nn.informed for nn in self.neighbors]
            if True in neighbor_status:
                self.informed = True
            else:
                return

        eps_i = self.model.error_distro()
        self._next_y_i = self.b_0 + np.dot(self.b_j, self.x_ij) +
            eps_i

    def advance(self):
        """Complete the SimultaneousActivation-style step."""
        if self.informed is False:
            return

        self.y_i = self._next_y_i
```

## A.3 Full model results

Table A.2. Regression results (ordinary least-squares) for the preferred main-effects only model on MA-$R_{adj}^2$ at $t = 500$. This model uses the batch quantity g_level, $b_{ij}$ distribution, how $b_{i0}$ and $b_{ij}$ are normalized, whether or not $y_i$ is normalized, and the maximum $d_i$ for the network num_X_vars (a proxy factor for network structure).

| Dep. Variable: | MA-$R_{adj}^2$ | R-squared: | 0.556 |
|---|---|---|---|
| Model: | OLS | Adj. R-squared: | 0.532 |
| Method: | Least Squares | F-statistic: | 22.68 |
| No. Observations: | 173 | Prob (F-statistic): | 1.11e-24 |
| Df Residuals: | 163 | Log-Likelihood: | 5.2028 |
| Df Model: | 9 | AIC: | 9.594 |
| | | BIC: | 41.13 |

| | coef | std err | t | P> \|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|

| | coef | std err | t | P>t | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 0.2016 | 0.067 | 2.987 | 0.003 | 0.068 | 0.335 |
| g_level[T.5] | 0.3342 | 0.056 | 6.002 | 0.000 | 0.224 | 0.444 |
| g_level[T.N] | 0.4057 | 0.065 | 6.268 | 0.000 | 0.278 | 0.534 |
| g_level[T.N/4] | 0.3771 | 0.056 | 6.761 | 0.000 | 0.267 | 0.487 |
| b_distro[T.U-11] | -0.0365 | 0.047 | -0.771 | 0.442 | -0.130 | 0.057 |
| b_distro[T.U01] | 0.2628 | 0.053 | 4.921 | 0.000 | 0.157 | 0.368 |
| normalize_bs[T.network] | -0.3148 | 0.044 | -7.171 | 0.000 | -0.401 | -0.228 |
| normalize_bs[T.no] | -0.1623 | 0.050 | -3.276 | 0.001 | -0.260 | -0.064 |
| normalize_yis[T.True] | -0.0994 | 0.040 | -2.502 | 0.013 | -0.178 | -0.021 |
| num_X_vars | 0.0021 | 0.001 | 2.260 | 0.025 | 0.000 | 0.004 |

| Omnibus: | 2.188 | Durbin-Watson: | 1.943 |
|---|---|---|---|
| Prob(Omnibus): | 0.335 | Jarque-Bera (JB): | 2.181 |
| Skew: | 0.219 | Prob(JB): | 0.336 |
| Kurtosis: | 2.667 | Cond. No. | 141. |

Table A.3. Regression results (ordinary least-squares) for the preferred main-effects and two-way interactions model on MA-$R^2_{adj}$ at $t = 500$. This model uses seven of the ten original design factors and uses backward elimination to remove low $p$-values interaction terms. Diagnostics of this model are graphed in Figure A.1.

| Dep. Variable: | MA-$R^2_{adj}$ | R-squared: | 0.823 |
|---|---|---|---|
| Model: | OLS | Adj. R-squared: | 0.754 |
| Method: | Least Squares | F-statistic: | 12.01 |
| No. Observations: | 173 | Prob (F-statistic): | 1.01e-28 |
| Df Residuals: | 124 | Log-Likelihood: | 84.743 |
| Df Model: | 48 | AIC: | -71.49 |
| | | BIC: | 83.02 |

| | coef | std err | t | P>t | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 0.9822 | 0.191 | 5.136 | 0.000 | 0.604 | 1.361 |
| g_level[T.5] | -0.1799 | 0.155 | -1.162 | 0.248 | -0.486 | 0.127 |
| g_level[T.N] | -0.0868 | 0.170 | -0.510 | 0.611 | -0.424 | 0.250 |
| g_level[T.N/4] | 0.0243 | 0.157 | 0.155 | 0.877 | -0.287 | 0.335 |
| b_distro[T.U-11] | -0.5490 | 0.169 | -3.242 | 0.002 | -0.884 | -0.214 |
| b_distro[T.U01] | -0.2380 | 0.169 | -1.405 | 0.162 | -0.573 | 0.097 |
| norm_yis[T.True] | -0.5095 | 0.142 | -3.579 | 0.000 | -0.791 | -0.228 |
| rand_update_seq[T.True] | 0.0756 | 0.134 | 0.566 | 0.572 | -0.189 | 0.340 |
| norm_bs[T.network] | -0.8281 | 0.129 | -6.416 | 0.000 | -1.084 | -0.573 |
| norm_bs[T.no] | -0.4398 | 0.159 | -2.771 | 0.006 | -0.754 | -0.126 |
| b_0_is_zero[T.True] | 0.0956 | 0.054 | 1.773 | 0.079 | -0.011 | 0.202 |
| g_level[T.5]:b_distro[T.U-11] | 0.1824 | 0.110 | 1.653 | 0.101 | -0.036 | 0.401 |
| g_level[T.N]:b_distro[T.U-11] | 0.0302 | 0.137 | 0.221 | 0.825 | -0.240 | 0.301 |
| g_level[T.N/4]:b_distro[T.U-11] | 0.0535 | 0.115 | 0.467 | 0.641 | -0.173 | 0.280 |
| g_level[T.5]:b_distro[T.U01] | 0.2706 | 0.119 | 2.275 | 0.025 | 0.035 | 0.506 |
| g_level[T.N]:b_distro[T.U01] | 0.0940 | 0.136 | 0.690 | 0.492 | -0.176 | 0.364 |
| g_level[T.N/4]:b_distro[T.U01] | 0.2667 | 0.120 | 2.221 | 0.028 | 0.029 | 0.504 |
| g_level[T.5]:norm_yis[T.True] | 0.1849 | 0.101 | 1.832 | 0.069 | -0.015 | 0.385 |
| g_level[T.N]:norm_yis[T.True] | 0.2338 | 0.116 | 2.022 | 0.045 | 0.005 | 0.463 |

| | coef | std err | t | P>t| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| g_level[T.N/4]:norm_yis[T.True] | 0.2732 | 0.103 | 2.644 | 0.009 | 0.069 | 0.478 |
| g_level[T.5]:rand_update_seq[T.True] | 0.0285 | 0.090 | 0.316 | 0.753 | -0.150 | 0.207 |
| g_level[T.N]:rand_update_seq[T.True] | 0.1010 | 0.104 | 0.975 | 0.331 | -0.104 | 0.306 |
| g_level[T.N/4]:rand_update_seq[T.True] | 0.1458 | 0.091 | 1.599 | 0.112 | -0.035 | 0.326 |
| b_distro[T.U-11]:norm_bs[T.network] | 0.1444 | 0.095 | 1.519 | 0.131 | -0.044 | 0.333 |
| b_distro[T.U01]:norm_bs[T.network] | -0.1666 | 0.102 | -1.634 | 0.105 | -0.368 | 0.035 |
| b_distro[T.U-11]:norm_bs[T.no] | 0.0868 | 0.103 | 0.840 | 0.402 | -0.118 | 0.291 |
| b_distro[T.U01]:norm_bs[T.no] | 0.2225 | 0.120 | 1.851 | 0.067 | -0.015 | 0.460 |
| b_distro[T.U-11]:norm_yis[T.True] | 0.2346 | 0.101 | 2.322 | 0.022 | 0.035 | 0.435 |
| b_distro[T.U01]:norm_yis[T.True] | 0.3520 | 0.108 | 3.248 | 0.001 | 0.138 | 0.567 |
| b_distro[T.U-11]:rand_update_seq[T.True] | -0.0583 | 0.082 | -0.709 | 0.479 | -0.221 | 0.104 |
| b_distro[T.U01]:rand_update_seq[T.True] | -0.1605 | 0.090 | -1.776 | 0.078 | -0.339 | 0.018 |
| norm_bs[T.network]:norm_yis[T.True] | 0.4496 | 0.081 | 5.547 | 0.000 | 0.289 | 0.610 |
| norm_bs[T.no]:norm_yis[T.True] | 0.1863 | 0.099 | 1.886 | 0.062 | -0.009 | 0.382 |
| norm_bs[T.network]:rand_update_seq[T.True] | 0.0360 | 0.073 | 0.491 | 0.624 | -0.109 | 0.181 |
| norm_bs[T.no]:rand_update_seq[T.True] | 0.1574 | 0.078 | 2.010 | 0.047 | 0.002 | 0.312 |
| norm_bs[T.network]:b_0_is_zero[T.True] | -0.2001 | 0.071 | -2.824 | 0.006 | -0.340 | -0.060 |
| norm_bs[T.no]:b_0_is_zero[T.True] | -0.1081 | 0.078 | -1.391 | 0.167 | -0.262 | 0.046 |
| norm_yis[T.True]:rand_update_seq[T.True] | -0.2956 | 0.067 | -4.435 | 0.000 | -0.428 | -0.164 |
| rand_update_seq[T.True]:b_0_is_zero[T.True] | 0.0824 | 0.061 | 1.341 | 0.182 | -0.039 | 0.204 |
| error_variance | -0.2415 | 0.102 | -2.359 | 0.020 | -0.444 | -0.039 |
| g_level[T.5]:error_variance | 0.1987 | 0.081 | 2.456 | 0.015 | 0.039 | 0.359 |
| g_level[T.N]:error_variance | 0.2082 | 0.090 | 2.317 | 0.022 | 0.030 | 0.386 |
| g_level[T.N/4]:error_variance | -0.0011 | 0.084 | -0.014 | 0.989 | -0.168 | 0.166 |
| b_distro[T.U-11]:error_variance | 0.1460 | 0.073 | 2.005 | 0.047 | 0.002 | 0.290 |
| b_distro[T.U01]:error_variance | 0.1175 | 0.078 | 1.515 | 0.132 | -0.036 | 0.271 |
| norm_bs[T.network]:error_variance | 0.2271 | 0.070 | 3.229 | 0.002 | 0.088 | 0.366 |
| norm_bs[T.no]:error_variance | -0.0081 | 0.085 | -0.096 | 0.924 | -0.176 | 0.159 |
| norm_yis[T.True]:error_variance | -0.0937 | 0.067 | -1.403 | 0.163 | -0.226 | 0.038 |
| rand_update_seq[T.True]:error_variance | 0.0777 | 0.063 | 1.241 | 0.217 | -0.046 | 0.202 |

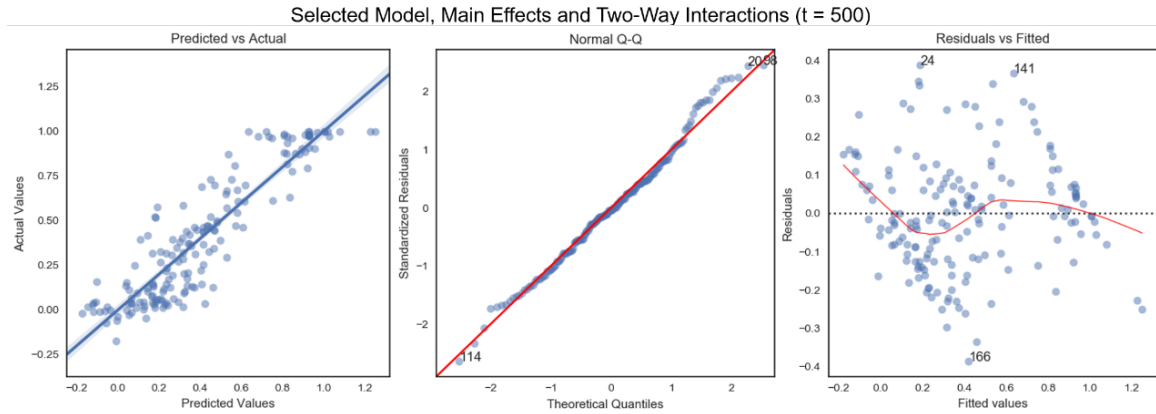| | | | | |
|---|---|---|---|---|
| Omnibus: | 1.727 | Durbin-Watson: | 1.759 |
| Prob(Omnibus): | 0.422 | Jarque-Bera (JB): | 1.647 |
| Skew: | 0.238 | Prob(JB): | 0.439 |
| Kurtosis: | 2.948 | Cond. No. | 77.3 |

Figure A.1. Diagnostic plots of model with main effects and two-way interactions of experimental design regressed on MA-$R^2_{adj}$. This model uses seven experimental factors and achieves $R^2_{adj} = 0.754$.
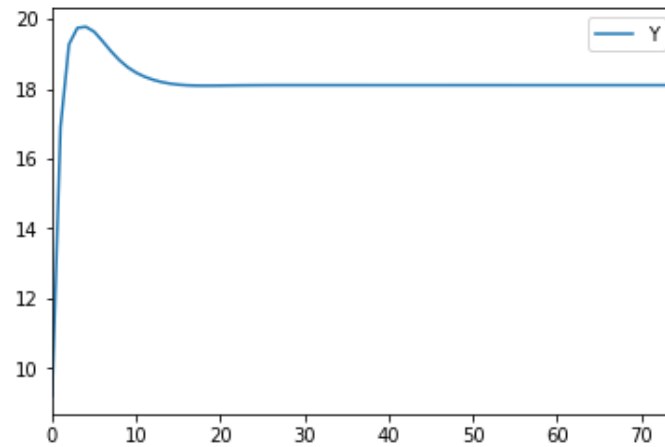


Figure A.2. Simulation configured for standard DeGroot-style interaction shows convergence in total network belief, as expected. This is one part of validating the simulation code.
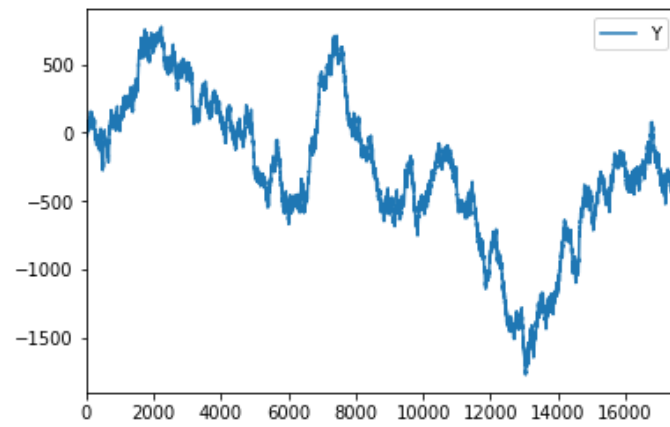
Figure A.3. Simulation configured for standard DeGroot-style interaction is modified to include Normal(0, 1) noise as part of belief update step. Total network belief no longer converges.

# B. SUPPLEMENTAL MATERIAL FOR CHAPTER 3

This appendix provides supplemental information for the article, "Social influence network simulation design affects behavior of aggregated entropy."

## B.1  Methods

The simulation described in this section was implemented in Python 3 using an agent-based modeling approach. Noteworthy libraries include the NetworkX for graph structures and algorithms [65], RandomGen for improved control of random number generation and stream management [120], and Statsmodels for statistical analysis [66]. Complete source code is available at [91].

Replications within a trial use different seeds for the random number generator to ensure independence of responses; replications across trials (e.g., the third replication for every trial) use the same seeds to reduce uncontrolled randomness between trials.

### B.1.1  Experimental design

We used a full factorial design of 1800 trials, based on five design factors: population size, network structure model, agent activation regime, influence model, and influence error term. The design matrix is summarized in Table B.1

Network structure models were drawn from Erdős-Rényi random graphs, small-world graphs, and graphs created through preferential attachment, and parameters were varied within each of these types as detailed in Section B.3. Each of these have different parameters affecting the resulting graph. Erdős-Rényi random graphs have parameter $p$ for the probability of adding each possible edge to the graph. When $p \geq \frac{\log N}{N}$, the probability the graph being connected tends to one. We prefer to study

Table B.1.

The experimental design for this study is a $3^2 \times 4^1 \times 5^1 \times 10^1$ full factorial design on the five factors specified in this table, resulting in 1800 distinct trials. Note that two of the network structure model levels are compressed here but represent all combinations of their respective parameters. References are cited for factors and levels where appropriate; see text for further explanation. A complete listing of trials is available online [91].

| Factor | Level | References |
|---|---:|---:|
| Population size $N$ (3 levels) | | |
| | 100 | |
| | 1000 | [85] |
| | 10,000 | |
| Network structure model (10 levels) | | |
| Erdős-Rényi random, $p = \log N/N$ | | [51] |
| small-world, $k \in 3, 10$, $p \in 0, 0.33, 0.66$ (6 levels) | | [36, 85] |
| preferential attachment, $m \in 1, 3, 5$ (3 levels) | | [84] |
| Agent activation regime (3 levels) | | [52] |
| | synchronous | |
| | uniform | |
| | random | |
| Influence model (5 levels) | | |
| | standard model | [2] |
| | similarity bias | [53] |
| | attractive-repulsive | [53] |
| | individual random adoption | [83] |
| | nonlinear | [Chan 2017][*] |
| Influence error terms (4 levels) | | [53, 57] |
| | none | [29] |
| | $N(0, \sigma = 0.05)$ | |
| | $N(0, 0.1)$ | |
| | $N(0, 0.2)$ | |

[*]WKV Chan, personal communication, 2017

a single connected component in our experiment, so we set $p$ equal to this threshold value. The small-world model takes two parameters: the degree of nodes in the initial ring lattice $k$ and the probability of "rewiring" $p$. We chose $k \in \{3, 10\}$ based on existing work [36, 85]. We selected $p \in \{0, 0.33, 0.66\}$ to span the allowed range $p \in [0, 1]$; the value $p = 1$ was omitted because it produces a purely random graph, which is already supplied by the Erdős-Rényi random graph model. Finally, the preferential attachment model from Barabasi and Albert [84] uses a single parameter, $m$, the number of edges used to attach a new node to the existing network. In their original paper, they use $m = 1$ and $m = 5$. We used those values here and added $m = 3$ as a midpoint. In total, we had ten levels for network structure model: one level for Erdős-Rényi random graphs, six levels for small-world (each pairwise combination of $k$ and $p$), and three levels for networks created via preferential attachment.

Four of the five influence models are detailed in the main text. The final influence model for this design factor was an explicitly nonlinear function of neighbor weights:

$$o_i(t+1) = \frac{b_i}{1 + \exp\left(w_{i1}o_1(t) + w_{i2}o_2(t) + \ldots\right)}, \tag{B.1}$$

where $b_i$ is a $U(0, 1)$-distributed individual bias and all other terms are as defined for the standard model (Equation 3.4). Our use of this model was motivated by (Wai K. V. Chan, personal communication, April 20, 2018).

The values for the influence error distribution standard deviation $\sigma$ were chosen such that 95% of the time, the magnitude of the error term accounts for no more than 10%, 20%, or 40% of the possible range of opinion values, respectively.

Synchronous activation has all agents take their actions "simultaneously in one discrete time step" [53] by using the system state from the end of the previous time step; this approach is used in DeGroot's original social learning model [29]. Uniform activation triggers agents one at a time, and the sequence is shuffled after each pass through the full population, akin to sampling without replacement. For both synchronous and uniform, agents have an equal number of activations. Finally, random activation samples the population with replacement, so agents may see an unequal number of activations.

Figure B.1. Symbolic approach for transfer entropy with a sub-sample length of 3 leads to 6 possible patterns or symbols. (Figure adapted from [75], Supplementary Materials.)

## B.1.2 Response variables

Each continuous data technique has several implementation details for computing entropy for individual agents. For binning, we constructed histograms on the opinion data using a fixed number of bins, depending on the entropy measure. Bin counts were chosen to balance the number of bins with the expected number of data points per bin. We used fixed bins rather than an adaptive partitioning method so that comparisons between trials were not influenced by the specification of bins. The histograms for an entropy calculation at time $t$ were built with opinion values from time 0 to $t$. Then, we built probability mass functions on the histogram values to be used with the discrete forms of the entropy measures.

The *symbolic approach* transforms the input data by first mapping the elements of a time series into patterns of relative orderings between values (e.g., Figure B.1). It required us to select the number of data points (time steps) used in one symbol. More data points per symbol can more faithfully capture the dynamics of the time series, but computational costs rapidly increase due to the growing probability state-space in the entropy calculations. Since the goal is to study changes in entropy rather than optimally measure it, we used the minimum recommended number of data points per symbol, three [75]. This gave an alphabet of $3! = 6$ symbols (breaking ties between consecutive equal values at random).

Note: The symbolic method on all entropy measures and transfer entropy in general use agent opinions from multiple time steps in order to calculate the entropy value for a single time step. As a result, the final few time steps are not available for some RVs. We truncate them uniformly at $t = 490$, while the unaffected RVs have data through $t = 500$. These are considered the final time steps for their respective RVs.

## B.2   Results

The agent-based simulation experiment produced response variable time series data for 1800 independent trials.

Table B.2.
In many of the results graphics, experimental levels are encoded as single letters for better readability.

|  | $N$ | structure | influence model | error | activation |
|---|---|---|---|---|---|
| a | 100 | erdos_renyi_random(N) | standard_model | none | synchronous |
| b | 1000 | small_world(N, 0.0, 3) | similarity_bias | N(0, $\sigma = 0.05$) | uniform |
| c | 10000 | small_world(N, 0.0, 10) | attractive_repulsive | N(0, 0.1) | random |
|  |  |  |  |  |  |
| d |  | small_world(N, 0.33, 3) | random_adoption | N(0, 0.2) |  |
| e |  | small_world(N, 0.33, 10) | nonlinear |  |  |
| f |  | small_world(N, 0.66, 3) |  |  |  |
| g |  | small_world(N, 0.66, 10) |  |  |  |
|  |  |  |  |  |  |
| h |  | scale_free(N, 1) |  |  |  |
| i |  | scale_free(N, 3) |  |  |  |
| j |  | scale_free(N, 5) |  |  |  |

### B.2.1 Experimental design analysis

The Kruskal-Wallace test is used to determine if varying the level of each factor has a statistically significant effect on the response variable. At the 0.05 level of significance, when the data is split into levels for population size $N$ and for activation regime, the values for RE-B appear to come from the same population. Practically, this suggests that varying these two factors—over the levels specified in our experiment—does not have a significant effect on the RV. This agrees with what we observe in the previous figures. These results are summarized in Table 3.1, item iii. We see similar agreement for the other RVs between the Kruskal-Wallace test results and the main effect plots.

Table B.3.
The Kruskal-Wallace test is conducted on RE-B values for trials at $t = 500$ to test if changing the level for a factor has a statistical effect on the response value. The asterisks indicate that population size $N$ and activation regime show no significant impact on RE-B.

| factor | K-W test stat | p-value |
|---|---|---|
| population size $N$ | 2.05 | * 3.57e-01 |
| network structure | 29.37 | 5.60e-04 |
| influence model | 622.07 | 2.58e-133 |
| error distribution | 892.12 | 4.51e-193 |
| activation regime | 0.13 | * 9.33e-01 |

Although the Kruskal-Wallace test indicates that for RE-B, differences exist among levels for the structure, influence model, and error factors, it does not identify where those differences are. For that, we use the Mann-Whitney U test on each pair of levels for a factor (Figure B.2). A significant ($< 0.05$) p-value indicates a statistical difference between the tested pair of experimental levels. This shows that all influence models are significantly different from one another, while only some network structures differ. The Mann-Whitney U test also finds no significant differences between

Figure B.2. We use the Mann-Whitney U test for post-hoc comparison testing to determine which levels are statistically different within each factor. The numbers in cells for the pairs with a significant test statistic ($< 0.05$) express the p-value as a percentage (e.g. 3* means the p-value is at least 0.03 and less than 0.04).

levels for $N$ and activation regime, further supporting their lack of importance for RE-B. These results are summarized in Table 3.1, item iv.

Across all RVs, network structures tend to be statistically different from structures in different density groups (higher/lower density; detailed in Section B.3). Interestingly, TE-B shows only three out of 45 pairs of network structures to be statistically different, while other RVs have many more differing pairs; that difference is not understood at this time.

Table B.4.
Each Evaluation row from the summary tables (e.g., Table 3.1) for research question 1 are combined to allow comparisons among all RVs. Only the activation regime changes based on the entropy measure.

*\* (Evaluation) Is the response variable sensitive to changes*

*in the level for the factor?*

Factor (number of levels)

| RV | $N$ (3) | structure (10) | influence model (5) | error (4) | activation (3) |
|------|------|------|------|------|------|
| RE-B | no | yes | yes | yes | no |
| MI-B | no | yes | yes | yes | yes |
| TE-B | no | yes | yes | yes | yes |
| | | | | | |
| RE-S | no | yes | yes | yes | no |
| MI-S | no | yes | yes | yes | yes |
| TE-S | no | yes | yes | yes | yes |

**B.2.2   Cluster analysis**

Table B.5 summarizes the total and meaningful cluster quantities for each RV and distance measure.

Table B.5.
Cluster analysis using dynamic time warping (DTW) as the distance measure can produce at least one meaningful cluster, with respect to the experimental design, for each RV, while clusters based on Pearson's correlation coefficient are meaningless.

|  | DTW | | Pearson corr. | |
| --- | --- | --- | --- | --- |
|  | total clusters | meaningful clusters | total clusters | meaningful clusters |
| RE-B | 4 | 2 | 2 | 0 |
| MI-B | 3 | 2 | 12 | 0 |
| TE-B | 2 | 1 | 2 | 0 |
| RE-S | 2 | 1 | 2 | 0 |
| MI-S | 5 | 4 | 2 | 0 |
| TE-S | 4 | 3 | 3 | 0 |

DTW for RE-B produced four fairly dense clusters (Figure B.3), while Pearson's correlation produced two homogeneous clusters (Figure B.4). In the clustered time series plots for the other RVs, DTW produces dense groups for nearly every cluster (suggesting potentially high quality), and Pearson's correlation produces clusters spread rather uniformly across the range (suggesting low quality).

**B.2.3   Comparison among response variables**

Figure B.6 shows the profiles of all trials for each RV. Two general patterns appear: one for relative entropy and another for mutual information and transfer entropy.

relative entropy, binning (RE-B), all trials, DTW clusters



Figure B.3. When using dynamic time warping (DTW) as the distance measure between pairs of time series for RE-B, the consensus method produces four clusters, each highlighted here on the original time series plot (Figure 3.1, left). The densely grouped nature of these clusters suggest a potentially high degree of cluster quality.

relative entropy, binning (RE-B), all trials, Pearson clusters



Figure B.4. When using Pearson's correlation as the distance measure, the consensus method produces only two clusters. The results are visually indistinct and may indicate low cluster quality.

Figure B.5. Clusters produced with Pearson's correlation are completely undifferentiated.

These two patterns also manifest in the main effect violin plots for the corresponding RVs (Figure B.7).

normalized response variables for all trials



Figure B.6. RVs for all trials reveal two distinct behaviors: one for relative entropy (RE) and a second for mutual information (MI) and transfer entropy (TE). Data in each subplot is normalized to its maximum value for that RV.

Figure B.7. The two characteristic behaviors in the complete time-series data (Figure B.6) appear in the main effect violin plots, represented here with RE-B (left, repeated from Figure 3.2) and MI-S (right). RE-B shows generally wide distributions while MI-S features compact distributions with long, thin tails.

### B.3  Network structure information

We use three network structure models in our experimental design, as introduced in Section 3.3.2: Erdős-Rényi random, small-world $\beta$, and preferential attachment. We also use an assortment of parameters for these models, detailed below. In total, this provides ten levels in the experimental design for the network structure factor.

Each level is represented as one of the following:

- *erdos_renyi_random(N)* produces an Erdős-Rényi random graph with density $\frac{\ln N}{N}$. This density equation is chosen as it is percolation threshold above which the probability of the network being connected approaches one.

- *small_world(N, p, k)* uses the $\beta$-model of Watts and Strogatz [85], where $p$ is the probability of rewiring each edge in a ring lattice network that initially has each node connected to its $k$ nearest neighbors.

- *scale_free(N, m)* produces a graph using the Barabasi-Albert preferential attachment process [84], where nodes are added to the network one at a time using $m$ edges.

The ten network structure levels in the experimental design are:

a. *erdos_renyi_random(N)*,

b. *small_world(N, 0.0, 3)*,

c. *small_world(N, 0.0, 10)*,

d. *small_world(N, 0.33, 3)*,

e. *small_world(N, 0.33, 10)*,

f. *small_world(N, 0.66, 3)*,

g. *small_world(N, 0.66, 10)*,

h. *scale_free(N, 1)*,

i. *scale_free(N, 3)*, and

j. *scale_free(N, 5)*.

Figure B.8. Network structures for the first replication of $N = 100$ trials. Based on the visual layout, the structures exhibit three patterns. Structures d, f, and h are tree-like; structure b and c are ring lattices; and the remaining structures (a, e, g, i, and j) appear as denser networks. The graphics are created using the yFiles radial layout in Cytoscape.

$N$ is the population size of the network. In our design, $N \in \{100, 1000, 10{,}000\}$.

Figure B.8 shows an instance of each of the the network structures for $N = 100$. Based on the shapes of the networks, three structural patterns are apparent: tree-like, ring lattices, and denser networks. Network density proved to be the most compelling metric for understanding (some of) the similarities and differences in simulation response variable distributions across each network structure (Figure B.11).

Each trial replication uses a different random number seed to create the network structure, so the structures differ slightly across the replications of a single trial. However, the seeds are controlled such that for every trial, replications use the same sequence of seeds. Therefore, the first replication for all trials with a given setting

(a)
erdos_renyi_random(1000)

(b)
small_world(1000, 0.0, 3)

(c)
small_world(1000, 0.0, 10)

(d)
small_world(1000, 0.33, 3)

(e)
small_world(1000, 0.33, 10)

(f)
small_world(1000, 0.66, 3)

(g)
small_world(1000, 0.66, 10)

(h)
scale_free(1000, 1)

(i)
scale_free(1000, 3)

(j)
scale_free(1000, 5)

Figure B.9. Network structures for the first replication of $N = 1000$ trials. Despite the larger network population, these visual layouts fall into the same three categories as with $N = 100$ (Figure B.8) and are a qualitative match with their counterparts that have the smaller population.

for $N$ and network structure model have the same network structure, and so on for each replication. Figure B.8 diagrams the network structures for $N = 100$ using the random number seed for the first replication; diagrams for structures created using the seed for the 100th replication (not pictured) show the same three patterns while differing in exact details.

These three groups are preserved when increasing the population size. Figure B.9 shows the results for $N = 1000$. Unfortunately, the networks for $N = 10,000$ exceeded the rendering software's capabilities, but we predict the same three patterns to be present.

Degree distributions (Figure B.10) did not exhibit the expected similarity between distribution of the same visual pattern, nor did average degree connectivity or

Out-degree densities for $N = 100$, replication 1



Figure B.10. Out-degree density distributions of the network structures in Figure B.8 do not fully align with the three visual patterns (tree-like, ring lattice, and denser networks), where we would expect more similarity between structures h and f, as well as between j and g.

shortest paths. However, network density proved to be the most compelling metric for understanding (some of) the similarities and differences in simulation response variable distributions across each network structure (Figure B.11).

Figure B.11. The groupings of network densities more closely align with the response variable behavior observed in Section C, so density may have an effect on the entropies measured in this research.

# C. DETAILED RESULTS PER RESPONSE VARIABLE FOR CHAPTER 3

In this part, each of the six response variables are analyzed independently of one another with respect to research questions 1 and 2. In C.1, we present detailed analysis of a single response variable with accompanying rationale for why the analysis was performed that way and explanation of some of the tests used. C.2-C.6 repeat that analysis with the remaining response variables but will omit the rationale and explanation elements for brevity; otherwise, all following sections will proceed identically. All readers are encouraged to read C.1 to orient themselves to the processes and figures.

We use a full factorial design for the experiment, so every combination of the experimental levels is equally represented in the data used for the following analysis. Table C.1 maps experimental level names to single letters that are used in many of the figures, in order to conserve space.

## C.1  Response variable 1 - relative entropy, binning (RE-B)

Relative entropy, binning (RE-B) assigns agent opinion to one of a set of equal-width bins and computes the relative entropy (Equation 3.1) of the resulting distribution $p(x)$ with respect to the uniform distribution $q(x)$, averaging across each agent and each replication to produce the trial-level response. Figure C.1 shows the time series of RE-B for each trial and an associated kernel density estimate (KDE) for the final time step. This shows some groupings within trials—hinting at possible cluster analysis outcomes—and shows the data set as a whole to not be normally distributed—limiting the relevance of mean values and analysis tools like ANOVA.

Table C.1.
In the following plots, experimental levels are mapped to single letters for
spacing reasons.

| | $N$ | structure | influence model | error | activation |
|---|---|---|---|---|---|
| a | 100 | erdos_renyi_random(N) | standard_model | none | synchronous |
| b | 1000 | small_world(N, 0.0, 3) | similarity_bias | N(0, 0.05) | uniform |
| c | 10000 | small_world(N, 0.0, 10) | attractive_repulsive | N(0, 0.1) | random |
| d | | small_world(N, 0.33, 3) | random_adoption | N(0, 0.2) | |
| e | | small_world(N, 0.33, 10) | nonlinear | | |
| f | | small_world(N, 0.66, 3) | | | |
| g | | small_world(N, 0.66, 10) | | | |
| h | | scale_free(N, 1) | | | |
| i | | scale_free(N, 3) | | | |
| j | | scale_free(N, 5) | | | |

relative entropy, binning (RE-B), all trials



Figure C.1. The time series of RE-B values for each trial are plotted on the same axes to reveal visual clusters at the extremes and center of the range. A gaussian kernel density estimate for the RE-B values at $t = 500$ shows the data to be somewhat tri-modal and also reinforces the visual clustering.

Each trial undergoes an initial transient period before becoming monotonic. For some trials, RE-B approaches zero (the minimum bound for relative entropy) but does not reach it within the length of the simulation. The upper extreme appears to correspond to trials where individual agent opinion converged; we use fifty bins for RE-B, so the upper limit for relative entropy with respect to the uniform distribution occurs when its opinion takes on only a single value:

$$D_X(p \parallel q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)} = 1 \log_2 \frac{1}{1/50} \approx 5.644. \tag{C.1}$$

**Research question 1: Which system design factors contribute most to aggregated entropy?**

For this research question, we explore the one-way sensitivity of each entropy response variable to changes in the levels of individual experimental design factors. This exploration includes qualitative comparisons of RV distributions when the trial data is grouped by experimental levels and statistical tests for differences between levels. These methods support a subjective evaluation of whether an RV is sensitive to changes in the level of each design factor. In Table C.2, we summarize the analysis results for the current response variable for this research question.

Analysis for design of experiments (DoE) typically focuses on changes to mean responses. However, the non-normal distribution of the data (Figure C.1) shows multiple modes, so the mean value is not salient. Instead, we look for main effects using the full distributions of the data when grouped by experimental level. Figure C.2 presents these distributions for RE-B at the final time step, $t = 500$, using a half-violin plot. Differences between distributions among the levels for a single factor qualitatively show the effect each level has on the response. For example, the distributions for population size $N$ are almost identical, so we infer that $N$ is not important (i.e., does not have a significant effect on the response variable), at least over the range of levels used in the experimental design; the same holds for the agent activation regime. On the other hand, strong differences between distributions are visible for the influence model and error distribution, marking these as important to RE-B. Structure models appear to fall into at least two groups of similar response distributions, and these groups resemble the higher/lower density networks (identified in B.3, Figure B.11).

Figure C.2 uses data for only a single time step. To reveal the effect of time on the response variable, Figures C.3-C.7 plot the medians of the grouped data over the full length of the simulation. The two inner quartiles (25th to 75th percentiles) are indicated by the shaded regions around each line. Overall, these figures reinforce the

Table C.2.

The findings for research question 1 on RE-B are summarized to support
the overall evaluation of each experimental design factor (final table row).

| | Factor (number of levels) | | | | |
|---|---|---|---|---|---|
| | $N$ (3) | structure (10) | influence model (5) | error (4) | activation (3) |
| i. | *(Main effect plot) What differences are present between the response variable distributions for each level at the final time step?* | | | | |
| | negligible | 2 or 3 patterns | 3 or 4 patterns | 3 patterns | negligible |
| ii. | *(Grouped time series) What differences are present between the median response values over the duration of the simulation?* | | | | |
| | negligible | overall similar shapes; small divide affected by density | 3 patterns; nonlinear+standard, similarity+random paired | 3 patterns; high+medium variance identical | negligible |
| iii. | *(K-W test) Does the Kruskal-Wallace test indicate statistical differences in the response variable between each level at the final time step? (i.e., is the p-value <0.05?)* | | | | |
| | no | yes | yes | yes | no |
| iv. | *(M-W U test) How many pairs of levels are statistically different (p-value <0.05) according to the Mann-Whitney U test?* | | | | |
| | 0/3 | 28/45 | 10/10 | 5/6 | 0/3 |
| * | *(Evaluation) Is the response variable sensitive to changes in the level for the factor?* | | | | |
| | no | yes | yes | yes | no |

Figure C.2. Each half-violin of this design of experiments (DoE) main effect plot represents the distribution of RE-B at $t = 500$ for all trials with the corresponding level on the horizontal axis, and its median is indicated with a horizontal dash; the grand median is shown for reference. For the network structure factor, horizontal space separates the three model families (Erdős-Rényi random, small world, and scale-free). Refer to Table C.1 to identify all experimental levels, and see text for further discussion.

This plot suggests that $N$ and activation are unimportant to the response variable, while zero error (level a) leads to significantly different outcomes than the other error distributions.

median relative entropy, binning (RE-B), all trials, grouped by $N$



Figure C.3. All trials are grouped by experimental level as in Figure C.2 and the groups' median response value over time is plotted. Shaded regions around each line enclose the 25th to 75th percentiles of the data. For population size $N$, these regions almost entirely overlap, which reinforces the low importance of $N$ shown in the DoE main effect plot.

similarities and differences observed in the main effect plot. They also show that the (median) response values are fairly stable over time, after an initial transient, so any observations made at $t = 500$ should be informative about the system over a longer period of time.

Thus far, we have use qualitative approaches to show the effect of varying individual design factors. Now, we turn to statistical measures. ANOVA is the classical tool for analyzing the results of a designed experiment, but it assumes the data to be normally distributed, which is not the case here (Figure C.1). Further, Levene's test shows that three of the five design factors violate ANOVA's assumption of homogeneity of variance. While ANOVA is robust to violations of these assumptions, we instead adopt a non-parametric approach to measuring differences between experimental levels.

median relative entropy, binning (RE-B), all trials, grouped by structure



Figure C.4. Following Figure C.3 in design, this plot shows some differentia-tion between network models but many are very similar. The line labels are allowed to overlap to reinforce the small differences between experimental levels. The scale free $(N, 1)$ model does stand out above the rest.

median relative entropy, binning (RE-B), all trials, grouped by influence model



Figure C.5. Some grouping is apparent based on the level of the influence model. One explanation is that nonlinear and the standard model both use weighted averages of neighbor opinion, while similarity bias and random adoption interact with (at most) one neighbor at a time.

median relative entropy, binning (RE-B), all trials, grouped by error



Figure C.6. With the influence error distribution, we observe clear separations in the response variable, again reinforcing the findings from Figure C.2.

median relative entropy, binning (RE-B), all trials, grouped by activation



Figure C.7. As with population size $N$, the agent activation regime appears to be unimportant to RE-B, although some differences are present in the early time steps.

Table C.3.
The Kruskal-Wallace test is ran on trial-level RE-B values at $t = 500$ to
test if changing the level for a factor has a statistical effect on the response
value. The asterisks indicate that population size $N$ and activation regime
show no significant impact on RE-B.

| factor | K-W test stat | p-value |
| --- | --- | --- |
| $N$ | 2.05 | * 3.57e-01 |
| structure | 29.37 | 5.60e-04 |
| influence model | 622.07 | 2.58e-133 |
| error | 892.12 | 4.51e-193 |
| activation | 0.13 | * 9.33e-01 |

The Kruskal-Wallace test lets us statistically determine if varying the level of each
factor has a significant effect on the response variable. Based on the p-values from
the Kruskal-Wallace test (Table C.3) on RE-B at $t = 500$, when the data is split into
levels for both population size $N$ and activation regime, the data appears to come
from the same population. Practically, this suggests that varying these factors—
over the levels specified in our experiment—does not have a significant effect on the
response variable. This agrees with what we observe in the previous figures.

Although the Kruskal-Wallace test indicates that differences exist among levels
for the structure, influence model, and error factors, it does not identify where those
differences are. For that, we use the Mann-Whitney U test on each pair of levels for
a factor. A significant ($< 0.05$) p-value indicates a statistical difference between the
tested pair of experimental levels. Figure C.8 aggregates the results. (We include $N$
and activation regime in the results to show consistency between the Mann-Whitney
U test results and Kruskal-Wallace.)

Some interesting observations can be made about what pairwise results are/are
not significant. For example, in the network structure models, structures with higher

Figure C.8. We use the Mann-Whitney U test for post-hoc comparison testing to determine which levels are statistically different within each factor. The numbers in cells for the pairs with a significant test statistic ($< 0.05$) express the p-value as a percentage (e.g. 3* means $0.03 \leq$ p-value $< 0.04$). The non-significant results for $N$ and activation are consistent with the previous findings. The results for influence model are somewhat unexpected since similar pairs of models appear in Figures C.2 and C.5. See text for further discussion.

density test as similar to nearly all other higher density structures, while structures with lower density test as different from most other low density structures (see Figure B.11 for density values). Also, this test finds a significant difference between the standard model (level a) and nonlinear (level e) influence models, which is not apparent from the DoE main effect or grouped time series plots.

Overall, for the RE-B response variable, varying population size $N$ and the agent activation regime has no significant effect; influence model error terms set to zero leads to higher relative entropy than normally distributed, zero mean error, and that effect varies slightly with the variance of the distribution; differences in the response due to changing the network structure model are slight, though tree-like structures have a higher minimum response than dense graphs; and influence models that cause agents to interact with one neighbor at a time lead to lower response values than those that interact with all neighbors at once.

## Research question 2: How is system design related to the response space of entropy time-series values?

Cluster analysis has three elements: clustering algorithm, distance (or dissimilarity) measure, and evaluation criteria of the results [87]. Few guidelines exist for designing a cluster analysis *a priori*, so we used an assortment of options to search for meaningful clusters in our entropy time-series data. For the distance (or dissimilarity) measure, we used dynamic time warping (DTW) and Pearson's correlation coefficient between each pair of trial-level time series for the response variable. Then, we used fourteen clustering methods provided by the R library NbClust.[1] Each clustering method suggested an optimal number of clusters (which we bounded between two and twelve, inclusive), then the number of clusters with the most "votes" was passed to an agglomerative hierarchical clustering method to assign each trial to a cluster. This process was applied to both distance measures.

Cluster assignments are summarized in the following figures. DTW for RE-B produced four clusters (Figure C.9), while Pearson's correlation produced two clusters (Figure C.10). With respect to the time series plots, DTW led to rather differentiated clusters, while Pearson's correlation produced highly homogeneous clusters.

---

[1]NbClust provides more than fourteen methods, but ten produced errors due to the dimensions of the data, and several were omitted due to their overwhelming computation time.

relative entropy, binning (RE-B), all trials, DTW clusters



Figure C.9. Using dynamic time warping (DTW) as the distance measure between pairs of response variable time series, the consensus method produces four clusters, each highlighted here using the original time series plot (Figure C.1). The densely grouped nature of these clusters suggest a high level of cluster quality.

relative entropy, binning (RE-B), all trials, Pearson clusters



Figure C.10. Using Pearson's correlation as the distance measure, the consensus method produces only two clusters. The results are less visually satisfying than DTW (Figure C.9) and may indicate less meaningful clusters.

Figure C.11. For each cluster produced through DTW, the trials assigned to the cluster are grouped by experimental level in order to find the percentage of a cluster associated with each experimental level. Within a cluster (row), the percentages for a single factor sum to one; within a experimental level (column), there are no such constraints. For example, all trials assigned to cluster 1 use error level a (no error), and trials assigned to cluster 4 use either influence model b or d.

However, clusters differentiated in the response variable space are not necessarily meaningful. In Figures C.11 and C.12, we conduct a "census" of the trials assigned to each cluster, with respect to the experimental design factors. For DTW, cluster 1 contains exclusively trials with no influence error term (error level a), though other clusters do contain a small number of such trials, as well. Cluster 4 is strong in the similarity bias and random adoption influence models (levels b and d), perhaps because those models have agents interact with a single neighbor at a time, while the other models have agents interact with all neighbors at once. Interestingly, cluster 4 is also very weak in lower-density network structures (levels b, d, f, and h). The levels for population size $N$ and activation regime are uniformly distributed within each cluster, further reinforcing the earlier evidence that those factors are not important to RE-B.

In summary, the variation in system design studied here can produce meaningful clusters, with respect to the experimental design factors, in the response space

Figure C.12. Clusters produced through Pearson's correlation are completely undifferentiated.

mutual information, binning (MI-B), all trials



Figure C.13. The time series of MI-B values for each trial are plotted on the same axes to reveal a visual cluster at the bottom of the range.

for RE-B. This effect is achieved when using dynamic time warping as the distance measure, but not when using Pearson's correlation coefficient.

## C.2  Response variable 2 - mutual information, binning (MI-B)

Mutual information, binning (MI-B) assigns agent opinion to one of a set of equal-width bins and computes the mutual information (Equation 3.2) between each agent-neighbor pair, averages across the neighbors for each agent, and then averages across each agent and each replication to produce the trial-level response. Figure C.13 shows the time series of MI-B for each trial and an associated kernel density estimate (KDE) for the final time step. This shows a grouping near zero mutual information, and that most—but not all—trials decrease over time.

**Research question 1: Which system design factors contribute most to aggregated entropy?**

For this research question, we explore the one-way sensitivity of each entropy response variable to changes in the levels of individual experimental design factors. This exploration includes qualitative comparisons of RV distributions when the trial data is grouped by experimental levels and statistical tests for differences between levels. These methods support a subjective evaluation of whether an RV is sensitive to changes in the level of each design factor. In Table C.4, we summarize the analysis results for the current response variable for this research question.

Figure C.14 presents the distributions grouped by experimental level for MI-B at the final time step, $t = 500$, using a half-violin plot. Differences between distributions among the levels for a single factor qualitatively show the effect each level has on the response. For example, the distributions for population size $N$ are almost identical, so we infer that $N$ is not important (i.e., does not have a significant effect on the response variable), at least over the range of levels used in the experimental design. On the other hand, strong differences between distributions are visible for the influence model, marking it as important to MI-B.

Figure C.14 uses data for only a single time step. To reveal the effect of time on the response variable, Figures C.15-C.19 plot the medians of the grouped data over the full length of the simulation. The two inner quartiles (25th to 75th percentiles) are indicated by the shaded regions around each line. Overall, these figures reinforce the similarities and differences observed in the main effect plot.

Thus far, we have use qualitative approaches to show the effect of varying individual design factors. We now adopt a non-parametric approach to measuring differences between experimental levels, using the Kruskal-Wallace test and Mann-Whitney U test.

Based on the p-values from the Kruskal-Wallace test (Table C.5) on MI-B at $t = 500$, when the data is split into levels for population size $N$, the data appears to

Table C.4.

The findings for research question 1 on MI-B are summarized to support the overall evaluation of each experimental design factor (final table row).

| | | | Factor (number of levels) | | |
|---|---|---|---|---|---|
| | N (3) | structure (10) | influence model (5) | error (4) | activation (3) |
| i. | *(Main effect plot) What differences are present between the response variable distributions for each level at the final time step?* | | | | |
| | negligible | 3 or 4 patterns | 5 patterns | 2 patterns; error vs no error | synchronous different than others |
| ii. | *(Grouped time series) What differences are present between the median response values over the duration of the simulation?* | | | | |
| | negligible | initial grouping by density, then partial convergence | 4 patterns; nonlinear+standard similar | most initially distinct, then partial convergence | minor early variation, then partial convergence |
| iii. | *(K-W test) Does the Kruskal-Wallace test indicate statistical differences in the response variable between each level at the final time step? (i.e., is the p-value <0.05?)* | | | | |
| | no | yes | yes | yes | yes |
| iv. | *(M-W U test) How many pairs of levels are statistically different (p-value <0.05) according to the Mann-Whitney U test?* | | | | |
| | 0/3 | 24/45 | 10/10 | 3/6 | 3/3 |
| * | *(Evaluation) Is the response variable sensitive to changes in the level for the factor?* | | | | |
| | no | yes | yes | yes | yes |

DoE main effect plot for
mutual information, binning (MI-B), all trials, $t = 500$

Figure C.14. Each half-violin of this design of experiments (DoE) main effect plot represents the distribution of MI-B at $t = 500$ for all trials with the corresponding level on the horizontal axis, and its median is indicated with a horizontal dash; the grand median is shown for reference. This plot suggests that $N$ is unimportant to the response variable, while influence model leads to highly varied outcomes.

median mutual information, binning (MI-B), all trials, grouped by $N$



Figure C.15. All trials are grouped by experimental level as in Figure C.14 and the groups' median response value over time is plotted. Shaded regions around each line enclose the 25th to 75th percentiles of the data. For population size $N$, these regions almost entirely overlap due to the closeness of the median lines, which reinforces the low importance of $N$ shown in the DoE main effect plot.

median mutual information, binning (MI-B), all trials, grouped by structure



Figure C.16. Following Figure C.15 in design, this plot shows some differentiation between network models but many are very similar. An interesting split into two groups occurs during the initial transient between lower density (lower group) and higher density (upper group) networks.

median mutual information, binning (MI-B), all trials, grouped by influence model



Figure C.17. The standard model and nonlinear model are closely aligned in this plot, but all other levels are well-differentiated from each other. One explanation is that nonlinear and the standard model both use weighted averages of neighbor opinion, while the other models do not.

median mutual information, binning (MI-B), all trials, grouped by error



Figure C.18. With the influence error distribution, we observe initial differences in the response variable, but the median lines converge after $t = 300$.

median mutual information, binning (MI-B), all trials, grouped by activation



Figure C.19. The median lines for the three activation regimes are nearly identical, but their variability differs, as also indicated in the main effect plot.

Table C.5.

The Kruskal-Wallace test is ran on trial-level MI-B values at $t = 500$ to test if changing the level for a factor has a statistical effect on the response value. The asterisk indicates that population size $N$ has no significant impact on MI-B.

|  | test stat | p-value |
|---|---|---|
| $N$ | 1.12 | * 5.69e-01 |
| structure | 28.80 | 6.99e-04 |
| influence model | 1317.66 | 4.93e-284 |
| error | 47.96 | 2.16e-10 |
| activation | 41.50 | 9.70e-10 |

come from the same population. Practically, this suggests that varying this factor— over the levels specified in our experiment—does not have a significant effect on the response variable. This agrees with what we observe in the previous figures.

Figure C.20 aggregates the results of the Mann-Whitney U test applied to each pair of levels within a factor. Some interesting observations can be made about what pairwise results are/are not significant. In the network structure models, only some of the similar pairs of levels (i.e., those with grey cells in the figure) have similar densities.

Overall, for the MI-B response variable, varying population size $N$ and agent activation regime has no significant effect; differences in network structure models and error terms have greater effects early in a run but very minor effects in the long term; and each influence model has a distinct signature.

Figure C.20. We use the Mann-Whitney U test to determine which levels are statistically different within each factor. The numbers in cells for the pairs with a significant test statistic ($< 0.05$) express the p-value as a percentage (e.g. 3* means $0.03 \leq$ p-value $< 0.04$). The non-significant results for $N$ are consistent with the previous findings.

mutual information, binning (MI-B), all trials, DTW clusters



Figure C.21. Using dynamic time warping (DTW) as the distance measure between pairs of response variable time series, the consensus method produces three clusters, each highlighted here using the original time series plot (Figure C.13). The densely grouped nature of these clusters suggest a reasonable level of cluster quality.

**Research question 2: How is system design related to the response space of entropy time-series values?**

Using the cluster analysis process described in C.1, trials are assigned to clusters for both DTW and Pearson's correlation. These assignments are summarized in the following figures. DTW for MI-B produced three clusters (Figure C.21), while Pearson's correlation produced twelve clusters (Figure C.22), the maximum number of clusters considered by our analysis process. (Conventional guidance says that if multiple clustering methods call for the minimum/maximum number of clusters, then the selected methods or distance metric may be unsuitable for clustering the data.) With respect to the time series plots, DTW led to rather differentiated clusters, while Pearson's correlation produced highly homogeneous clusters.

mutual information, binning (MI-B), all trials, Pearson clusters



Figure C.22. Using Pearson's correlation as the distance measure, the consensus method produces twelve clusters. The results show no clear pattern and may indicate less meaningful clusters.

membership of cluster per factor by fraction of trials,
DTW distance on mutual information, binning (MI-B)



Figure C.23. For each cluster produced through DTW, the trials assigned to the cluster are grouped by experimental level in order to find the percentage of a cluster associated with each experimental level. For example, all trials assigned to cluster 2 use influence model d (random adoption) and include more of the higher density network structures.

In Figures C.23 and C.24, we conduct a "census" of the trials assigned to each cluster, with respect to the experimental design factors. For DTW, cluster 1 contains exclusively trials with the random adoption influence model (level d) and is low in trials with tree-like network structures For Pearson's correlation, cluster membership is homogeneous and indistinct.

In summary, the variation in system design studied here can produce meaningful clusters, with respect to the experimental design factors, in the response space for MI-B. This effect is achieved when using dynamic time warping as the distance measure, but not when using Pearson's correlation coefficient.

## C.3   Response variable 3 - transfer entropy, binning (TE-B)

Transfer entropy, binning (TE-B) assigns agent opinion to one of a set of equal-width bins and computes the transfer entropy (Equation 3.3) between each agent-neighbor pair, averages across the neighbors for each agent, and then averages across each agent and each replication to produce the trial-level response. Figure C.25 shows the time series of TE-B for each trial and an associated kernel density estimate (KDE)

Figure C.24. Clusters produced through Pearson's correlation are mostly undifferentiated, suggesting this distance measure is unsuitable for the response variable.

transfer entropy, binning (TE-B), all trials



Figure C.25. The time series of TE-B values for each trial are plotted on the same axes to reveal a visual cluster at the bottom of the range.

for the final time step.[2] This shows a grouping near zero transfer entropy, and that most—but not all—trials decrease over time.

## Research question 1: Which system design factors contribute most to aggregated entropy?

For this research question, we explore the one-way sensitivity of each entropy response variable to changes in the levels of individual experimental design factors. This exploration includes qualitative comparisons of RV distributions when the trial data is grouped by experimental levels and statistical tests for differences between levels. These methods support a subjective evaluation of whether an RV is sensitive

---

[2]Because transfer entropy as calculated here uses two time steps per calculation, the final time step of the simulation has no response value. We trim the data to $t = 490$ for ease of reading but observe no unusual behavior after 490.

to changes in the level of each design factor. In Table C.6, we summarize the analysis results for the current response variable for this research question.

Table C.6.
The findings for research question 1 on TE-B are summarized to support
the overall evaluation of each experimental design factor (final table row).

| | | Factor (number of levels) | | | |
|---|---|---|---|---|---|
| | $N$ (3) | structure (10) | influence model (5) | error (4) | activation (3) |
| i. | *(Main effect plot) What differences are present between the response variable distributions for each level at the final time step?* | | | | |
| | negligible | 2 or 3 patterns | 5 patterns | 2 patterns; error vs no error | 3 patterns |
| ii. | *(Grouped time series) What differences are present between the median response values over the duration of the simulation?* | | | | |
| | negligible | initial grouping by density, then partial convergence | 3 patterns; random adoption is outlier | 3 patterns; high+medium variance identical | minor early variation, then partial convergence |
| iii. | *(K-W test) Does the Kruskal-Wallace test indicate statistical differences in the response variable between each level at the final time step? (i.e., is the p-value <0.05?)* | | | | |
| | no | no | yes | yes | yes |
| iv. | *(M-W U test) How many pairs of levels are statistically different (p-value <0.05) according to the Mann-Whitney U test?* | | | | |
| | 0/3 | 3/45 | 10/10 | 5/6 | 2/3 |
| * | *(Evaluation) Is the response variable sensitive to changes in the level for the factor?* | | | | |
| | no | yes | yes | yes | yes |

Figure C.26 presents the distributions grouped by experimental level for TE-B at the final time step, $t = 490$, using a half-violin plot. Differences between distributions among the levels for a single factor qualitatively show the effect each level has on the response. For example, the distributions for population size $N$ are almost identical,

Figure C.26. Each half-violin of this design of experiments (DoE) main effect plot represents the distribution of TE-B at $t = 500$ for all trials with the corresponding level on the horizontal axis, and its median is indicated with a horizontal dash; the grand median is shown for reference. This plot suggests that $N$ is unimportant to the response variable, while influence model leads to highly varied outcomes.

so we infer that $N$ is not important (i.e., does not have a significant effect on the response variable), at least over the range of levels used in the experimental design. On the other hand, strong differences between distributions are visible for the influence model, marking it as important to TE-B.

Figure C.26 uses data for only a single time step. To reveal the effect of time on the response variable, Figures C.27-C.31 plot the medians of the grouped data over the full length of the simulation. The two inner quartiles (25th to 75th percentiles) are indicated by the shaded regions around each line. Overall, these figures reinforce the similarities and differences observed in the main effect plot.

median transfer entropy, binning (TE-B), all trials, grouped by $N$



Figure C.27. All trials are grouped by experimental level as in Figure C.26 and the groups' median response value over time is plotted. Shaded regions around each line enclose the 25th to 75th percentiles of the data. For population size $N$, these regions almost entirely overlap due to the closeness of the median lines, which reinforces the low importance of $N$ shown in the DoE main effect plot.

median transfer entropy, binning (TE-B), all trials, grouped by structure



Figure C.28. Following Figure C.27 in design, this plot shows some differentiation between network models but many are very similar. A distinct split into two groups occurs during the initial transient between lower density (lower group) and higher density (upper group) network structures.

median transfer entropy, binning (TE-B), all trials, grouped by influence model



Figure C.29. The standard model, attractive-repulsive model, and nonlinear model are closely aligned in this plot; similarity bias has a different characteristic shape, and random adoption is an extreme outlier.

median transfer entropy, binning (TE-B), all trials, grouped by error



Figure C.30. With the influence error distribution, we observe clear differences in the response variable, but the differences are very small relative to the total range of transfer entropy observed in the trials.

median transfer entropy, binning (TE-B), all trials, grouped by activation



Figure C.31. The median lines for the three activation regimes are very similar, but their variability differs, as also indicated in the main effect plot.

Table C.7.
The Kruskal-Wallace test is ran on trial-level TE-B values at $t = 490$ to test if changing the level for a factor has a statistical effect on the response value. The asterisk indicates that population size $N$ has no significant impact on TE-B.

|  | test stat | p-value |
| --- | --- | --- |
| $N$ | 0.04 | * 9.76e-01 |
| structure | 7.11 | * 6.24e-01 |
| influence model | 1284.07 | 9.41e-277 |
| error | 267.21 | 1.23e-57 |
| activation | 12.38 | 2.04e-03 |

Thus far, we have use qualitative approaches to show the effect of varying individual design factors. We now adopt a non-parametric approach to measuring differences between experimental levels, using the Kruskal-Wallace test and Mann-Whitney U test.

Based on the p-values from the Kruskal-Wallace test (Table C.7) on TE-B at $t = 490$, when the data is split into levels for population size $N$ and network structure model, the data appears to come from the same population. Practically, this suggests that varying these factors—over the levels specified in our experiment—does not have a significant effect on the response variable. This agrees with what we observe in the previous figures.

Figure C.32 aggregates the results of the Mann-Whitney U test applied to each pair of levels within a factor.

Overall, for the TE-B response variable, varying population size $N$ and agent activation regime have no real effect; the different densities of the network structure models have an effect on the response; and influence models and error distributions have distinct signatures.

Figure C.32. We use the Mann-Whitney U test to determine which levels are statistically different within each factor. The numbers in cells for the pairs with a significant test statistic ($< 0.05$) express the p-value as a percentage (e.g. 3* means $0.03 \leq$ p-value $< 0.04$). The non-significant results for $N$ are consistent with the previous findings. The low number of differences in the network structure models is unexpected, since the main effect plot showed marked differences in distribution tail length among some levels.

transfer entropy, binning (TE-B), all trials, DTW clusters



Figure C.33. Using dynamic time warping (DTW) as the distance measure between pairs of response variable time series, the consensus method produces two clusters, each highlighted here using the original time series plot (Figure C.25). The densely grouped nature of these clusters suggest a reasonable level of cluster quality.

## Research question 2: How is system design related to the response space of entropy time-series values?

Using the cluster analysis process described in C.1, trials are assigned to clusters for both DTW and Pearson's correlation. These assignments are summarized in the following figures. Both DTW and Pearson's correlation produced two clusters for TE-B (Figures C.34 and Figure C.33). (Conventional guidance says that if multiple clustering methods call for the minimum/maximum number of clusters, then the selected methods or distance metric may be unsuitable for clustering the data.) With respect to the time series plots, DTW led to rather differentiated clusters, while Pearson's correlation did not.

In Figures C.35 and C.36, we conduct a "census" of the trials assigned to each cluster, with respect to the experimental design factors. For DTW, one cluster contains exclusively trials with influence model d (random adoption), but is also low in trials with tree-like network structures. For Pearson's correlation, cluster membership is almost entirely homogeneous.

In summary, the variation in system design studied here does produce a somewhat meaningful cluster, with respect to the experimental design factors, in the response

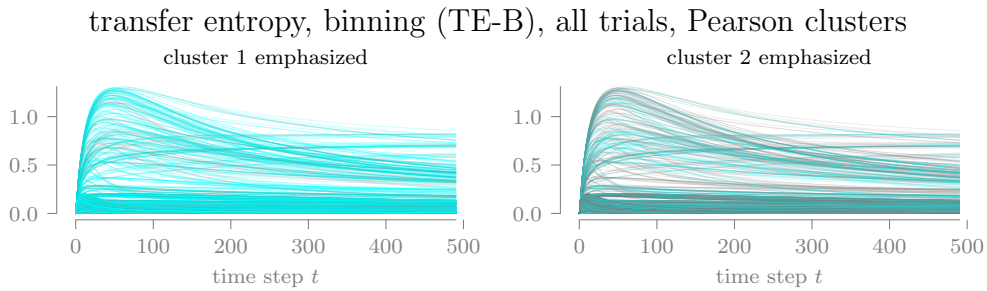transfer entropy, binning (TE-B), all trials, Pearson clusters

Figure C.34. Using Pearson's correlation as the distance measure, the consensus method produces two clusters. The results show no clear pattern and may indicate less meaningful clusters. Difference in perceived brightness is due to the z-ordering of the lines and different cluster sizes.
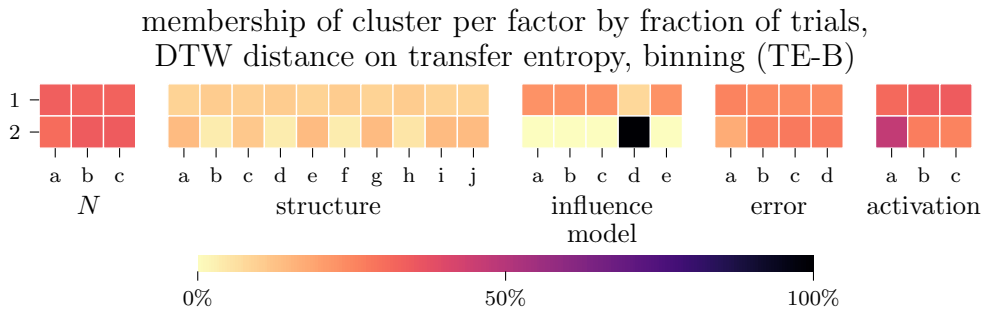


membership of cluster per factor by fraction of trials, DTW distance on transfer entropy, binning (TE-B)

Figure C.35. For each cluster produced through DTW, the trials assigned to the cluster are grouped by experimental level in order to find the percentage of a cluster associated with each experimental level. For example, all trials assigned to cluster 2 use influence model d (random adoption) and fewer trials with lower density networks (b, d, f, and h).

membership of cluster per factor by fraction of trials,
Pearson distance on transfer entropy, binning (TE-B)

Figure C.36. Clusters produced through Pearson's correlation are completely undifferentiated, suggesting this distance measure is unsuitable for the response variable.

space for TE-B. This effect is achieved when using dynamic time warping as the distance measure, but not when using Pearson's correlation coefficient. However, in both cases, only two clusters were created, which limits the usefulness of cluster analysis for this response variable.

## C.4  Response variable 4 - relative entropy, symbolic approach (RE-S)

Relative entropy, symbolic approach (RE-S) transforms each agent's sequence of opinion values into a pattern of relative orderings (e.g., Figure B.1) and computes the relative entropy (Equation 3.3) of the resulting distribution $p(x)$ with respect to the uniform distribution $q(x)$, averaging across each agent and each replication to produce the trial-level response. Figure C.37 shows the time series of RE-S for each trial and an associated kernel density estimate (KDE) for the final time step.[3] The upper extreme appears to correspond to trials where individual agent opinion converged; we use six symbols (patterns) for RE-B, so the upper limit for relative entropy with respect to the uniform distribution occurs when its opinion takes on only a single value:

$$D_X(p \parallel q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)} = 1 \log_2 \frac{1}{1/6} \approx 2.585. \tag{C.2}$$

## Research question 1:  Which system design factors contribute most to aggregated entropy?

For this research question, we explore the one-way sensitivity of each entropy response variable to changes in the levels of individual experimental design factors. This exploration includes qualitative comparisons of RV distributions when the trial data is grouped by experimental levels and statistical tests for differences between levels. These methods support a subjective evaluation of whether an RV is sensitive

---

[3]Because the symbolic method uses multiple time steps per calculation, the final time steps of the simulation have no response value. We trim the data to $t = 490$ for ease of reading.

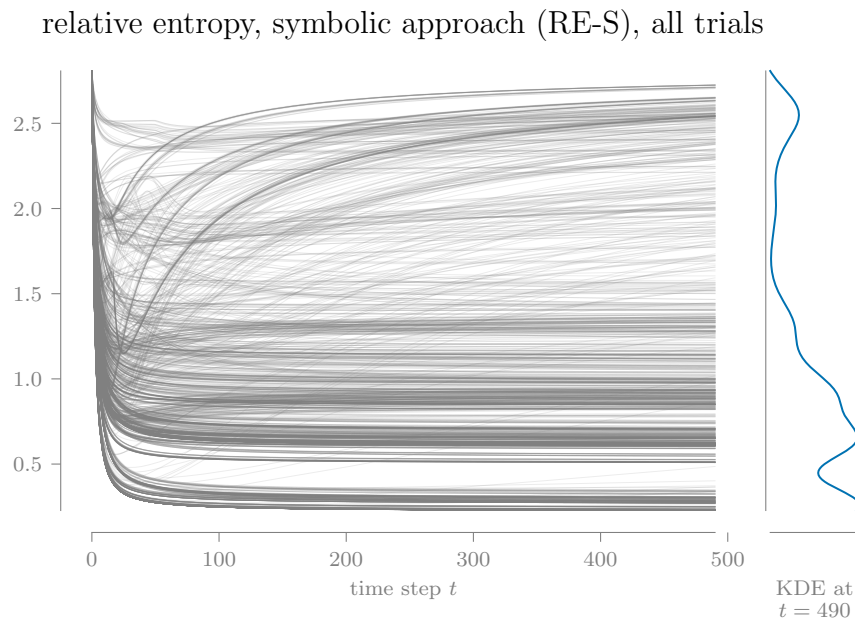relative entropy, symbolic approach (RE-S), all trials



Figure C.37. The time series of RE-S values for each trial are plotted on the same axes to reveal several visual clusters near the bottom of the range and a small grouping near the maximum value.

to changes in the level of each design factor. In Table C.8, we summarize the analysis results for the current response variable for this research question.

Table C.8.
The findings for research question 1 on RE-S are summarized to support the overall evaluation of each experimental design factor (final table row).

| | Factor (number of levels) | | | | |
|---|---|---|---|---|---|
| | $N$ (3) | structure (10) | influence model (5) | error (4) | activation (3) |
| i. | *(Main effect plot) What differences are present between the response variable distributions for each level at the final time step?* | | | | |
| | negligible | 2 patterns | 2 or 3 patterns; attractive-repulsive significantly different | 2 patterns; error vs no error | random adoption different than others |
| ii. | *(Grouped time series) What differences are present between the median response values over the duration of the simulation?* | | | | |
| | negligible | overall similar shapes; small divide affected by density | 2 patterns; attractive-repulsive distinct | no error significantly different from rest | 2 patterns |
| iii. | *(K-W test) Does the Kruskal-Wallace test indicate statistical differences in the response variable between each level at the final time step? (i.e., is the p-value <0.05?)* | | | | |
| | no | yes | yes | yes | yes |
| iv. | *(M-W U test) How many pairs of levels are statistically different (p-value <0.05) according to the Mann-Whitney U test?* | | | | |
| | 0/3 | 37/45 | 7/10 | 3/6 | 2/3 |
| * | *(Evaluation) Is the response variable sensitive to changes in the level for the factor?* | | | | |
| | no | yes | yes | yes | no |

Figure C.38 presents the distributions grouped by experimental level for RE-S at the final time step, $t = 490$, using a half-violin plot. Differences between distributions
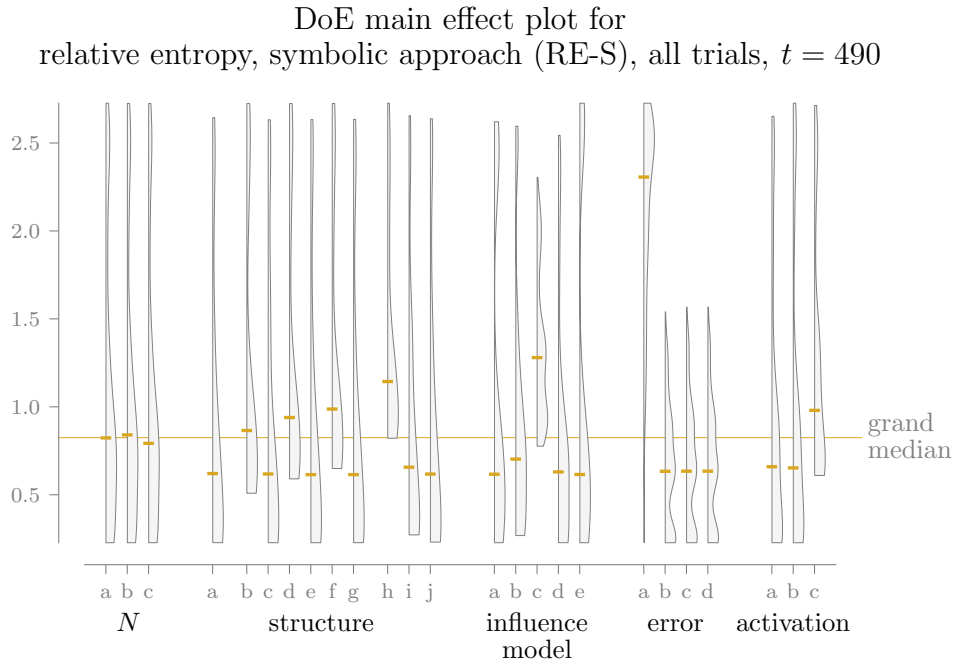
Figure C.38. Each half-violin of this design of experiments (DoE) main effect plot represents the distribution of RE-S at $t = 490$ for all trials with the corresponding level on the horizontal axis, and its median is indicated with a horizontal dash; the grand median is shown for reference. This plot suggests that $N$ is unimportant to the response variable, while the other factors lead to varied outcomes, to greater or lesser extents.

among the levels for a single factor qualitatively show the effect each level has on the response. For example, the distributions for population size $N$ are almost identical, so we infer that $N$ is not important (i.e., does not have a significant effect on the response variable), at least over the range of levels used in the experimental design. On the other hand, strong differences between distributions are visible for the influence model and error term, marking them as important to RE-S. The results for the network structure shows four levels above the grand median line, which happen to be the four network structures with lower density (Figure B.11).

Figure C.38 uses data for only a single time step. To reveal the effect of time on the response variable, Figures C.39-C.43 plot the medians of the grouped data over

median relative entropy, symbolic approach (RE-S), all trials, grouped by $N$



Figure C.39. All trials are grouped by experimental level as in Figure C.38 and the groups' median response value over time is plotted. Shaded regions around each line enclose the 25th to 75th percentiles of the data. For population size $N$, these regions almost entirely overlap due to the closeness of the median lines, which reinforces the low importance of $N$ shown in the DoE main effect plot.

the full length of the simulation. The two inner quartiles (25th to 75th percentiles) are indicated by the shaded regions around each line. Overall, these figures reinforce the similarities and differences observed in the main effect plot.

Thus far, we have use qualitative approaches to show the effect of varying individual design factors. We now adopt a non-parametric approach to measuring differences between experimental levels, using the Kruskal-Wallace test and Mann-Whitney U test.

Based on the p-values from the Kruskal-Wallace test (Table C.9) on RE-S at $t = 490$, when the data is split into levels for population size $N$, the data appears to come from the same population. Practically, this suggests that varying this factor—over the levels specified in our experiment—does not have a significant effect on the response variable. This agrees with what we observe in the previous figures.

median relative entropy, symbolic approach (RE-S), all trials, grouped by structure
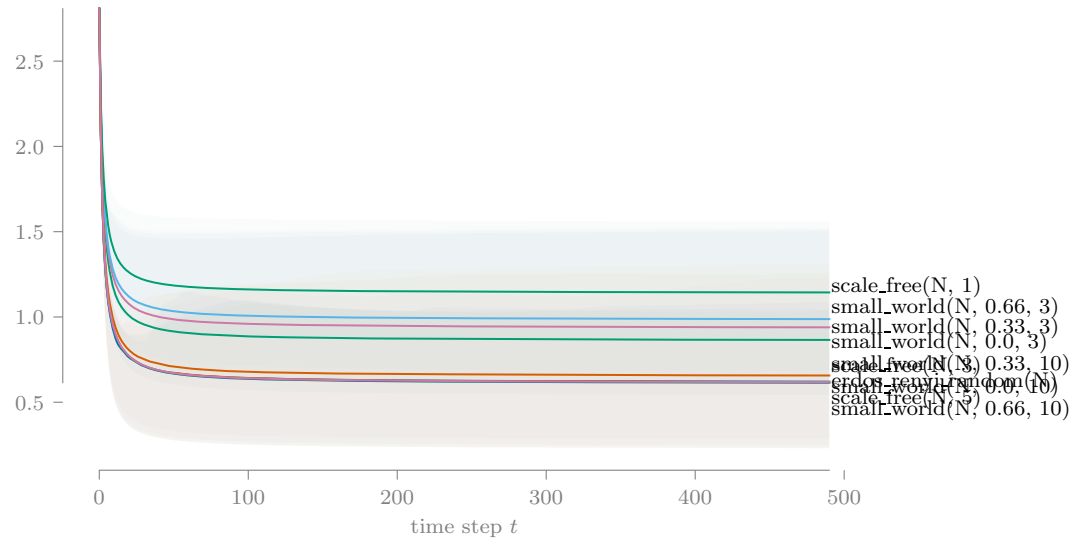


Figure C.40. Following Figure C.39 in design, this plot shows some differences in magnitude between network models but all have the same characteristic shape. Lower density structures have higher median values for RE-S.

Table C.9.

The Kruskal-Wallace test is ran on trial-level RE-S values at $t = 490$ to test if changing the level for a factor has a statistical effect on the response value. The asterisk indicates that population size $N$ has no significant impact on RE-S.

|  | test stat | p-value |
| --- | --- | --- |
| $N$ | 1.14 | * 5.65e-01 |
| structure | 234.41 | 1.92e-45 |
| influence model | 231.76 | 5.51e-49 |
| error | 859.47 | 5.46e-186 |
| activation | 159.54 | 2.26e-35 |

median relative entropy, symbolic approach (RE-S), all trials, grouped by influence model



Figure C.41. The attractive-repulsive model is clearly differentiated from the other influence models in both magnitude and shape of the median response line.

median relative entropy, symbolic approach (RE-S), all trials, grouped by error



Figure C.42. With the influence error distribution, we observe clear differences between no error term and the normally distributed error terms, while the median lines for the three normally distributed terms are indistinguishable.

median relative entropy, symbolic approach (RE-S), all trials, grouped by activation



Figure C.43. The median lines for the three activation regimes are very similar in shape, but the random regime converges to a slightly higher entropy value than the others.

Figure C.44. We use the Mann-Whitney U test to determine which levels are statistically different within each factor. The numbers in cells for the pairs with a significant test statistic ($< 0.05$) express the p-value as a percentage (e.g. 3* means $0.03 \leq$ p-value $< 0.04$). The non-significant results for $N$ are consistent with the previous findings.

Figure C.44 aggregates the results of the Mann-Whitney U test applied to each pair of levels within a factor. Level pairs for network structure that are similar (grey cells) are almost all for pairs of higher density graphs.

Overall, for the RE-S response variable, varying population size $N$ has no significant effect; lower density and higher density network structure models lead to different outcomes; trials with no error term have much higher RE-S than any trial

relative entropy, symbolic approach (RE-S), all trials, DTW clusters



Figure C.45. Using dynamic time warping (DTW) as the distance mea-
sure between pairs of response variable time series, the consensus method
produces two clusters, each highlighted here using the original time series
plot (Figure C.37). The densely grouped nature of these clusters suggest a
reasonable level of cluster quality.

with an error term; the random activation regime yields somewhat higher RE-S; and
results for each influence model, except attractive-repulsive, are quite similar.

**Research question 2: How is system design related to the response space
of entropy time-series values?**

Using the cluster analysis process described in C.1, trials are assigned to clusters
for both DTW and Pearson's correlation. These assignments are summarized in the
following figures. Both DTW and Pearson's correlation produced two clusters for
RE-S (Figures C.46 and Figure C.45). (Conventional guidance says that if multiple
clustering methods call for the minimum/maximum number of clusters, then the
selected methods or distance metric may be unsuitable for clustering the data.) With
respect to the time series plots, DTW led to rather differentiated clusters, while
Pearson's correlation did not.

In Figures C.47 and C.48, we conduct a "census" of the trials assigned to each
cluster, with respect to the experimental design factors. For DTW, one cluster con-
tains exclusively trials with error term a (no error). For Pearson's correlation, cluster
membership is almost entirely homogeneous.

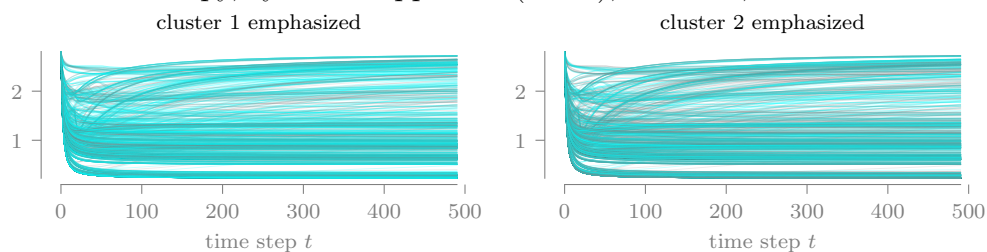relative entropy, symbolic approach (RE-S), all trials, Pearson clusters



Figure C.46. Using Pearson's correlation as the distance measure, the consensus method produces two clusters. The results show no clear pattern and may indicate less meaningful clusters.
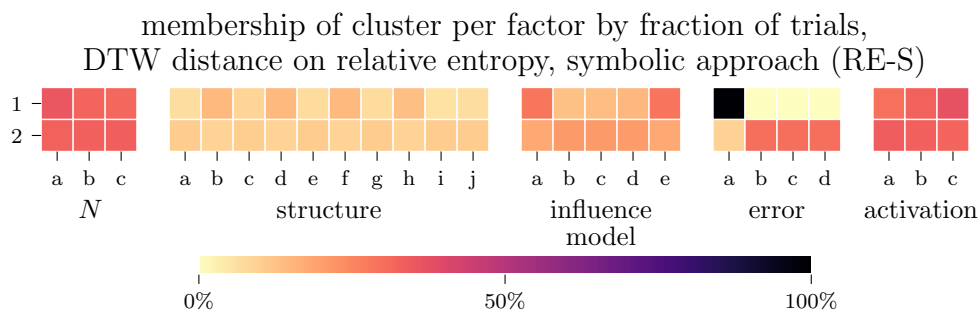


Figure C.47. For each cluster produced through DTW, the trials assigned to the cluster are grouped by experimental level in order to find the percentage of a cluster associated with each experimental level. For example, all trials assigned to cluster 2 use error term a (no error) and have a slight preference for lower density networks (b, d, f, and h).
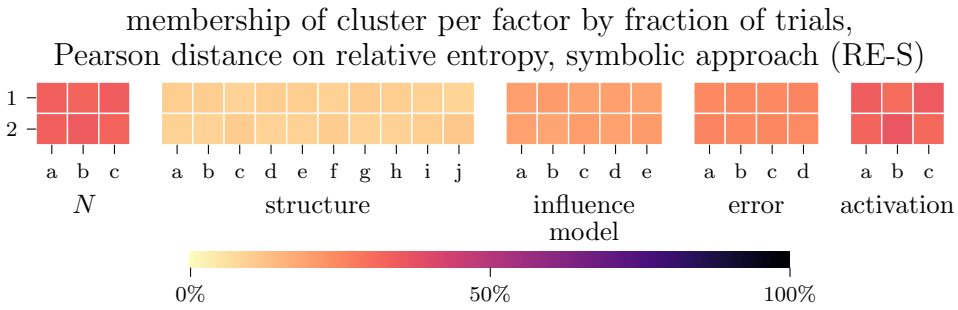
Figure C.48. Clusters produced through Pearson's correlation are completely undifferentiated, suggesting this distance measure is unsuitable for the response variable.

In summary, the variation in system design studied here does produce a somewhat meaningful cluster, with respect to the experimental design factors, in the response space for RE-S. This effect is achieved when using dynamic time warping as the distance measure, but not when using Pearson's correlation coefficient. However, in both cases, only two clusters were created, which limits the usefulness of cluster analysis for this response variable.

## C.5   Response variable 5 - mutual information, symbolic approach (MI-S)

Mutual information, symbolic approach (MI-S) transforms each agent's sequence of opinion values into a pattern of relative orderings (e.g., Figure B.1) and computes the mutual information (Equation 3.2) between each agent-neighbor pair, averages across the neighbors for each agent, and then averages across each agent and each replication to produce the trial-level response. Figure C.49 shows the time series of MI-S for each trial and an associated kernel density estimate (KDE) for the final time step.[4]

**Research question 1:  Which system design factors contribute most to aggregated entropy?**

For this research question, we explore the one-way sensitivity of each entropy response variable to changes in the levels of individual experimental design factors. This exploration includes qualitative comparisons of RV distributions when the trial data is grouped by experimental levels and statistical tests for differences between levels. These methods support a subjective evaluation of whether an RV is sensitive to changes in the level of each design factor. In Table C.10, we summarize the analysis results for the current response variable for this research question.

---

[4]Because the symbolic method uses multiple time steps per calculation, the final time steps of the simulation have no response value. We trim the data to $t = 490$ for ease of reading.

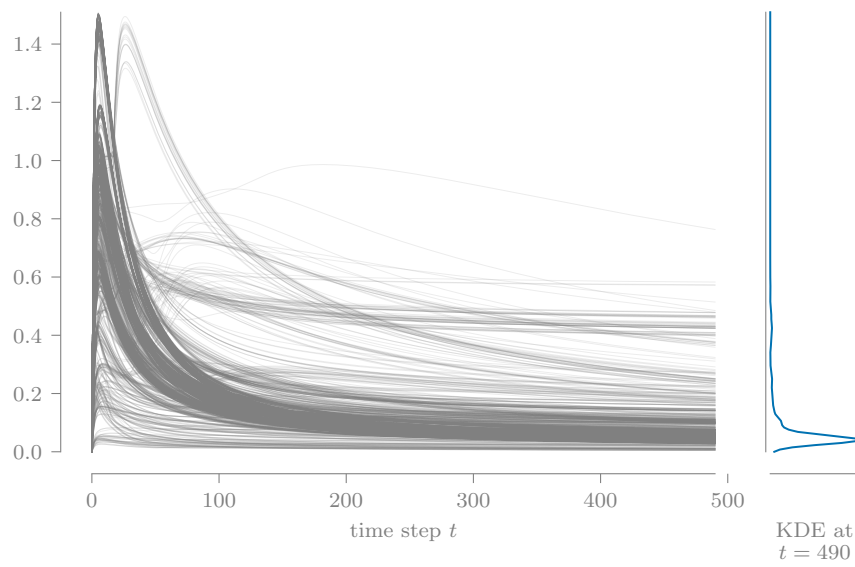mutual information, symbolic approach (MI-S), all trials



Figure C.49. The time series of MI-S values for each trial are plotted on the same axes to reveal a dense visual cluster near the bottom of the range.

Table C.10.

The findings for research question 1 on MI-S are summarized to support
the overall evaluation of each experimental design factor (final table row).

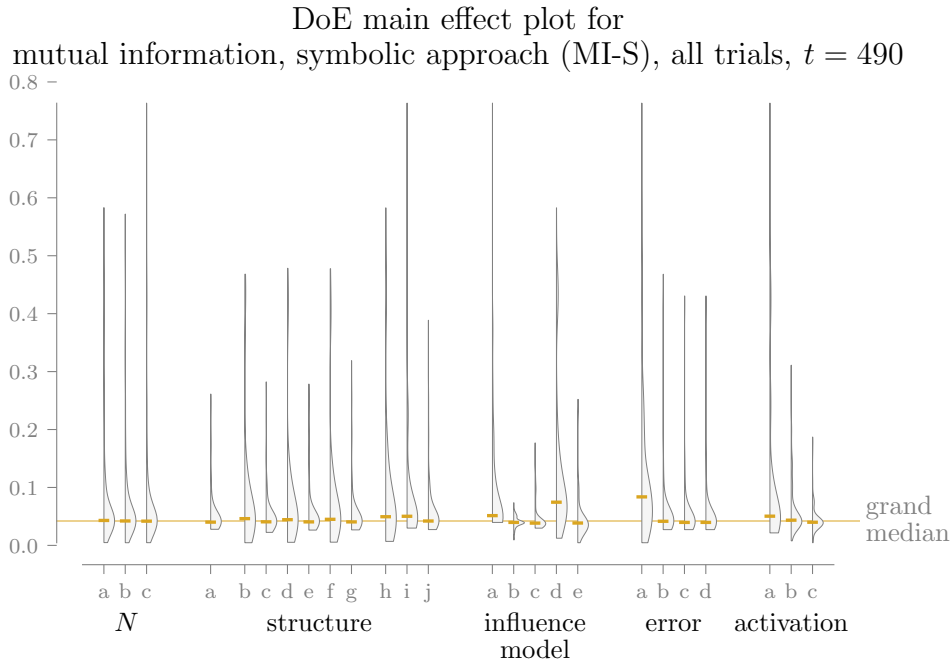| | Factor (number of levels) | | | | |
|---|---|---|---|---|---|
| | *N* (3) | structure (10) | influence model (5) | error (4) | activation (3) |
| i. | *(Main effect plot) What differences are present between the response variable distributions for each level at the final time step?* | | | | |
| | $N = 10k$ has much longer tail | 2 or 3 patterns | 5 patterns | 2 patterns; error vs no error | 3 patterns |
| ii. | *(Grouped time series) What differences are present between the median response values over the duration of the simulation?* | | | | |
| | negligible | initial grouping by density, then partial convergence | initial separation, then partial convergence | no error somewhat different from rest | minor early variation, then partial convergence |
| iii. | *(K-W test) Does the Kruskal-Wallace test indicate statistical differences in the response variable between each level at the final time step? (i.e., is the p-value <0.05?)* | | | | |
| | no | yes | yes | yes | yes |
| iv. | *(M-W U test) How many pairs of levels are statistically different (p-value <0.05) according to the Mann-Whitney U test?* | | | | |
| | 0/3 | 12/45 | 9/10 | 5/6 | 3/3 |
| * | *(Evaluation) Is the response variable sensitive to changes in the level for the factor?* | | | | |
| | no | yes | yes | yes | yes |

Figure C.50. Each half-violin of this design of experiments (DoE) main effect plot represents the distribution of MI-S at $t = 490$ for all trials with the corresponding level on the horizontal axis, and its median is indicated with a horizontal dash; the grand median is shown for reference. This plot suggests that $N$ is fairly unimportant to the response variable, while changes in influence model lead to more varied outcomes.

Figure C.50 presents the distributions grouped by experimental level for MI-S at the final time step, $t = 490$, using a half-violin plot. Differences between distributions among the levels for a single factor qualitatively show the effect each level has on the response. For example, the distributions for population size $N$ are almost identical (except in the length of the upper tails), so we infer that $N$ is not very important (i.e., does not have a significant effect on the response variable), at least over the range of levels used in the experimental design. On the other hand, strong differences between distributions are visible for the influence model, marking it as important to MI-S.

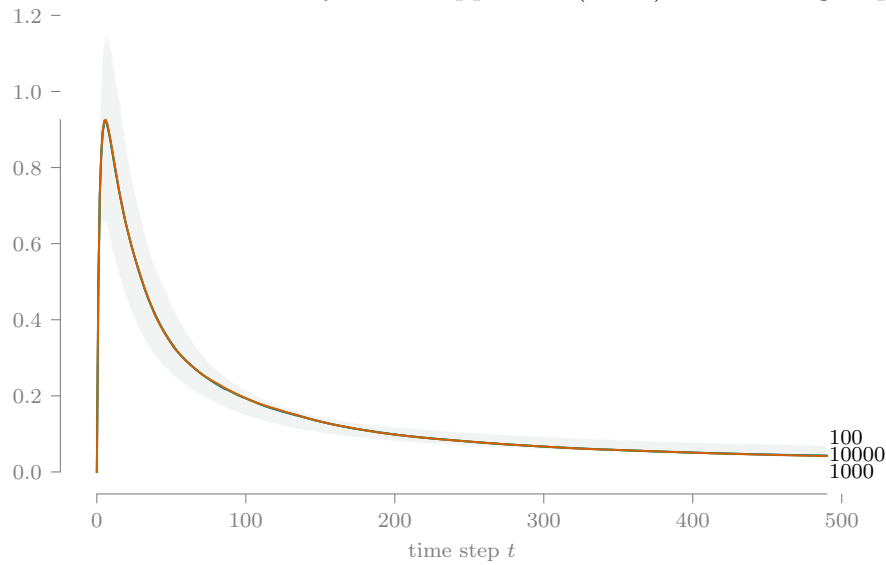median mutual information, symbolic approach (MI-S), all trials, grouped by $N$



Figure C.51. All trials are grouped by experimental level as in Figure C.50 and the groups' median response value over time is plotted. Shaded regions around each line enclose the 25th to 75th percentiles of the data. For population size $N$, these regions almost entirely overlap due to the closeness of the median lines, which reinforces the low importance of $N$ shown in the DoE main effect plot.

Figure C.50 uses data for only a single time step. To reveal the effect of time on the response variable, Figures C.51-C.55 plot the medians of the grouped data over the full length of the simulation. The two inner quartiles (25th to 75th percentiles) are indicated by the shaded regions around each line. Overall, these figures reinforce the similarities and differences observed in the main effect plot.

Thus far, we have use qualitative approaches to show the effect of varying individual design factors. We now adopt a non-parametric approach to measuring differences between experimental levels, using the Kruskal-Wallace test and Mann-Whitney U test.

Based on the p-values from the Kruskal-Wallace test (Table C.11) on MI-S at $t = 490$, when the data is split into levels for population size $N$, the data appears to come from the same population. Practically, this suggests that varying this factor—

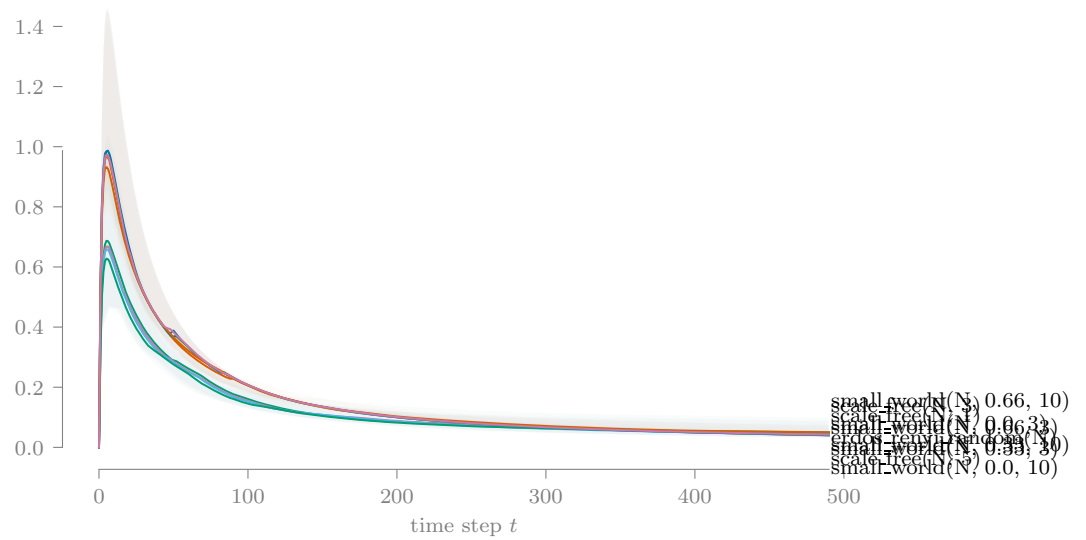median mutual information, symbolic approach (MI-S), all trials, grouped by structure



Figure C.52. Following Figure C.51 in design, an interesting split into two groups occurs during the initial transient between lower density (lower group) and upper density (upper group) networks, but all converge by the end.

median mutual information, symbolic approach (MI-S), all trials, grouped by influence model
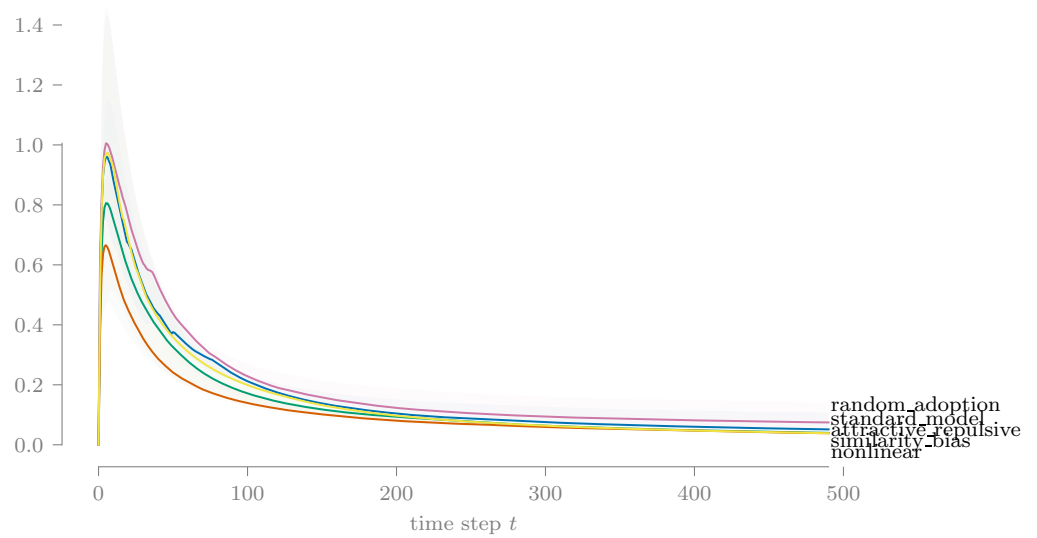


Figure C.53. The various influence models show some initial differences, but all nearly converge by the end of the run.

median mutual information, symbolic approach (MI-S), all trials, grouped by error
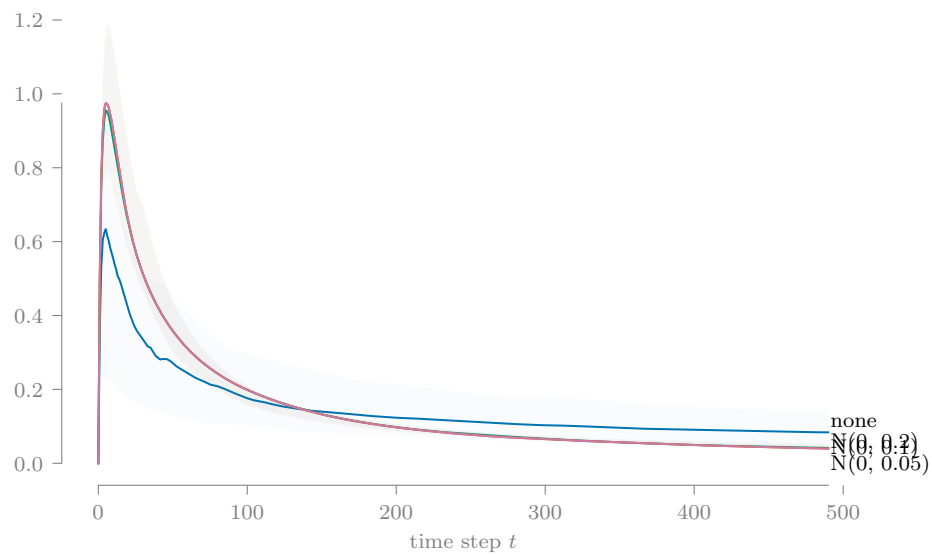


Figure C.54. With the influence error distribution, we observe clear differences between no error term and the normally distributed error terms, while the median lines for the three normally distributed terms are indistinguishable.

median mutual information, symbolic approach (MI-S), all trials, grouped by activation
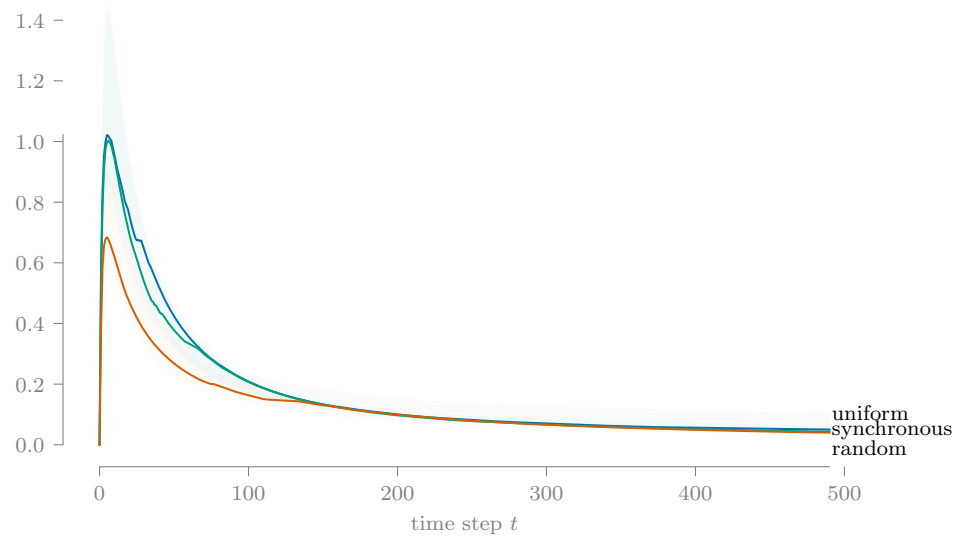


Figure C.55. The median lines for the three activation regimes experience initial differences but rapidly converge.

Table C.11.
The Kruskal-Wallace test is ran on trial-level MI-S values at $t = 490$ to test if changing the level for a factor has a statistical effect on the response value. The asterisk indicates that population size $N$ has no significant impact on MI-S.

|  | test stat | p-value |
| --- | --- | --- |
| $N$ | 0.39 | * 8.20e-01 |
| structure | 21.18 | 1.18e-02 |
| influence model | 720.43 | 1.30e-154 |
| error | 154.74 | 2.49e-33 |
| activation | 112.10 | 4.52e-25 |

over the levels specified in our experiment—does not have a significant effect on the response variable. This agrees with what we observe in the previous figures.

Figure C.56 aggregates the results of the Mann-Whitney U test applied to each pair of levels within a factor. Only network structure levels i and j (the two higher-density scale-free networks) show any differentiation. Similarity bias (level b) and attractive-repulsive (c) influence models test as similar here, despite their very different functional forms; however, they do both use the magnitude of opinion difference to vary the strength of the interaction.

Overall, for the MI-S response variable, varying population size $N$ and agent activation regime has no significant effect; density differences in network structure models have greater effects early in a run but very minor effects in the long term; and each influence model has a different response distribution but nearly identical median responses in the long term.
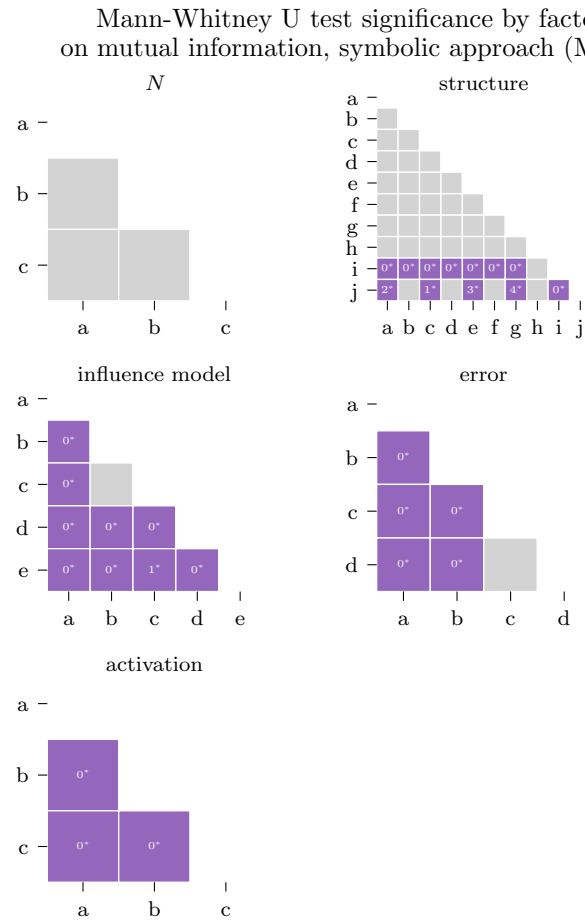
Figure C.56. We use the Mann-Whitney U test to determine which levels are statistically different within each factor. The numbers in cells for the pairs with a significant test statistic ($< 0.05$) express the p-value as a percentage (e.g. 3* means $0.03 \leq$ p-value $< 0.04$). The non-significant results for $N$ are consistent with the previous findings.

**Research question 2: How is system design related to the response space of entropy time-series values?**

Using the cluster analysis process described in C.1, trials are assigned to clusters for both DTW and Pearson's correlation. These assignments are summarized in the following figures. DTW for MI-S produced five clusters (Figure C.57), while Pearson's correlation produced two clusters (Figure C.58). With respect to the time series plots, DTW led to somewhat differentiated clusters, while Pearson's correlation did not.

In Figures C.59 and C.60, we conduct a "census" of the trials assigned to each cluster, with respect to the experimental design factors. For DTW, cluster 2 contains exclusively trials with the synchronous activation regime (level a), while cluster 3 contains only trials with no error term (level a); cluster 5 omits random activation and lower density network structures. For Pearson's correlation, cluster membership is almost entirely homogeneous.

In summary, the variation in system design studied here does produce a somewhat meaningful cluster, with respect to the experimental design factors, in the response space for MI-S. This effect is achieved when using dynamic time warping as the distance measure, but not when using Pearson's correlation coefficient.

## C.6 Response variable 6 - transfer entropy, symbolic approach (TE-S)

Transfer entropy, symbolic approach (TE-S) transforms each agent's sequence of opinion values into a pattern of relative orderings (e.g., Figure B.1) and computes the transfer entropy (Equation 3.3) between each agent-neighbor pair, averages across the neighbors for each agent, and then averages across each agent and each replication to produce the trial-level response. Figure C.61 shows the time series of TE-S for each trial and an associated kernel density estimate (KDE) for the final time step.[5]

---

[5]Because the symbolic method uses multiple time steps per calculation, the final time steps of the simulation have no response value. We trim the data to $t = 490$ for ease of reading.

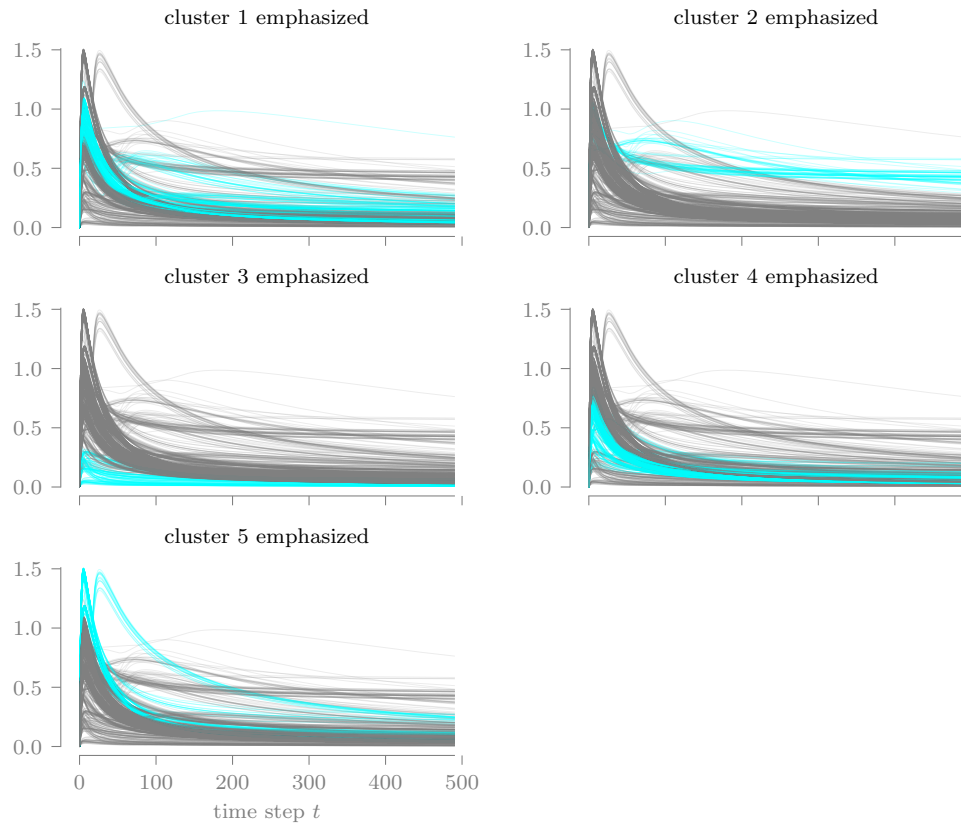mutual information, symbolic approach (MI-S), all trials, DTW clusters



Figure C.57. Using dynamic time warping (DTW) as the distance measure between pairs of response variable time series, the consensus method produces two clusters, each highlighted here using the original time series plot (Figure C.49). The densely grouped nature of these clusters suggest a moderate level of cluster quality.
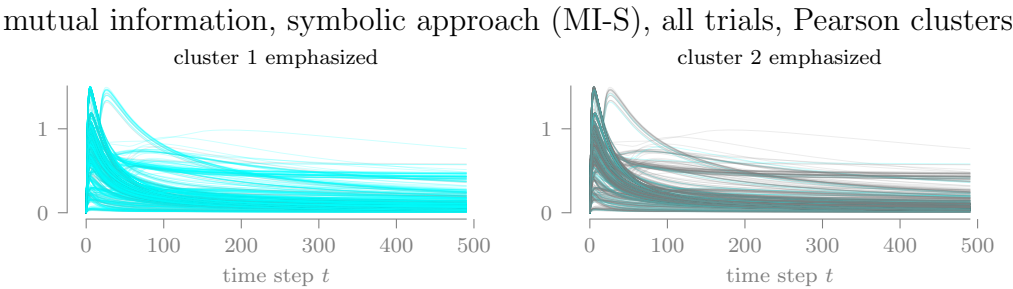
mutual information, symbolic approach (MI-S), all trials, Pearson clusters



Figure C.58. Using Pearson's correlation as the distance measure, the consensus method produces two clusters. The results show no clear pattern and may indicate less meaningful clusters.
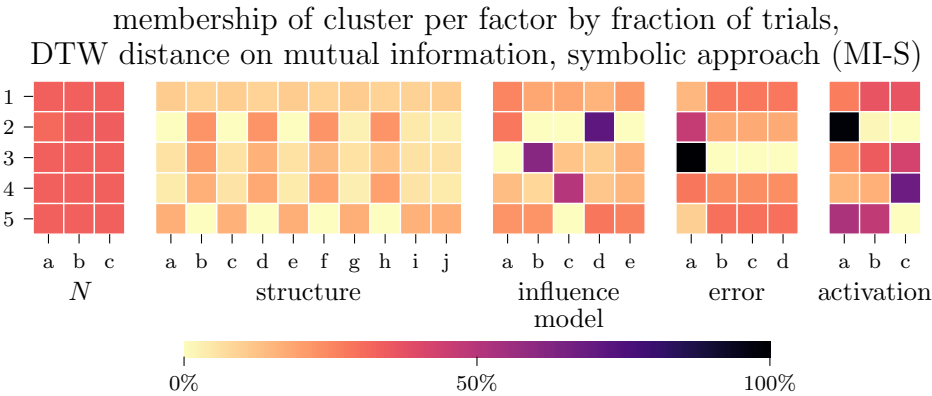


Figure C.59. For each cluster produced through DTW, the trials assigned to the cluster are grouped by experimental level in order to find the percentage of a cluster associated with each experimental level. For example, all trials assigned to cluster 3 use error term a (no error).

membership of cluster per factor by fraction of trials,
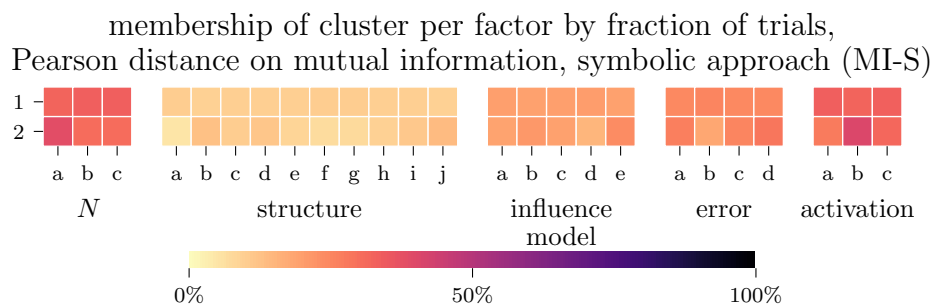Pearson distance on mutual information, symbolic approach (MI-S)



Figure C.60. Clusters produced through Pearson's correlation are completely undifferentiated, suggesting this distance measure is unsuitable for the response variable.

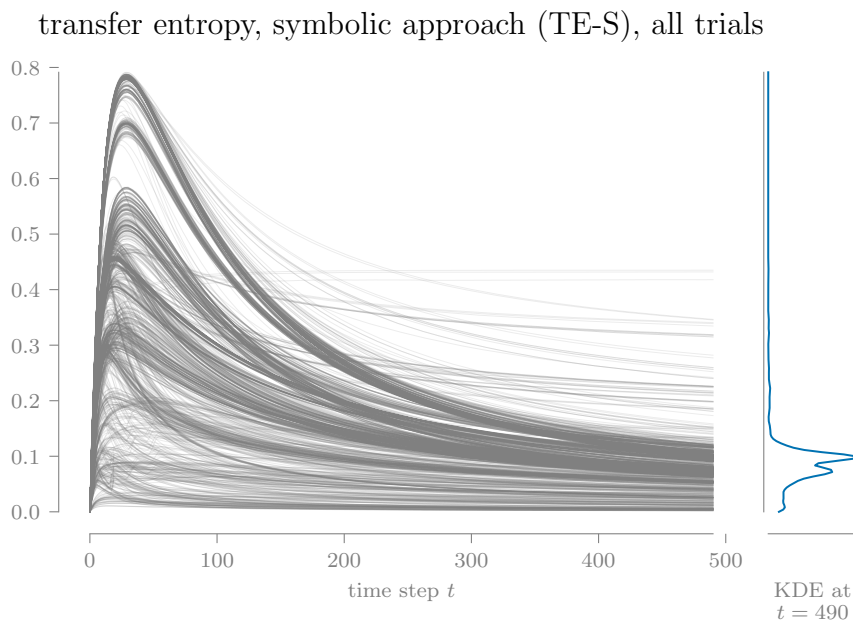transfer entropy, symbolic approach (TE-S), all trials



Figure C.61. The time series of TE-S values for each trial are plotted on the same axes to reveal a dense visual cluster near the bottom of the range.

**Research question 1: Which system design factors contribute most to aggregated entropy?**

For this research question, we explore the one-way sensitivity of each entropy response variable to changes in the levels of individual experimental design factors. This exploration includes qualitative comparisons of RV distributions when the trial data is grouped by experimental levels and statistical tests for differences between levels. These methods support a subjective evaluation of whether an RV is sensitive to changes in the level of each design factor. In Table C.12, we summarize the analysis results for the current response variable for this research question.

Figure C.62 presents the distributions grouped by experimental level for TE-S at the final time step, $t = 490$, using a half-violin plot. Differences between distributions among the levels for a single factor qualitatively show the effect each level has on the response. For example, the distributions for population size $N$ are almost identical (except in the length of the upper tails), so we infer that $N$ is not very important (i.e., does not have a significant effect on the response variable), at least over the range of levels used in the experimental design. On the other hand, strong differences between distributions are visible for the influence model, marking it as important to TE-S. While the shape of the distributions for the lower density networks (b, d, f, and h) are inconsistent, they do happen to be the four levels with medians below the grand median.

Figure C.62 uses data for only a single time step. To reveal the effect of time on the response variable, Figures C.63-C.67 plot the medians of the grouped data over the full length of the simulation. The two inner quartiles (25th to 75th percentiles) are indicated by the shaded regions around each line. Overall, these figures reinforce the similarities and differences observed in the main effect plot.

Thus far, we have use qualitative approaches to show the effect of varying individual design factors. We now adopt a non-parametric approach to measuring differences

Table C.12.

The findings for research question 1 on TE-S are summarized to support
the overall evaluation of each experimental design factor (final table row).

| | $N$ (3) | structure (10) | influence model (5) | error (4) | activation (3) |
|---|---|---|---|---|---|
| | | | Factor (number of levels) | | |
| i. | *(Main effect plot) What differences are present between the response variable distributions for each level at the final time step?* | | | | |
| | negligible | 4 or 5 patterns | 3 patterns | 2 patterns; error vs no error | significant differences between all levels |
| ii. | *(Grouped time series) What differences are present between the median response values over the duration of the simulation?* | | | | |
| | negligible | overall similar shapes; small divide affected by density | initial separation between levels, then partial convergence | no error significantly different from rest | random adoption initially different, then partial convergence |
| iii. | *(K-W test) Does the Kruskal-Wallace test indicate statistical differences in the response variable between each level at the final time step? (i.e., is the p-value <0.05?)* | | | | |
| | no | yes | yes | yes | yes |
| iv. | *(M-W U test) How many pairs of levels are statistically different (p-value <0.05) according to the Mann-Whitney U test?* | | | | |
| | 0/3 | 25/45 | 10/10 | 3/6 | 3/3 |
| * | *(Evaluation) Is the response variable sensitive to changes in the level for the factor?* | | | | |
| | no | yes | yes | yes | yes |

DoE main effect plot for
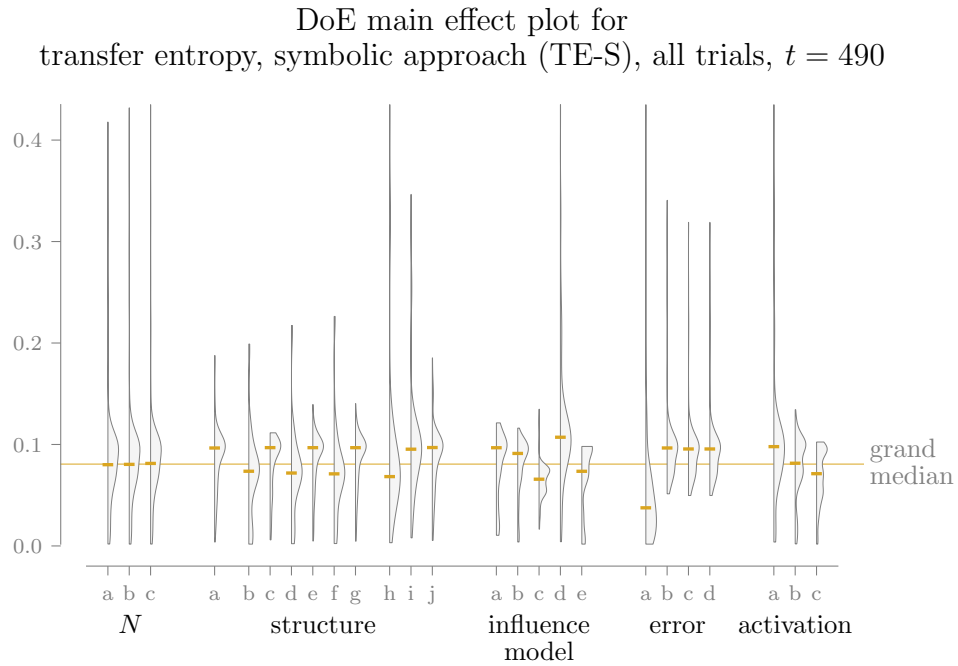transfer entropy, symbolic approach (TE-S), all trials, $t = 490$

Figure C.62. Each half-violin of this design of experiments (DoE) main effect plot represents the distribution of TE-S at $t = 490$ for all trials with the corresponding level on the horizontal axis, and its median is indicated with a horizontal dash; the grand median is shown for reference. This plot suggests that $N$ is fairly unimportant to the response variable, while changes in influence model lead to more varied outcomes.

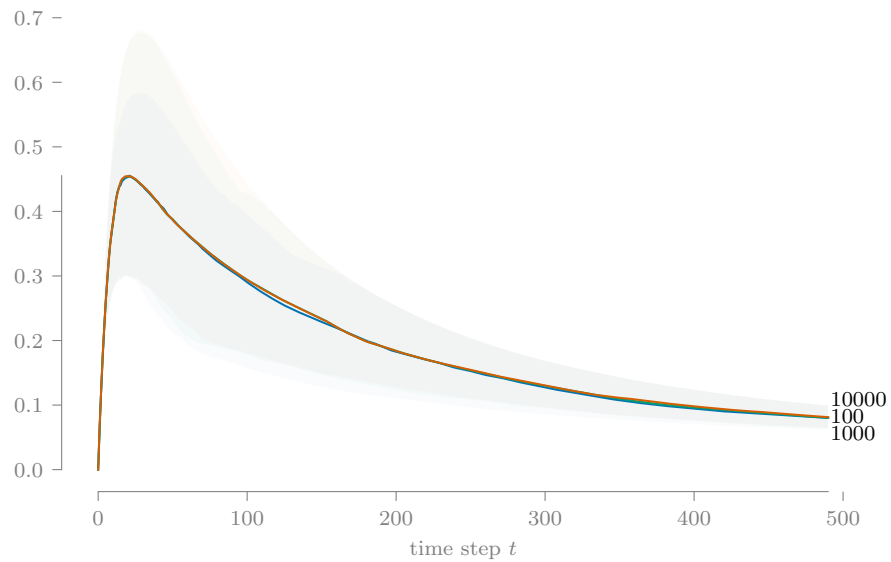median transfer entropy, symbolic approach (TE-S), all trials, grouped by $N$

Figure C.63. All trials are grouped by experimental level as in Figure C.62 and the groups' median response value over time is plotted. Shaded regions around each line enclose the 25th to 75th percentiles of the data. For population size $N$, these regions almost entirely overlap due to the closeness of the median lines, which reinforces the low importance of $N$ shown in the DoE main effect plot.

median transfer entropy, symbolic approach (TE-S), all trials, grouped by structure

small_world(N, 0.66, 10)
small_world(N, 0.33, 10)
scale-free(N, 3)
scale-free(N, 10)
small_world(N, 0.33, 3)
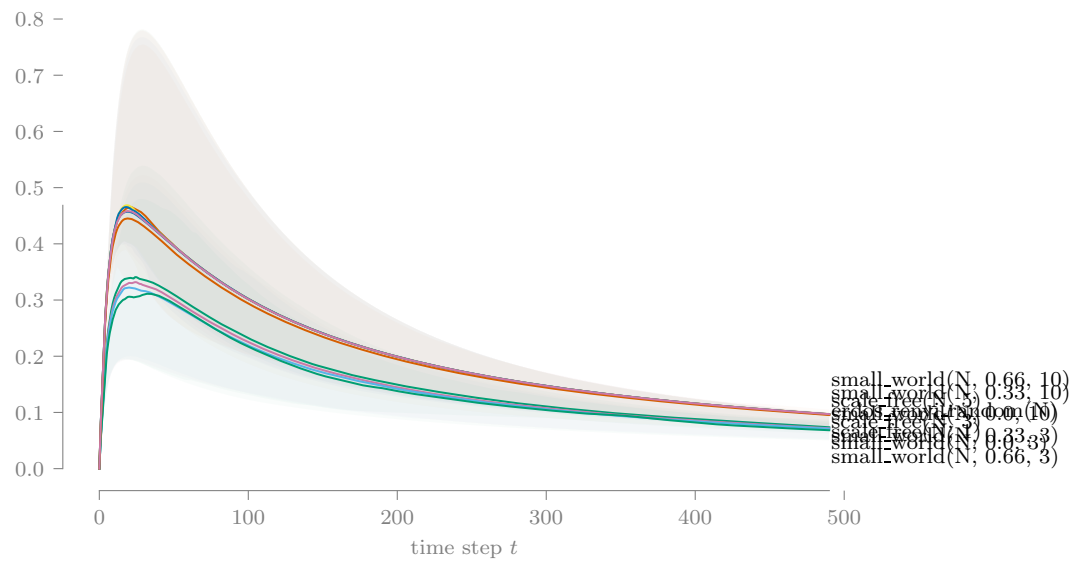small_world(N, 0.66, 3)

time step $t$

Figure C.64. Following Figure C.63 in design, this plot shows two groupings of network models: lower density (lower group) and higher density (upper group).

median transfer entropy, symbolic approach (TE-S), all trials, grouped by influence model
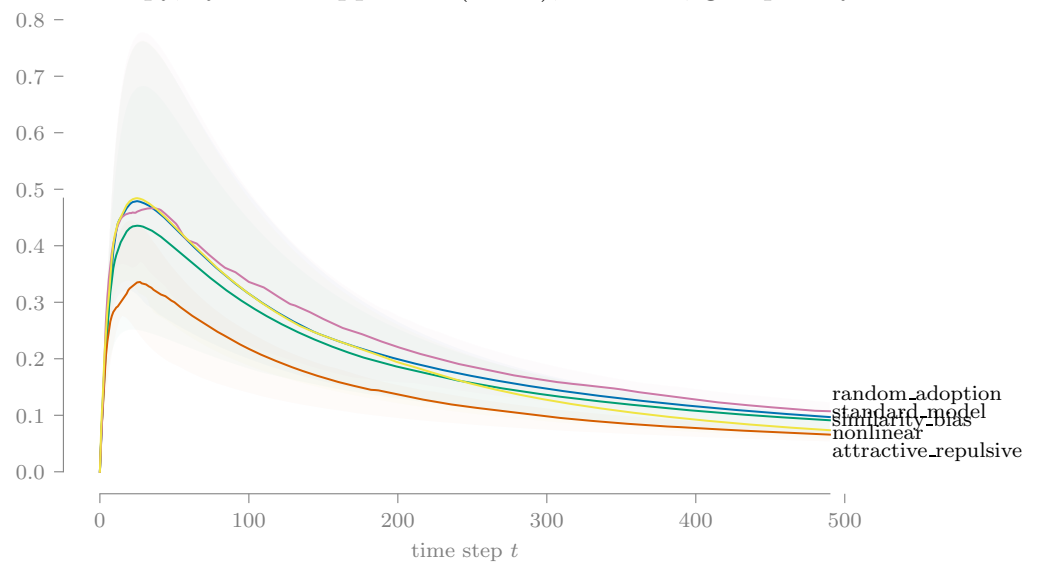


Figure C.65. The various influence models show some initial differences, but begin to converge by the end of the run.

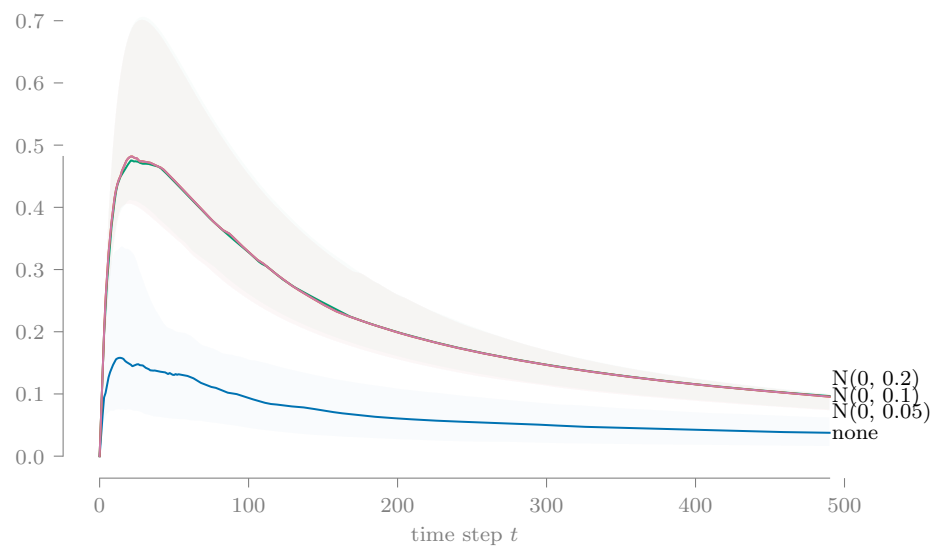median transfer entropy, symbolic approach (TE-S), all trials, grouped by error

Figure C.66. With the influence error distribution, we observe clear differences between no error term and the normally distributed error terms, while the median lines for the three normally distributed terms are indistinguishable.

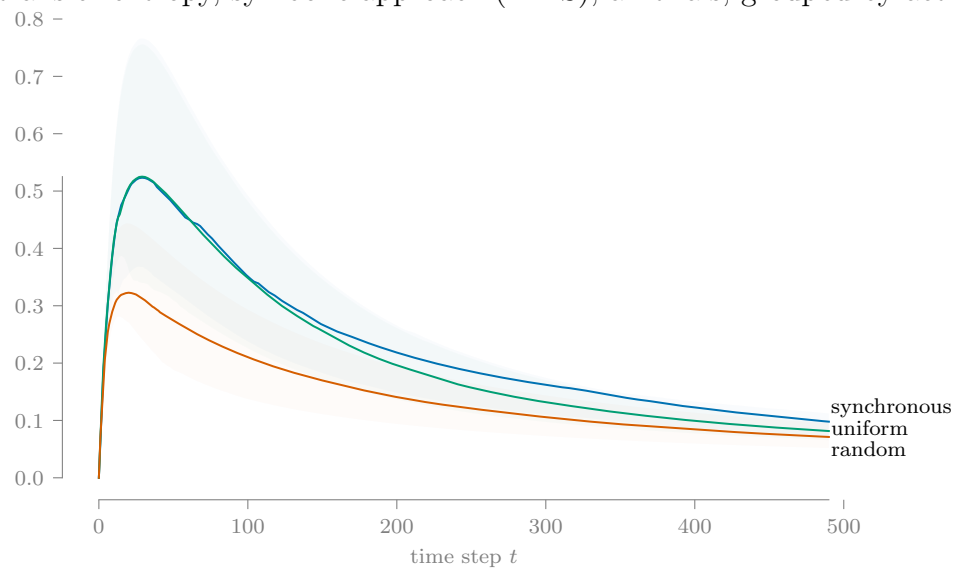median transfer entropy, symbolic approach (TE-S), all trials, grouped by activation



Figure C.67. The median lines for the three activation regimes experience initial differences but begin to converge at the end of the run.

Table C.13.
The Kruskal-Wallace test is ran on trial-level TE-S values at $t = 490$ to test
if changing the level for a factor has a statistical effect on the response value.
The asterisk indicates that population size $N$ has no significant impact on
TE-S.

|  | test stat | p-value |
|---|---|---|
| $N$ | 0.23 | * 8.91e-01 |
| structure | 133.57 | 2.17e-24 |
| influence model | 458.17 | 7.41e-98 |
| error | 558.63 | 9.35e-121 |
| activation | 206.61 | 1.36e-45 |

between experimental levels, using the Kruskal-Wallace test and Mann-Whitney U
test.

Based on the p-values from the Kruskal-Wallace test (Table C.13) on TE-S at
$t = 490$, when the data is split into levels for population size $N$, the data appears to
come from the same population. Practically, this suggests that varying this factor—
over the levels specified in our experiment—does not have a significant effect on the
response variable. This agrees with what we observe in the previous figures.

Figure C.68 aggregates the results of the Mann-Whitney U test applied to each
pair of levels within a factor. Structure pairs are primarily different if they are have
different relative network densities (lower vice higher density, with respect to all
network structures in the experiment).

Overall, for the TE-S response variable, varying population size $N$ has no signifi-
cant effect, while varying all other factors produce noticeable differences in response
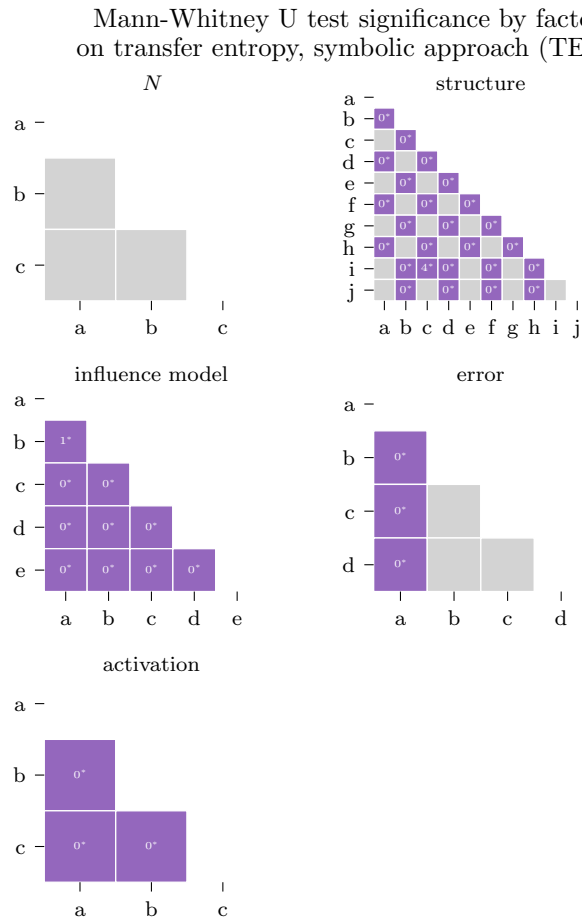distributions.

Figure C.68. We use the Mann-Whitney U test to determine which levels are statistically different within each factor. The numbers in cells for the pairs with a significant test statistic ($< 0.05$) express the p-value as a percentage (e.g. $3^*$ means $0.03 \leq$ p-value $< 0.04$). The non-significant results for $N$ are consistent with the previous findings.

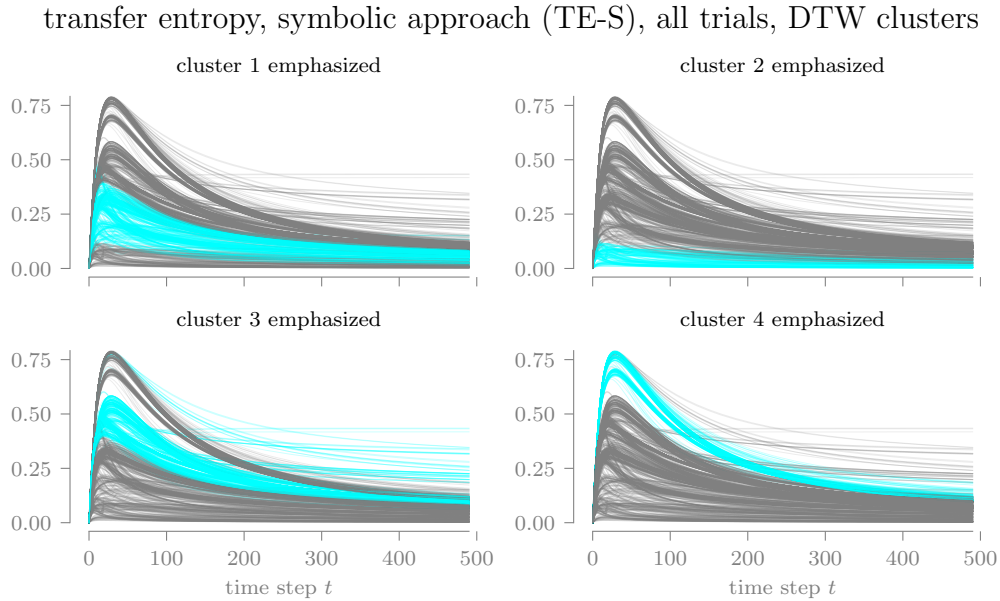transfer entropy, symbolic approach (TE-S), all trials, DTW clusters



Figure C.69. Using dynamic time warping (DTW) as the distance measure between pairs of response variable time series, the consensus method produces four clusters, each highlighted here using the original time series plot (Figure C.61). The densely grouped nature of these clusters suggest a moderate level of cluster quality.

## Research question 2: How is system design related to the response space of entropy time-series values?

Using the cluster analysis process described in C.1, trials are assigned to clusters for both DTW and Pearson's correlation. These assignments are summarized in the following figures. DTW for TE-S produced four clusters (Figure C.69), while Pearson's correlation produced three clusters (Figure C.70). With respect to the time series plots, DTW led to somewhat differentiated clusters, while Pearson's correlation did not.

In Figures C.71 and C.72, we conduct a "census" of the trials assigned to each cluster, with respect to the experimental design factors. For DTW, cluster 2 contains exclusively trials with no error term (level a) and is somewhat higher in tree-like network structures; cluster 1 is strong in the random activation regime (level c), while

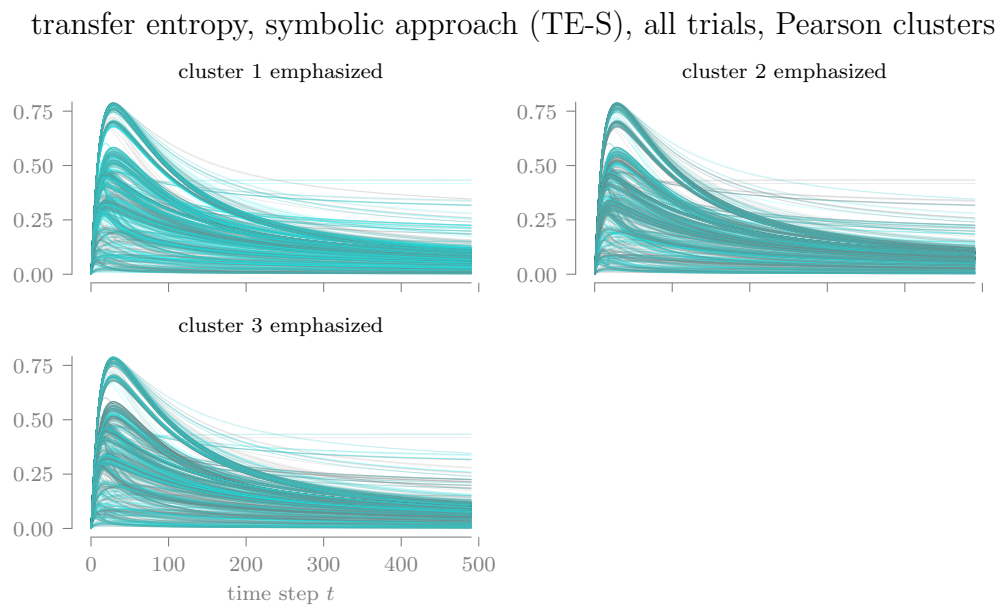transfer entropy, symbolic approach (TE-S), all trials, Pearson clusters

Figure C.70. Using Pearson's correlation as the distance measure, the consensus method produces three clusters. The results show no clear pattern and may indicate less meaningful clusters.
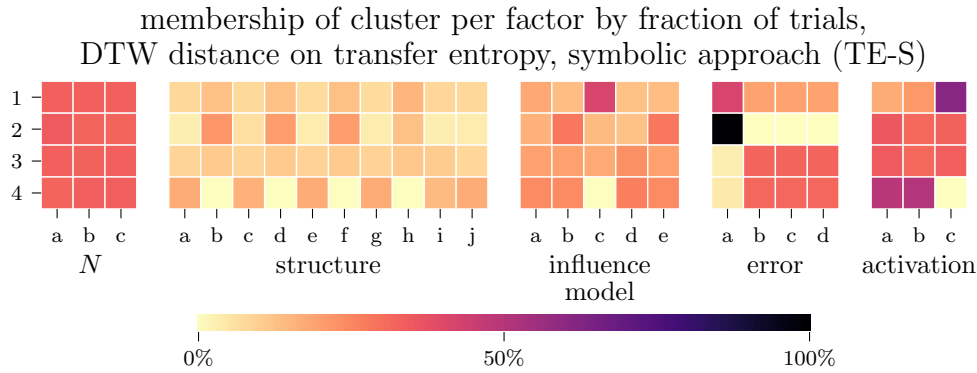
Figure C.71. For each cluster produced through DTW, the trials assigned to the cluster are grouped by experimental level in order to find the percentage of a cluster associated with each experimental level. For example, all trials assigned to cluster 2 use error term a (no error).



Figure C.72. Clusters produced through Pearson's correlation are completely undifferentiated, suggesting this distance measure is unsuitable for the response variable.

cluster 4 is strong in the other regimes and happens to omit lower density networks. For Pearson's correlation, cluster membership is almost entirely homogeneous.

In summary, the variation in system design studied here can produce meaningful clusters, with respect to the experimental design factors, in the response space for TE-S. This effect is achieved when using dynamic time warping as the distance measure, but not when using Pearson's correlation coefficient.

# D. SUPPLEMENTAL MATERIAL FOR CHAPTER 4

This appendix provides supplemental information for the article, "Effects of nonhomogeneous agents in social influence networks on system-level entropy."

## D.1  Experimental design

The original experimental design in [91] is reduced here to take advantage of their findings. Population size $N$ is fixed at $N = 1000$ for this experiment, because $N$ was found to be unimportant to their entropy analysis; thus $N$ is not a design factor in the current work. We delete three of the original ten network structure models due to them having highly similar response variable behavior as other structure models in the design. The nonlinear influence model performed very similarly to the standard model, so we omit the former. For the gaussian influence error term, $N(0, \sigma = 0.1)$ and $N(0, 0.2)$ are nearly indistinguishable in the original results, so we delete the latter. Finally, results for the uniform activation regime were very similar to at least one of the other two regimes across all response variables and is deleted from the design.

## D.2  Comparison between homogeneous and nonhomogeneous trials

As a first analysis step, the scenarios are compared to each other and the homogeneous base case by visualizing their raw RV time series (e.g., Figure D.1). Scenarios 2 and 3 omit certain design factors present in the base design, so they are compared to the appropriate subsets of the base case trials. Each scenario shows qualitatively similar behavior in the time series plots for every RV, with respect to the profile and visual distribution of the data.

Table D.1.
The experimental design from [91] is reduced to four factors and 168 total trials based on insights from their results. This design is the *base case* upon which the nonhomogeneous scenarios are constructed.

| structure | influence model | error | activation |
|---|---|---|---|
| Erdős-Rényi random | standard model | none | synchronous |
| small-world(0.0, 3) | similarity bias | $N(0, 0.05)$ | random |
| small-world(0.0, 10) | attractive-repulsive | $N(0, 0.10)$ | |
| small-world(0.66, 3) | random adoption | | |
| small-world(0.66, 10) | | | |
| scale-free(1) | | | |
| scale-free(5) | | | |

transfer entropy, binning (TE-B)



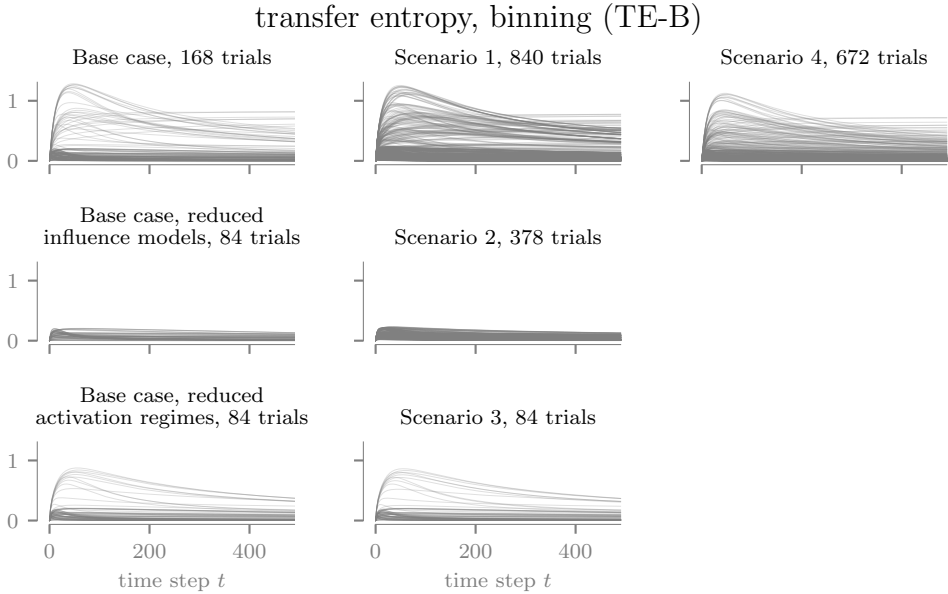Figure D.1. The raw RV data for the scenarios and the base case have similar profiles and visual distributions. Each line plots a single trial. Scenarios 2 and 3 omit some design factors from the base design and are each compared to a subset of base case trials. TE-B is shown here to emphasize the importance of comparing scenarios to an appropriate subset of homogeneous trials, but similar results appear for each RV.

Table D.2.
Design factor importance for each scenario and response variable is indicated with a $y$ (is important) or $n$ (is not important); dashes mark factors that are not applicable for a scenario. Importance is based on qualitative analyses of the grouped medians plots (e.g., Figure 4.3). The scenario-specific design factors are fraction (of population) uninformed (Scenario 1), fraction Concord-type & fraction left-biased (2), and fraction stubborn (4).

| Factor | *Scenario* 1 | 2 | 3 | 4 | base case | | *Scenario* 1 | 2 | 3 | 4 | base case |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **RE-B** | | | | | | **RE-S** | | | | | |
| structure | y | y | y | y | y | | y | y | y | y | y |
| influence model | y | - | y | y | y | | y | - | y | y | y |
| error | y | y | y | y | y | | y | y | y | y | y |
| activation | n | y | - | y | n | | y | y | - | y | y |
| scenario-specific | y | y,y | - | y | - | | y | n,n | - | y | - |
| | | | | | | | | | | | |
| **MI-B** | | | | | | **MI-S** | | | | | |
| structure | y | y | y | y | y | | n | n | n | n | n |
| influence model | y | - | y | y | y | | n | - | n | n | n |
| error | y | y | y | y | y | | y | n | n | n | y |
| activation | y | y | - | y | n | | n | n | - | n | n |
| scenario-specific | y | y,y | - | y | - | | n | n,n | - | n | - |
| | | | | | | | | | | | |
| **TE-B** | | | | | | **TE-S** | | | | | |
| structure | y | y | y | y | y | | y | y | y | y | y |
| influence model | y | - | y | y | y | | y | - | y | y | y |
| error | y | y | y | y | y | | y | y | y | y | y |
| activation | n | n | - | n | n | | y | y | - | y | y |
| scenario-specific | y | y,n | - | n | - | | n | n,n | - | y | - |

A selection of grouped medians plot observations for the scenario-specific design factors serve to illustrate the impact of those factors on the RVs (Figures D.2-D.4). Since each of the four scenario-specific factors are continuous variables, the median lines change between levels in a rather smooth way, unlike some of those for the categorical design factors (e.g., Figure 4.3).



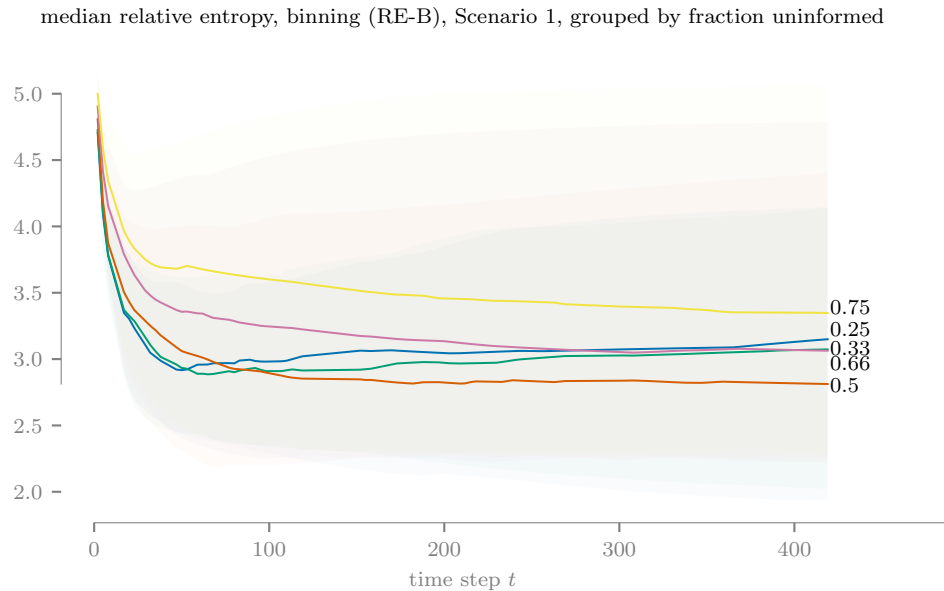median relative entropy, binning (RE-B), Scenario 1, grouped by fraction uninformed

Figure D.2. For RE-B, the median lines grouped by the fraction of initially uninformed agents in Scenario 1 for values below 0.5 show different behavior than those at 0.5 or greater. The separation between lines and the presence of different behaviors suggest this design factor to be important to RE-B.

## D.3 Cluster analysis for identifying scenario from response variable

The nonhomogeneity of the agent population may cause the response variables to behave in a distinctive way (e.g., Scenario 2 in Figure D.1). Specifically, the scenario used to generate the data may be identifiable using cluster analysis. As a test of this idea, we apply dynamic time warping (DTW) to the time series data for one response variable (RE-B) and use hierarchical clustering to assign trials to clusters. DTW is an

median relative entropy, binning (RE-B), Scenario 2, grouped by fraction concord

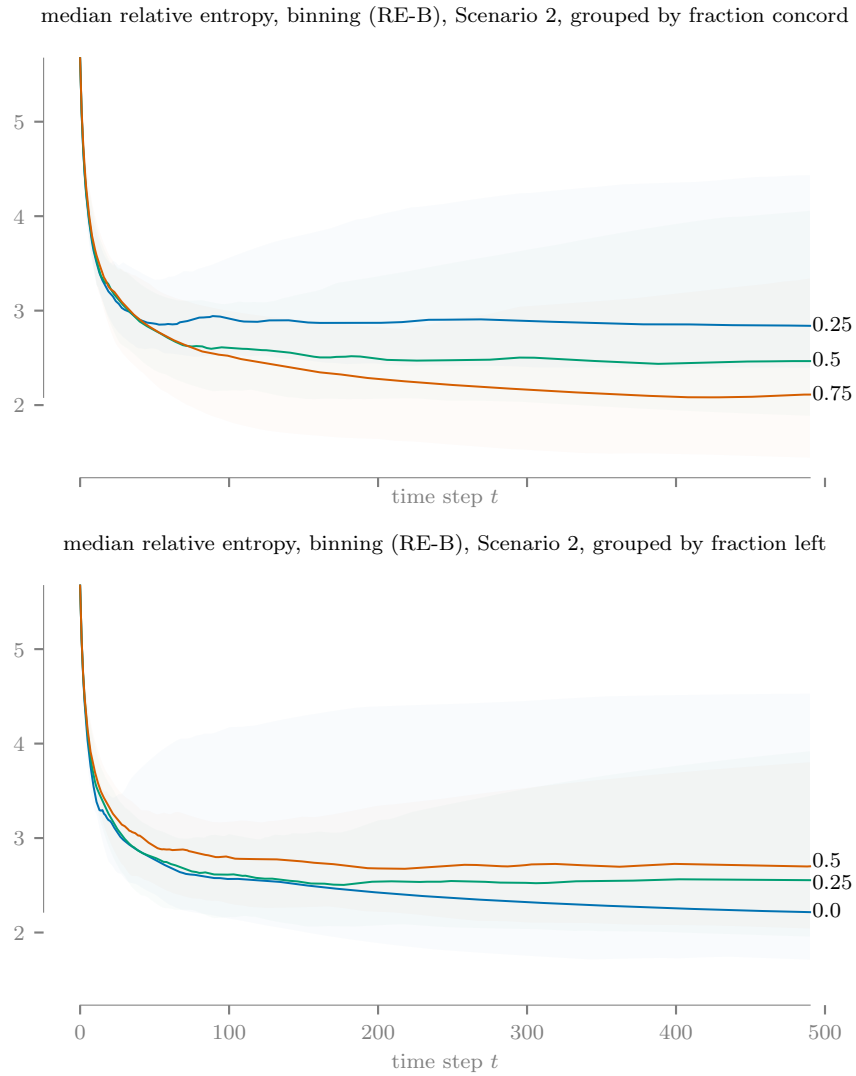median relative entropy, binning (RE-B), Scenario 2, grouped by fraction left

Figure D.3. Increasing the fraction of Concord-type agents (using the similarity bias influence model) in Scenario 2 decreases the median RE-B, while increasing the fraction of left-biased agents increases median RE-B.

median mutual information, binning (MI-B), Scenario 4, grouped by fraction stubborn
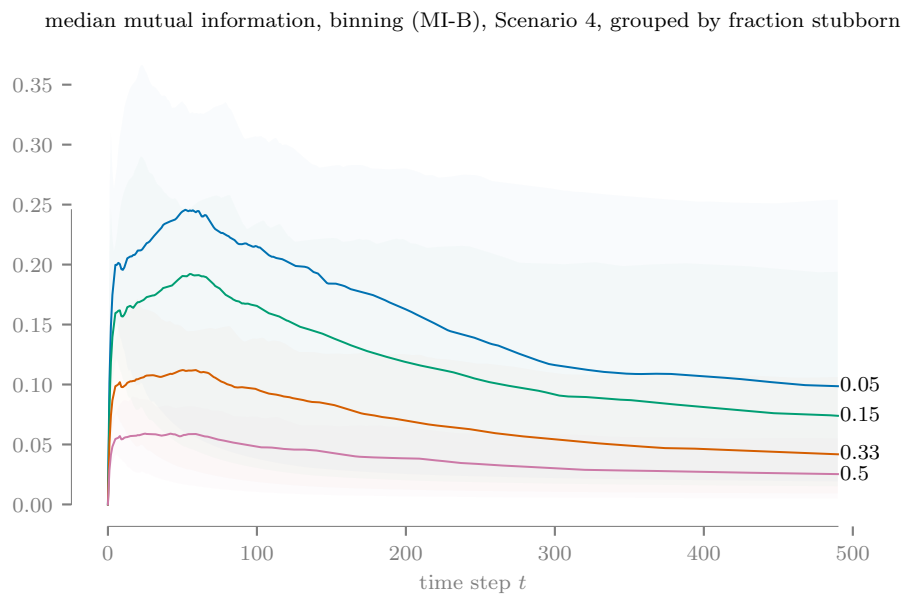
Figure D.4. Median MI-B decreases in Scenario 4 as the fraction of stubborn agents increases, since more stubborn agents leads to less information flow between agents. This design factor is important to MI-B.

appropriate dissimilarity measure for comparing time series data, while hierarchical clustering is selected due to its familiarity.

This clustering approach performs poorly overall, achieving a best-case identification accuracy of only 0.32. (Cluster labels are permuted to maximize scenario identification accuracy, since the clustering algorithm has no concept of scenario number.) The confusion matrix of the clustering results illustrates the accuracy of cluster assignment (Figure D.5). There, we see Scenario 4 (stubborn agents) is the most distinctive, with nearly half of its trials assigned to a single cluster, followed by Scenario 2 (concord/partial antagonism). The classification rates for Scenarios 2 and 4 also differ the most from those for the base case. The other scenarios show far less differentiation, both independently and relative to the base case. Some of this poor performance may attributable to the mismatch between actual trial counts and cluster size [121], summarized in Table D.3.
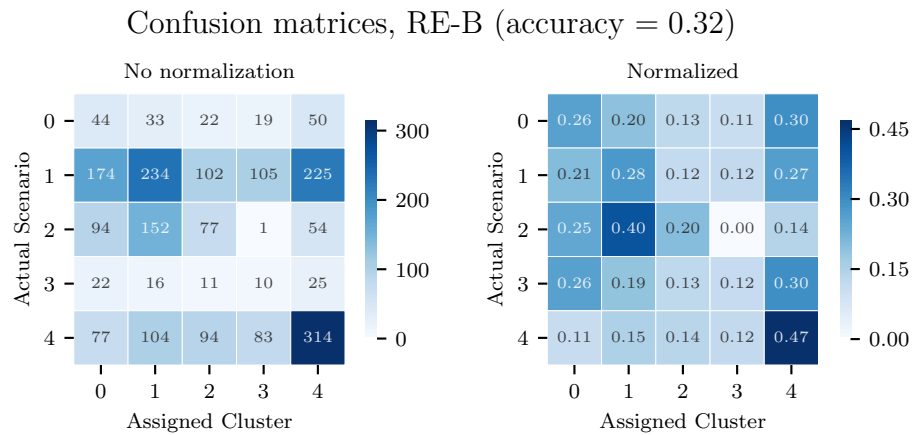


Figure D.5. (Left) The confusion matrix shows the number of trials for each scenario assigned to each cluster. The homogeneous base case is labeled as scenario/cluster 0. Cluster analysis is performed using dynamic time warping (DTW) and hierarchical clustering. DTW computes a dissimilarity value between the RE-B time series data for each pair of trials across all scenarios. (Right) The same data is row-normalized to show the fraction of a scenario's trials assigned to each cluster.

Table D.3.
The assigned cluster paired to Scenario 4 is nearly the correct size (though less than half of those trials are correctly identified); those for other scenarios differ greatly.

| Scenario | 1 | 2 | 3 | 4 | base case |
|---|---|---|---|---|---|
| Assigned cluster size | 539 | 306 | 218 | 668 | 411 |
| Actual trial count | 840 | 378 | 84 | 672 | 168 |
| Assigned : Actual | 0.64 | 0.81 | 2.60 | 0.99 | 2.45 |

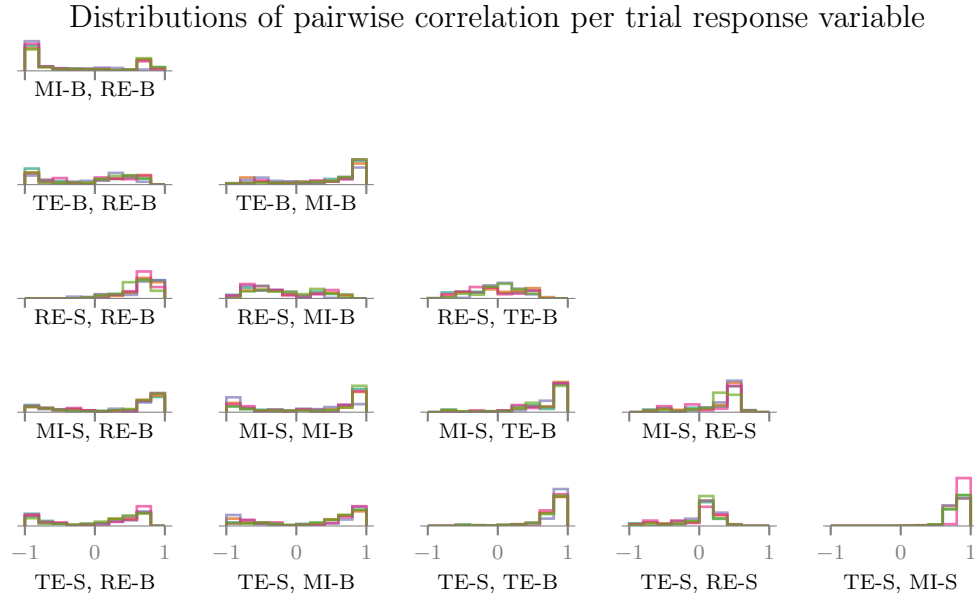Distributions of pairwise correlation per trial response variable



Figure D.6. MI and TE are somewhat positively correlated with each other
and not with RE. For each trial, we compute the correlation between the
time series of each response variable. The five lines in each subplot show the
density of correlation values for each scenario, and the base case, revealing
correlation patterns for the responses (i.e., each line outlines a histogram).
For example, TE-S and MI-S are very positively correlated for most trials.
The densities for each scenario exhibit the same general trend for each RV
pair, suggesting that the relationships among RVs are not strongly affected
by network homogeneity.

## D.4   Comparison between entropy measures

Finally, we compute the correlation between each pair of RV time series for each
trial and construct histograms on this data for each RV pair (Figure D.6). As in the
original study, some RV pairs are positively correlated for many trials, while other
pairs show a more even distribution. The correlation histograms show little variation
across scenarios, so network homogeneity and scenario design does not appear to have
a strong effect on the relationship between RVs for a single trial.

VITA

# Michael J. GAREE

*Air Force officer with a PhD in industrial engineering* m.garee@gmail.com

| | | | |
|---|---|---|---|
| May 2020 | PhD in industrial engineering | GPA 3.91 | Purdue University |
| Mar 2014 | MS in operations research | GPA 3.94 | Air Force Institute of Technology |
| May 2008 | BS in physics & mathematics | GPA 3.21 | Ohio Northern University |

## Military service    Edwards Air Force Base, CA

2014-2017    As interim division chief and operations analyst for the initial operational test and evaluation of the F-35 Joint Strike Fighter, I led accreditation efforts for modeling and simulation products totaling $450M. I also performed planning and analysis of live-fire weapons effectiveness testing.

NATO Kosovo Forces Headquarters, Kosovo

2016    For the Plans & Policy Branch while on deployment, I led the biannual review of manpower requirements for the 4000+ person NATO mission in Kosovo. I also provided technical editing on all reports created by the branch.

Nellis Air Force Base, NV

2010-2012    As a junior operations analyst, I created analysis products for test and evaluation of electronic warfare tactics, techniques, and procedures. I also jointly led the testing and development of a novel training aid for aircrews operating in environments with electronic jamming.

## Scientific contributions

[invited paper & talk]    M.J. Garee, W.K.V. Chan, and H. Wan, "Regression-based Social Influence Networks and the Linearity of Aggregated Belief," Proc. Winter Simulation Conference (2018)

M.J. Garee, R.R. Hill, D.K. Ahner, and G. Czarnecki, "Fragment capture simulation for MANPADS test arena optimization," Journal of Simulation (2017)

M.J. Garee, R.R. Hill, and B. Russell, "Fragment Capture Simulation for Missile Blast Test Optimization," Proc. IIE Annual Conference (2014)

R.R. Hill, D. Ahner, and M.J. Garee, "Using simulation to examine live-fire test configurations," Proc. Winter Simulation Conference (2014)

T.E. Sheridan, K.D. Wells, M.J. Garee, and A.C. Herrick, "Theoretical and experimental study of elliptical Debye clusters," Journal of Applied Physics (2007)

Posters

M.J. Garee and T.E. Sheridan, "Computational study of acoustic solitary waves in 2D complex plasma," APS Ohio Sections Spring Meeting (2008)

M.J. Garee and T.E. Sheridan, "Computational study of nonlinear waves in 2d complex plasma," APS Ohio Sections Fall Meeting (2007)

M.J. Garee and T.E. Sheridan, "Simulation of solitons in two-dimensional complex plasma," APS Ohio Sections Spring Meeting (2007)

K.D. Wells, M.J. Garee, A.C. Herrick, and T.E. Sheridan, "Breathing oscillation of elliptical Debye cluster," APS Meeting (2006)

M.J. Garee and T.E. Sheridan, "Sizing melamine-formaldehyde microspheres using an electron microscope," APS Ohio Sections and MAAPT Joint Spring Meeting (2006)

# Others

Reviewer for Transactions on Modeling and Computer Simulation

Program Committee member for Winter Simulation Conference 2020

Inductee in Tau Beta Pi national engineering honor society (2019)

Software Engineering (Development) Certificate from the Air Force Institute of Technology (2016)

Inductee in Omega Rho operations research honor society (2014)

Inductee in Sigma Pi Sigma physics honor society (2008)

Eagle Scout, Boy Scouts of America (2004)