SANDIA REPORT
SAND2022-14662
Printed September 2022

Sandia National Laboratories

A Model of Narrative Reinforcement on a Dual-Layer Social Network

Benjamin Freixas Emery

with technical consultation from
Christina Ting,
Jared Gearhart,
& J. Derek Tucker

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185
Livermore, California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods

ABSTRACT

Widespread integration of social media into daily life has fundamentally changed the way society communicates, and, as a result, how individuals develop attitudes, personal philosophies, and worldviews. The excess spread of disinformation and misinformation due to this increased connectedness and streamlined communication has been extensively studied, simulated, and modeled. Less studied is the interaction of many pieces of misinformation, and the resulting formation of attitudes. We develop a framework for the simulation of attitude formation based on exposure to multiple cognitions. We allow a set of cognitions with some implicit relational topology to spread on a social network, which is defined with separate layers to specify online and offline relationships. An individual's opinion on each cognition is determined by a process inspired by the Ising model for ferromagnetism. We conduct experimentation using this framework to test the effect of topology, connectedness, and social media adoption on the ultimate prevalence of and exposure to certain attitudes.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background

Although widely considered to be a modern phenomenon, disinformation has played a critical role in conspiracy theories and hate movements throughout history. The rise of the Nazis came in conjunction with widespread blame for Germany's economic downturn on Jewish Germans. Similarly, American racists in the Jim Crow era used falsified crime reports to justify lynchings. Today, many conspiracy theories with active believers make use of the same scapegoating of vulnerable factions of society as the historical examples.

While structural commonalities can be seen between historical and modern conspiratorial hate movements, the environment in which these spread has substantially changed within the past few decades, due to the emergence of the internet and social media. The removal of physical distance as a barrier to communication has amplified the total flux of information, including potentially false information, hateful messages, or incitements of violence. With the consequences of misinformation proliferation and its acceleration becoming increasingly apparent and deadly, it is critical to the wellbeing of society that these phenomena are well understood, as are potential methods of mediation. Pursuit of this understanding can and has taken the form of retrospective study, cognitive experimentation, and prospective modeling.

1.2. Existing research

The body of research surrounding social influence has been expanding for the past 100 years, and recent advancements in computing technology have allowed for the validation and disproving of the many psychological hypotheses regarding the phenomenon. As recently as the mid 2000's, we've seen generalized frameworks for modeling social influence on network structures [1–4]. These generalized models have in turn been used to uncover universal behaviors of influence cascades on networks [5–8]. With varying levels of validity, social influence theory has been applied to a set of real-world phenomena spanning from smoking to housing segregation [9, 10]. Adding to these areas of study, the past five years have seen the emergence of great interest in the echo-chamber-aided-proliferation of misinformation online, which has been studied empirically and synthetically [11–17].

With advancements made in understanding how and why misinformation spreads across our collective consciousness, the next step is to address the way that multiple misinformative articles, posts, and memes build and reinforce larger narratives. These narratives can serve as motivation for individuals to commit acts of terror without the need for a centralized leader or group, making accountability for such indoctrination nearly impossible. This question has been investigated in the past few years. In 2018, racial violence in Sri Lanka was traced to misinformation propagating on Facebook and Whatsapp [18]. Researchers have compared the narratives of false conspiracy theories to those of actual conspiracies, developed methods of detecting cultural schema in text data, and built algorithms to represent and extract narratives from event maps [19–21]. Frameworks have been created

for simulating belief interaction and agenda-setting, and some researchers have even modeled interaction between spreading beliefs, and found that given a sufficiently complex relationship between these diffusants, polarization emerges universally [20, 22].

This particular work will apply modified methods adopted from Dalege et al to model the interaction between beliefs as an Ising process [23, 24]. This approach lends itself well to misinformation, as it only requires two states per belief, which simplifies computation.

1.3. Focus of study

In this work, we introduce a simple framework for simulating the spread and interaction of information on a social network in such a way that the duality of online and offline relationships is explicitly included. This places an SIR-type contagion model on a dual-layer network topology, where the two layers can be interpreted as online and offline social connections. Critically, it allows for interaction between multiple beliefs, using the Ising model for ferromagnetism to capture a preference for cognitive consistency.

While the work presently serves as an introduction to a modeling and simulation tool, we wish to demonstrate potential use of this framework as part of the aforementioned introduction. We do this by conducting a simple parameter sweep and observe how the results deviate due to changes in these parameters. We discuss further experiments possible with this framework and potential expansions of the model.

2. METHODS

2.1. Model construction

2.1.1. Cognitive network belief model

We establish the weighted and undirected "cognitive network" $C = \{V_c, E_c\}$, which defines the reinforcement structure of a connected narrative. The network's nodes V_c are a set of cognitions, which one may imagine as statements which can either be believed or disbelieved by an agent. The edges E_c are reinforcement relationships between these cognitions. In essence, an edge represents a cognitive consistency between the statements, so it is more consistent for two connected cognitions to be both believed or both disbelieved. Such a cognitive network exists as an attribute of every agent within this model.

We use the mechanism for cognitive state adoption introduced by Dalege et al [23, 24]. Each cognitive node $v_{c,i} \in V_c$ has a value χ_i, which may be either 1 or -1, corresponding to spin up or spin down of a particle on a ferromagnetic lattice. We define χ to be the set of

all χ_i values for all nodes in the cognitive network, with each possible vector value of χ having Hamiltonian energy

$$H(\chi) = -\sum_i \tau_i \chi_i - \sum_{i,j} \omega_{i,j} \chi_i \chi_j,$$

where $\tau_i \in [-1, 1]$ is a predisposition toward $\chi_i = -1$ or 1, and $\omega_{i,j} \in (0, 1]$ is the weight of the edge connecting $v_{c,i}$ and $v_{c,j}$. Each energy value also has a corresponding energy-minimizing probability

$$P(X = \chi) = \frac{1}{z} e^{-H(\chi)},$$

with normalization factor z. We show an example cognitive network and configuration in Figure 2-1.

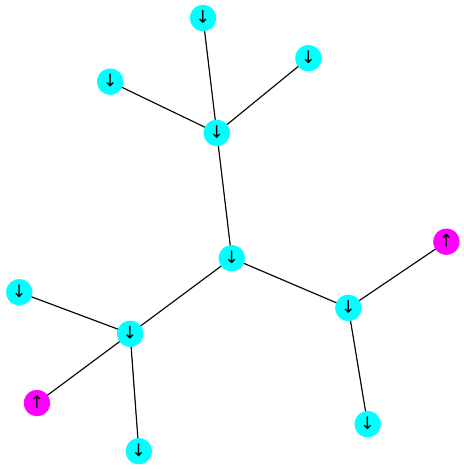

Figure 2-1 An example of a cognitive network occupying a possible set of belief (spin up) and disbelief (spin down) states. Connected nodes with the same state are in consistency, where as those in opposing states are dissonance. Consistent and dissonant pairs decrease and increase the Hamiltonian energy of the system, respectively.

2.1.2. Cognitive spread on a social network

We consider the social network S, in which nodes are agents within a simulation. For simulations using this model, one agent begins with a complete cognitive network C with all cognitive nodes in the spin-up state, and all other agents have a null cognitive network. At each timestep, a separate random event of probability ρ determines whether each agent receives each cognition that each neighbor holds in the spin-up state. Agent i's cognitive network at time t, $c_{i,t}$ is therefore a subgraph of C, with all links between pairs preserved.

Once the synchronous passage of cognitive nodes is determined for the timestep, the state of each agent's cognitive network must be determined. Agent i's predisposition toward belief of cognition j $\tau_{i,j}$ is computed as

$$\tau_{i,j,t} = s_{i,j,t}\chi_{i,j,t-1} + \phi p_{i,j}(1 - s_{i,j,t}),$$

where $s_{i,j,t}$ is the stickiness of agent i's belief state for cognition j at time t, $\chi_{i,j,t-1}$ is their belief state for that cognition at the previous timestep (assumed to be 0 if the agent was previously unexposed to that cognition), ϕ is a social parameter determining how much influence the neighbors of an agent have, and $p_{i,j}$ is the proportion of their neighbors from whom they've received the cognition.

Stickiness $s_{i,j,t}$ is calculated with a recursive processes. If the cognition is novel to the agent, or the agent has switched the state of that cognition, the stickiness assumes the value s, referred to as "base stickiness". Otherwise the value of stickiness grows as

$$s_t = s_{t-1} + r(1 - s_{t-1}),$$

where r is a predefined stickiness growth rate. With the values of $\tau_{i,j,t}$ computed, each agent has their cognitive network state reconfigured according to the probability distribution described in Section 2.1.1.

2.1.3. Extension to dual online/offline network

We extend the multiple-cognition social network spreading mechanism described in section 2.1.2 to a dual-layer network in order to capture the discrepancy between online and offline interpersonal relationships. The parameters for spreading are may thus be different between the layers, even between the same pair of neighbors should they share both an online and offline edge. Our offline network is specified as a Watts-Strogatz small-world network with $n = 20, k = 4, p = .2$, and our online network is specified with a two-community stochastic block model with $n = 12$ and variable density values.

In order to initialize an existing echo-chamber effect at the simulation start, we define one community of the stochastic block model for the online to be "credulous", meaning they are more likely to adopt the cognitions within the simulated narrative. We specify that the other community is "skeptical", meaning they are likely to reject the cognitions. This specification corresponds to τ_0 values of ± 0.2 if they receive the cognition via an offline connection,and ± 0.3 if they receive it via an online connection. Agents that only exist in the offline layer have $\tau_0 = 0$. These specifications only determine the initial behaviors of agents upon being exposed to a cognition. After the first exposure an agent's behavior with respect to that cognition is determined according to the spreading rules defined in section 2.1. An example of narrative spread over time with this dual-layer topology is shown in Figure 2-2.

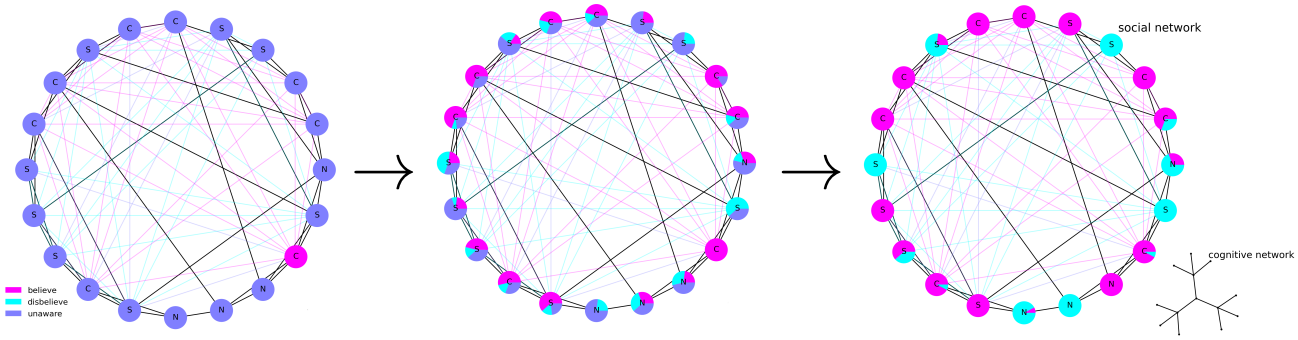

Figure 2-2 An example of the complete spread of a narrative on a social network projected into online and offline components. Agents are represented with pie charts that display the proportion of cognitions that they adopt (red), reject (blue), and have not yet been exposed to (blue). The online block of each node is denoted by the letter S for "skeptical", C for "credulous", and N for "neutral" (meaning the agent is offline-only). The three timesteps shown are the initial state ($t = 0$), partway through the spread ($t = 3$), and after every agent has been exposed to every cognition ($t = 13$).

2.2. Experiment: online network structure

While the intent of the work is to present a model for general use, we demonstrate its utility with a simple experiment. We design a scheme to investigate the role that different topologies for online connectedness play in the proliferation of narrative structures. We do this by conducting a parameter sweep on the within-group and between-group densities specified for the stochastic block model that represents our online social network structure. We run 10 simulations for 100 timesteps at each pair of values for within-group and between-group densities from 0.1 to 1.0 in steps of 0.1. This culminates in a total of 1,000 independent simulations over the 100 pairs of density values. For each independent run, we initialize a new random topology of each network layer, and choose a new random starting agent from the "credulous" community of the online network. We provide the values of all fixed parameters for this experiment in Table 2-1.

Parameter	Description	Value
ρ	Probability that an agent spreads their belief to a neighbor	0.2
ϕ	Influence of neighbors on belief	0.8 if belief received offline 0.4 if belief received online
s	Stickiness base	0.1
r	Stickiness growth rate	0.2
τ_0	Initial predisposition for online agents	$\pm$ 0.2 if belief received offline $\pm$ 0.3 if belief received online
ω	Cognitive link strength	0.8

Table 2-1 Fixed parameters for our simulations and their values. Value ranges for non-fixed parameters are specified in Section 2.2. Values that include a $\pm$ are positive for "credulous" agents and negative for "skeptical" agents.

3. RESULTS

While we observe differences in the details between behavior at different density values, the larger trend toward polarization is universal. Figure 3-1 displays a snapshot of every simulation at the tenth timestep, which is amid the transient period before full polarization for most simulation runs. We use the variance of these distributions as a proxy for aggregate polarization at each density value pair, which we show at four different timesteps in Figure 3-2.

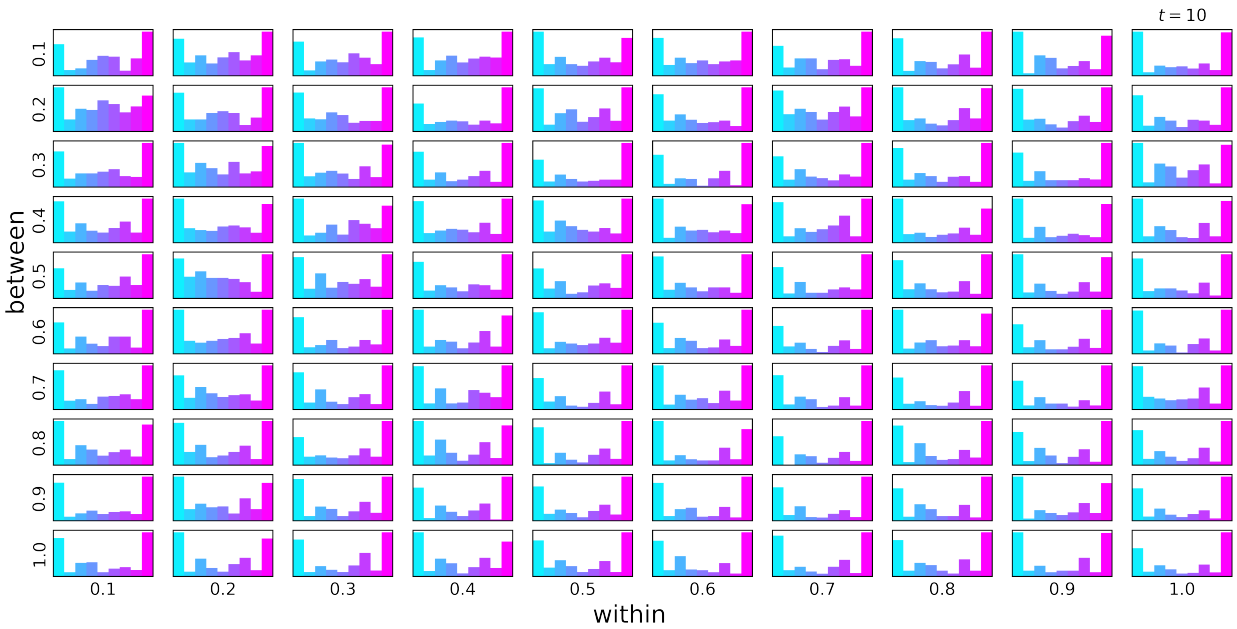

Figure 3-1 Snapshot at timestep 10 of the simulation states for each density pair. Each cell contains the histogram showing the proportion of cognitions believed by each agent in each simulation with the between-group density indicated by the row and the within-group density indicated by the column.

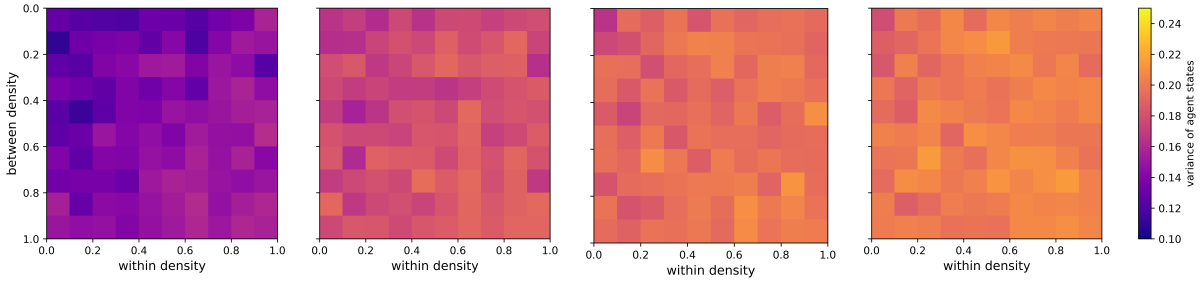

Figure 3-2 Variances of the distributions of the proportions of cognitions believed by agents for each density value pair at times $t = 10, 30, 50, 70$, from left to right.

The similarity between the distribution variances at $t = 50$ and $t = 70$ suggests that transient behavior of the system has largely concluded after 50 timesteps. To investigate the temporal length of this transient behavior, we compute the time taken in each simulation for the average reach of each cognition to surpass 17 agents. We present a

breakdown of the time taken to reach 17 agents on average by within and between density in Figure 3-3.

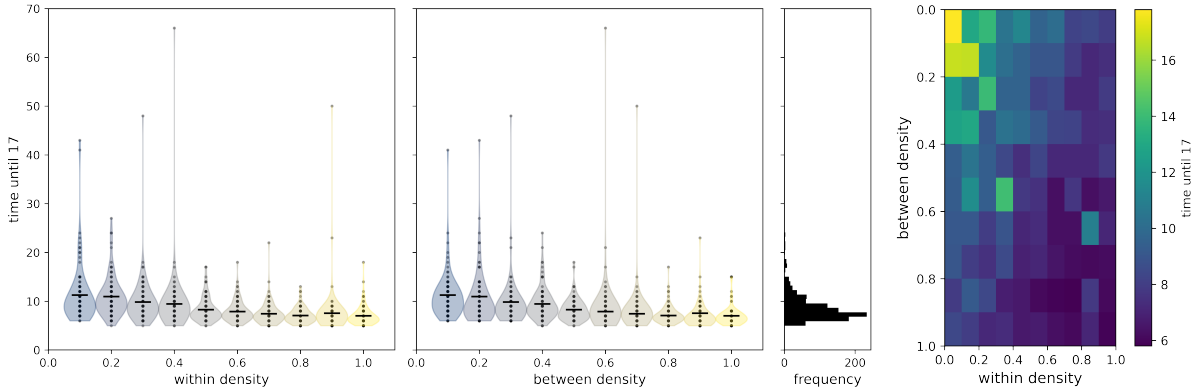

Figure 3-3 Time taken for the reach of cognitions in a simulation to surpass 17 agents. Broken down by within density, between density, in aggregate, and broken down by both within and between density, from left to right. The two left panels contain violin plots showing the distributions of times taken for cognitions within individual simulations to reach 17 agents at each density value. Points are also included to show results of individual cognitions, with an opacity of 0.3, allowing many points in the same position to appear darker. The third panel displays a histogram showing the aggregate distribution over every simulation at all density values. The final panel shows the average time to reach 17 agents over all cognitions, and all simulations at each value-pair for within and between density.

Figure 3-3 indicates a heavy-tailed distribution of cognition spreading rates, where most spread to 17 agents in between 5 and 20 timesteps. In a few rare instances, a cognition takes longer than 40 timesteps. As expected, higher densities result in faster spreading, although there isn't a clear difference in effect between the within-block density and between-block density.

In addition to the universal emergence of polarization stemming from this model, we observe the flipping of some agents from their initial dispositions to an opposing set of beliefs. To measure this, we recorded the number of agents at the 100th timestep whose proportion of spin-up cognitions was 1.0 if they began in the S block, or was 0.0 if they were in the C block. We display this final distribution and it's breakdown by density in Figure 3-4. We find that this "side-switching" is a universal behavior alongside polarization, and while it does vary between simulations, this variation is not impacted noticeably by online network density.

4. DISCUSSION

Misinformation, polarization, and hate movements are widely viewed as some of the most imminent and existential threats to national security, democracy, and human well-being today. Despite often being considered as separate issues, these phenomena are inseparably linked. While efforts are in place to combat these issues and to understand them, there is still a lack of holistic understanding necessary to optimally address them. Many forms of

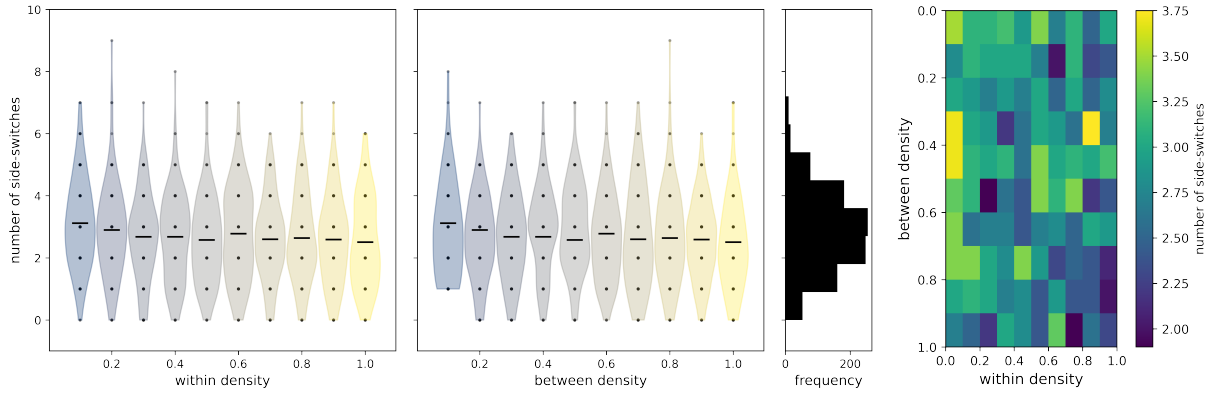

Figure 3-4 Number of agents whose final belief state opposed their initial predisposition toward the connected narrative. Broken down by within density, between density, in aggregate, and broken down by both within and between density, from left to right. Plotting details are identical to those in Figure 3-3.

study are necessary to fully understand these problems and their societal impacts, including retrospective data investigations, analyses of the systems that allow and propagate harm, and computational modeling to simulate these systems.

Among existing agent based modeling techniques are models of disinformation spreading and models of cognitive consistency and dissonance. The model we've presented intends to capture both spreading and cognitive consistency, as well as a deliberately specified duality between online and offline social interaction. Such an extension of both these methods brings the social and cognitive simulation community closer to capturing reality with agent-based modeling.

Experimental results using this model confirmed some expectations and defied others. As with previous models of cognitive consistency on social networks, polarization among agents emerged universally for parameter ranges included in this study. While the rate of transmission did become more rapid with higher social network density, distributions were long-tailed, with only a few extreme, slow-transmission events. Another universally observed behavior was side-switching, in which some collection of agents reached a homogenous final cognitive state that opposed their initial predisposition, which is by definition the same predisposition as their online community. This side-switching, however, did not change with density of the online network, and distributions of side-switching frequency were nearly symmetric.

Although these insights can be drawn from experimentation we've conducted, this writing serves to mainly present a useful model, and our tests so far have been limited. We do, however, encourage readers to run their own simulations using our model with focus their parameters of interest. In addition to exploring the large parameter space, there exist numerous extensions to the current model to better approximate the dynamics of online and offline social interaction. We've considered removing the permanence of links in the social network, allowing agents to strengthen links to others who share an attitude toward the cognitive network, and weaken those toward agents with opposing attitudes. This could mimic the real-life phenomenon of people cutting off family members and friends in

favor of an internet community, pushing them deeper and deeper into a conspiracy "rabbit-hole". The chosen topology also favors simplicity. Different behaviors of the system as a whole could be revealed by transferring the model to scale-free network or Lancichinetti–Fortunato–Radicchi benchmark. The ability to extend this model to large networks (thousands or tens of thousands of nodes) would be ideal for simulating a societal phenomenon accurately, but the exponential scaling of the model makes such an extension intractable with current computing technology.

This work sits upon a lineage of cognitive, psychological, and computational research to understand and address disinformation and conspiratorial narratives. As we continue into our next phase of research toward that end, we look forward to the advances made by our colleagues in this field at this and other institutions.

REFERENCES

[1] P. S. Dodds and D. J. Watts, "A generalized model of social and biological contagion," *Journal of theoretical biology*, vol. 232, no. 4, pp. 587–604, 2005.

[2] P. S. Dodds and D. J. Watts, "Universal behavior in a generalized model of contagion," *Physical review letters*, vol. 92, no. 21, p. 218 701, 2004.

[3] P. S. Dodds and J. L. Payne, "Analysis of a threshold model of social contagion on degree-correlated networks," *Phys. Rev. E*, vol. 79, p. 066 115, 2009.

[4] D. J. Watts and P. S. Dodds, "Threshold models of social influence," in *The Oxford Handbook of Analytical Sociology*, P. Hedström and P. Bearman, Eds., Oxford, UK: Oxford University Press, 2009, ch. 20, pp. 475–497.

[5] J. P. Gleeson and D. J. Cahalane, "Seed size strongly affects cascades on random networks," *Physical Review E*, vol. 75, no. 5, p. 056 103, 2007.

[6] D. J. Watts and P. S. Dodds, "Influentials, networks, and public opinion formation," *Journal of Consumer Research*, vol. 34, pp. 441–458, 2007.

[7] Y.-S. Chiang, "Birds of moderately different feathers: Bandwagon dynamics and the threshold heterogeneity of network neighbors," *Journal of Mathematical Sociology*, vol. 31, no. 1, pp. 47–69, 2007.

[8] D. M. Centola, "Homophily, networks, and critical mass: Solving the start-up problem in large group collective action," *Rationality and society*, vol. 25, no. 1, pp. 3–40, 2013.

[9] N. A. Christakis and J. H. Fowler, "The collective dynamics of smoking in a large social network," *New England journal of medicine*, vol. 358, no. 21, pp. 2249–2258, 2008.

[10] T. C. Schelling, "Dynamic models of segregation," *Journal of mathematical sociology*, vol. 1, no. 2, pp. 143–186, 1971.

[11] A. E. Marwick, "Why do people share fake news? a sociotechnical model of media effects," *Georgetown Law Technology Review*, vol. 2, no. 2, pp. 474–512, 2018.

[12] A. Guess, J. Nagler, and J. Tucker, "Less than you think: Prevalence and predictors of fake news dissemination on facebook," *Science advances*, vol. 5, no. 1, eaau4586, 2019.

[13] M. Del Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Echo chambers: Emotional contagion and group polarization on facebook," *Scientific reports*, vol. 6, p. 37 825, 2016.

[14] P. Törnberg, "Echo chambers and viral misinformation: Modeling fake news as complex contagion," *PloS one*, vol. 13, no. 9, e0203958, 2018.

[15] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, "The spreading of misinformation online," *Proceedings of the National Academy of Sciences*, vol. 113, no. 3, pp. 554–559, 2016.

[16] A. Bovet and H. A. Makse, "Influence of fake news in twitter during the 2016 us presidential election," *Nature communications*, vol. 10, no. 1, pp. 1–14, 2019.

[17] A. Boutyline and R. Willer, "The social structure of political echo chambers: Variation in ideological homophily in online networks," *Political Psychology*, vol. 38, no. 3, pp. 551–569, 2017.

[18] A. Amarasingham, "Terrorism on the teardrop island: Understanding the easter 2019 attacks in sri lanka," *CTC Sentinel*, vol. 12, no. 5, pp. 1–10, 2019.

[19] M. A. Taylor and D. S. Stoltz, "Concept class analysis: A method for identifying cultural schemas in texts," *Sociological Science*, vol. 7, no. 23, pp. 544–569, 2020, ISSN: 2330-6696. DOI: 10.15195/v7.a23. [Online]. Available: http://dx.doi.org/10.15195/v7.a23.

[20] J. P. Houghton, *Interdependent diffusion: The social contagion of interacting beliefs*, 2021. arXiv: 2010.02188 [cs.SI].

[21] B. Keith and T. Mitra, *Narrative maps: An algorithmic approach to represent and extract information narratives*, 2020. arXiv: 2009.04508 [cs.HC].

[22] K. Lakkaraju, "Modeling attitude diffusion and agenda setting: The mama model," *Social Network Analysis and Mining*, vol. 6, no. 1, pp. 1–13, 2016.

[23] J. Dalege, D. Borsboom, F. van Harreveld, H. van den Berg, M. Conner, and H. L. van der Maas, "Toward a formalized account of attitudes: The causal attitude network (can) model.," *Psychological review*, vol. 123, no. 1, p. 2, 2016.

[24] H. L. van der Maas, J. Dalege, and L. Waldorp, "The polarization within and across individuals: The hierarchical ising opinion model," *Journal of Complex Networks*, vol. 8, no. 2, cnaa010, 2020.

DISTRIBUTION

Email—Internal (encrypt for OUO)

Name	Org.	Sandia Email Address
Technical Library	01177	libref@sandia.gov

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.